

## VISIÓN DE MÁQUINA APLICADA A SISTEMAS DE SEGURIDAD

Jorge Enrique Ramírez Neira\*

**Resumen.** Dentro de las aplicaciones del procesamiento digital de señales, el trabajo con imágenes de vídeo y la incorporación de servomecanismos que faciliten una vía de retroalimentación, se han convertido en una rama denominada visión activa o visión de máquina. En este artículo se describe el desarrollo de un sistema de bajo costo que, con base en esta tecnología, provee gran funcionalidad y aplicabilidad en el área de vigilancia y seguridad de recintos con bajos niveles de movilidad. Funciones autónomas como el seguimiento de objetos en movimiento, grabación de imágenes en formato digital, control de dispositivos remotos, son algunos de los aspectos a resaltar. Se destaca el uso de herramientas de desarrollo de vanguardia y las técnicas de programación utilizadas.

**Abstract.** Among existing applications for Digital Signal Processing, the work done with video images and the use of servo-mechanisms to facilitate feedback have become a new field known as Active Vision or Machine Vision. This paper describes the development of a low-cost system based on this technology that is highly functional and has several applications in the areas of vigilance and security of rooms and spaces with low levels of movement. Autonomous functions, such as following moving objects, recording images in digital format and controlling remote devices, are important aspects to be mentioned; the use of cutting-edge development tools and programming techniques is outstanding.

### 1. INTRODUCCIÓN

El procesamiento de imágenes de vídeo como área de estudio del procesamiento digital de señales es un campo que se ha trabajado desde hace varios años. Su misma extensión ha generado diversas vertientes dentro de las que los sistemas de visión activa o visión de máquina ocupan un importante lugar en el desarrollo de automatismos y productos para supervisión industrial y seguridad. Estos sistemas se caracterizan por incluir bloques motorizados, que permiten modificar

---

\* Ingeniero electrónico, Pontificia Universidad Javeriana.

la posición de la cámara o dispositivo que se encarga de capturar la imagen. Su importancia en cuanto a sistemas de vigilancia, radica en su facilidad para trabajar de manera continua y autónoma.

A continuación se presenta un sistema de bajo costo en el campo de vigilancia y seguridad aprovechando las técnicas de visión activa, que de forma autónoma registra y supervisa hasta seis señales de vídeo simultáneas, provenientes de cámaras blanco y negro ubicadas en sitios estratégicos en recintos alejados con bajos niveles de movilidad de personas. El procesamiento de imagen permite que para cada cámara se pueda modificar su posición con un grado de libertad de forma manual o automática, utilizando para esto último un algoritmo para detección y seguimiento de objetos en movimiento. Algoritmos especializados para realizar compresión y decompresión de vídeos, conocidos como *video codecs*, instalados en el sistema operativo, permiten almacenar información de vídeo en tiempo real sobre medio magnético con una tasa de ocupación razonable. Seis módulos remotos de procesamiento autónomo soportan las cámaras y se comunican con un servidor principal a través de una red de datos utilizando el estándar eléctrico RS-485 y un protocolo de comunicaciones propietario con información adicional para detección de errores. Cada módulo está en capacidad de controlar el encendido/apagado de cuatro cargas de potencia de forma manual, programada o automática. Adicionalmente, estos módulos por su fabricación, garantizan alta resistencia y durabilidad tanto física como electrónica.

Un computador convencional tipo PC, operando como servidor, realiza el procesamiento, la visualización y el despliegue de funciones al usuario de manera independiente para cada cámara. Una serie de funciones adicionales garantiza un seguimiento constante de cada módulo, detectando acciones que puedan atentar contra la integridad de los mismos.

## 2. EL PROCESAMIENTO DE IMAGEN

Dentro del procesamiento de imágenes para detección y seguimiento de movimientos existen varias técnicas, dependiendo de la velocidad de procesamiento y el tipo de seguimiento que se desee realizar. Para este caso en particular se concentró el esfuerzo en desarrollar un algoritmo para detectar cualquier clase de movimiento y poder retornar a una posición precisa del lugar donde ocurrió.

Tomando un número constante de imágenes consecutivas en tiempo, se calcula, mediante operaciones de suma y resta, la zona que ha sufrido cambios (Figura 1). En la imagen resultante se observa el proceso de binarización mediante el cual la zona de movimiento se resalta con puntos blancos de valor  $P_w$  y la restante con puntos negros de valor  $P_B$ . El valor de umbral en este proceso determina el nivel de sensibilidad.

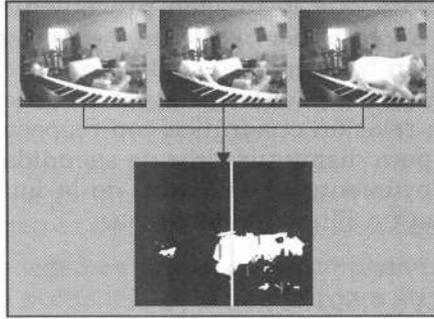
La imagen original tiene un tamaño de  $H = 640 \times V = 480$  puntos, tamaño que resulta difícil de manejar para operaciones de tiempo real. Por esta razón, durante el análisis de imagen solamente se toman

puntos equidistantes horizontal y verticalmente de la matriz original (Ecuación 1).

$$P[x,y] = M \left[ (x \times G) + \left( \frac{G-1}{2} \right), (y \times G) + \left( \frac{G-1}{2} \right) \right] \quad (1)$$

Donde P es la matriz de procesamiento resultante, M la matriz de imagen original, G, un número entero que representa la constante de espaciamiento y H y V deben ser múltiplos enteros de G. De esta manera se genera una grilla invisible y una imagen de procesamiento de resolución  $P_H = (H / G)$   $P_V = (V / G)$ , con  $P_H$  y  $P_V$  enteros.

Figura 1. Procesamiento para detección de movimiento



A medida que la iluminación en el recinto disminuye, la relación señal/ruido de la señal de vídeo de la cámara aumenta; esto conduce a un incremento en la probabilidad de una falsa detección. Para disminuir la relación señal/ruido proveniente de la cámara, se calculan promedios en dos dimensiones tomando un conjunto de puntos en la vecindad del punto de análisis (Ecuación 2). Este conjunto es un cuadrado de  $G \times G$ , con un valor de G siempre impar. La nueva matriz de procesamiento es entonces:

$$P[x,y] = \left( \sum_{j=-((G-1)/2)}^{(G-1)/2} \sum_{k=-((G-1)/2)}^{(G-1)/2} M \left[ (x \times G) + \left( \frac{G-1}{2} \right) + j, (y \times G) + \left( \frac{G-1}{2} \right) + k \right] \right) / G^2 \quad (2)$$

En la siguiente etapa se estima la posición hacia la que debe girar la cámara. Para ello se calcula un promedio de promedios en franjas horizontales solamente en las zonas de detección de movimiento (Ecuaciones 3 y 4).

$$F(k,y) = \begin{cases} P[k,y] = P_W, & k \\ P[k,y] = P_B, & 0 \end{cases} \quad (3)$$

$$X_{\text{Linea}}(y) = \left( \frac{P_H}{2} \right) - \left( \sum_{k=1}^{H/G} F(k,y) \right) / P_H \quad (4)$$

$$X_{Cuadro} = \left( \sum_{k=1}^{P_V} X_{Linea}(k) \right) / P_V \quad (4)$$

Donde  $X_{Linea}(y)$  es el valor promedio de posición de movimiento por línea y  $X_{Cuadro}$  es el valor promedio de posición para toda la imagen. Esta posición permitirá definir en qué dirección y con qué velocidad girará la cámara. Por su naturaleza este algoritmo es altamente eficiente en zonas en las que se estimen muy bajo niveles de movilidad de personas y objetos, como en bodegas, depósitos de almacenamiento de mercancía, recintos de trabajo de jornada no continua, entre otros.

Teniendo en cuenta que la imagen proveniente de la cámara se distorsiona por el efecto óptico del lente, se realizó una caracterización del mismo y, con base en ello, se diseñó una tabla de valores de conversión para la posición entregada por la función anterior. Esta tabla conserva una relación cuadrática con respecto a la velocidad de giro de la cámara, para disminuir giros de sacudida fuertes de la misma cuando hay movimiento en el centro de la imagen. (Ecuación 5) Así mismo se utiliza un filtro tipo mediana.

$$Vel(T_{POS}) = V\_MAX - 4,59136 \times 10^{-3} \cdot T_{POS}^2 \quad (5)$$

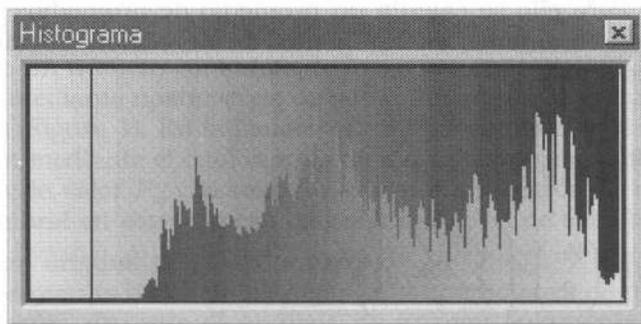
Donde

$T_{POS}$ : posición retornada por la tabla de conversión

$V\_MAX$ : velocidad máxima de giro de la cámara.

Con base en la información de tonalidades de gris en la imagen, se calcula un histograma con cuya distribución se realiza una estimación del nivel de luminancia y contraste (Figura 2). De esta forma, cada imagen se adapta automáticamente al nivel de luz en el recinto dentro de un rango superior al que provee el mecanismo de auto-iris incorporado en la cámara.

Figura 2. Histograma para cálculo de luminancia y contraste



### 3. EL PROGRAMA DE COMPUTADOR

El computador que trabaja como servidor realiza las operaciones de procesamiento y despliegue de información a través de un programa que se desarrolló utilizando la herramienta *Microsoft Visual C++ 6.0*, denominado *VMASIS versión 3.0*. Grandes ventajas se encuentran en la selección de la herramienta de desarrollo, entre otras: control de bajo nivel para recursos gráficos del sistema operativo, control detallado de ejecución para rutinas y secuencias de tratamiento matemático, optimización de código para mejora en la velocidad de ejecución y facilidad de crecimiento gracias a su diseño orientado a objetos.

El programa *VMASIS versión 3.0* permite al usuario desplegar simultáneamente hasta seis imágenes de vídeo, cada una con controles independientes que le permiten manipular todas las funciones de cada una de las cámaras, entre otras, giro manual y automático, encendido y apagado de puertos en forma manual, automática o programada, visualización de histograma, grabación de vídeo en forma manual y automática utilizando algoritmos de compresión de vídeo y configuración de todos los parámetros de procesamiento de imagen.

La grabación de vídeo es un proceso completamente automatizado de forma que cada secuencia se nombra automáticamente con fecha, hora, número de cámara en una ubicación dentro del sistema de archivos que es seleccionada con el uso de criterios de número de cámara y fecha. Adicionalmente, cada secuencia de vídeo incluye código de tiempo insertado, que permite obtener precisión de segundos en la fecha de la grabación.

La tasa de ocupación de cada secuencia de vídeo depende del compresor seleccionado, dentro de los que se encuentren instalados en el sistema operativo. Pruebas de desempeño mostraron que con el compresor *Intel Indeo 5.04* se obtiene un valor promedio de 1 Megabyte por minuto con una calidad del 85%, una resolución de 320 x 240 y una frecuencia de 10 a 15 cuadros por segundo. Esto es equivalente a una ocupación entre 50 y 100 veces menor a la que se tiene grabando la información sin utilizar compresión.

Para cada cámara es posible definir un umbral de detección diferente, número de cuadros por segundo y tiempos de espera, así como rutinas de giro que permiten automatizarlas completamente en los instantes en que no se detecta movimiento. Pensando en eventualidades comunes en seguridad, el sistema puede detectar corte de cables, falla en la alimentación o destrucción de cualquiera de los módulos. Igualmente puede detectar situaciones como la sustitución de señal de vídeo gracias a la utilización de un código de teletexto inserto en la señal.

Las características del sistema operativo permiten, asimismo, utilizar múltiples monitores en un solo computador (Figura 3). Para ello, un monitor adicional permite visualizar la cámara que se desee en la tota-

lidad de área de la pantalla, dando así mayor facilidad en la ubicación de detalles. Tal carga de recursos gráficos es enrutada a los sistemas de transferencia de imágenes que vienen incorporados en circuitos electrónicos en las tarjetas de control gráfico y que pueden ser aprovechados gracias a la utilización de últimas tecnologías de aceleración gráfica (*Hardware Abstraction Layer-HAL*) agrupadas por *Microsoft Direct X*. Esto permite que el procesador principal no se ocupe de los recursos gráficos mientras realiza las operaciones de procesamiento de imágenes.

Figura 3. Múltiples monitores y acercamiento digital



#### 4. LOS MÓDULOS ELECTRÓNICOS Y LOS DESARROLLOS ANEXOS

El sistema de captura de imagen *EllipsRio* embebido en el computador principal permite la adquisición simultánea de dos señales de vídeo no correlacionadas gracias a los dos procesadores de gran escala que incorpora. Es posible entonces, obtener una tasa de 30 cuadros por segundo en cada uno de estos procesadores.

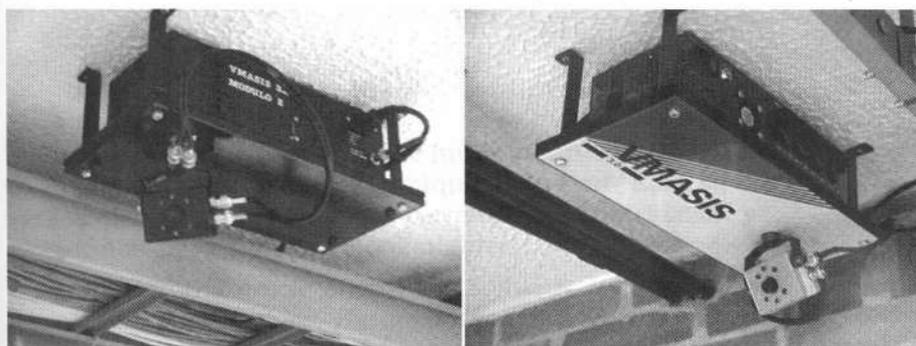
En el desarrollo del sistema se utilizaron cámaras de vídeo diseñadas para propósitos de vigilancia que incorporan emisores de luz infrarroja y circuitos CCD sensibles a la misma que incrementan su capacidad de visión en circunstancias de absoluta oscuridad, con lo cual supera el umbral de sensibilidad de un ojo humano.

Para lograr la adquisición de las seis imágenes ahora se desarrolla en el Departamento de Electrónica de la Pontificia Universidad Javeriana un sistema de sincronización de señales de vídeo con el cual se logra una velocidad de adquisición de hasta 10 cuadros por segundo en cada cámara. Adicionalmente, se incorporará un sistema

de inserción de código de teletexto compatible con el estándar para personas con discapacidad auditiva.

Para la identificación de cámara y detección de ausencia de señal se utilizó un sistema de *closed caption*. Mediante seis módulos remotos que soportan cada una de las cámaras se incorpora un sistema microcontrolado que sirve como interpretador de comandos y que incluye, además, un microcódigo diseñado utilizando técnicas de programación orientada a objetos. Estos módulos, diseñados para instalación en techo (Figura 4), disponen de los mecanismos necesarios para mover una cámara que puede pesar hasta 250 libras o, de igual manera, circuitos de potencia para el control de puertos remotos.

Figura 4. Módulos remotos instalados



Dichos puertos remotos pueden ser utilizados para incorporar al sistema control de luces, cantoneras eléctricas, sirenas, alarmas silenciosas y, en general, cualquier dispositivo eléctrico que utilice voltaje de línea para alimentarse y que no exceda un consumo de 360 watts.

La comunicación entre los módulos remotos y el computador principal se realiza a través de una red de datos privada que se encuentra también en desarrollo. Para esta red se utiliza el estándar eléctrico RS-485 y una comunicación serial multi-punto manejando un protocolo similar a *Frame Relay*. Esta red fue diseñada pensando en proyectos para seguridad y por ello se decidió robustecer su fiabilidad incorporando sistemas de detección y corrección de errores.

## 6. RESULTADOS OBTENIDOS

Después de tres años de desarrollo y puesta en marcha del sistema, éste se encuentra en operación con dos módulos completamente terminados que se han instalado en el cuarto y quinto piso de la Facultad de Ingeniería para propósitos de vigilancia. Una vez culminen los proyectos anexos se procederá a efectuar la instalación de cuatro módulos adicionales. De esta manera se incorporan últimas tecnologías en desarrollo electrónico y en programación, en una situación de gran utilidad en nuestro medio.

El sistema presenta importantes ventajas —aparte de su comparativamente reducido costo— que difícilmente se obtienen en sistemas de venta comercial; entre ellas se encuentran el control centralizado desde un computador de escritorio, el control de puertos, el seguimiento automático de objetos, la utilización de últimas tecnologías de aceleración gráfica y la grabación de vídeo en forma automática. Otra bondad del sistema radica en que las técnicas de programación orientada a objetos aplicadas tanto en el programa de computador como en el código de los microcontroladores permiten una fácil expansión y modificación de las funciones actuales del sistema.

