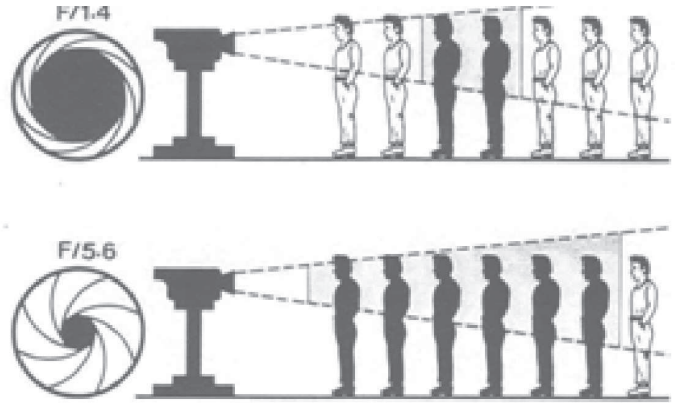
Sistema de visión artificial para conteo de objetos en movimiento

Andrés Felipe Loaiza Quintana[*]
David Andrés Manzano Herrera[**]
Luis Eduardo Múnera Salazar[***]



Resumen

En este artículo se hace una introducción a diferentes conceptos y metodologías utilizadas en la construcción de sistemas de visión artificial para el reconocimiento de patrones; así mismo, se describen los resultados obtenidos durante el estudio y desarrollo de un sistema de visión artificial para el conteo de personas que se mueven dentro de un espacio cerrado. El objetivo de este tipo de sistemas es el de obtener información de interés, que se pueda interpretar, a partir de un complejo procesamiento de imágenes obtenidas por medio de dispositivos ópticos. En este proceso se deben tener en cuenta varios aspectos: el método para obtener las imágenes, las diferentes técnicas de procesamiento de las mismas y la manera en que se va a interpretar la información obtenida a través de esos procesos de manipulación. Cualquiera sea el objeto que se desea identificar, la precisión de las aplicaciones y la calidad de los resultados dependerán: del nivel de conocimiento que se tenga acerca del objeto de estudio, la resolución de los dispositivos con los que

(*) Universidad Icesi, Cali, Colombia. Correo electrónico: andres.loaiza@correo.icesi.edu.co.
(**) Universidad Icesi, Cali, Colombia. Correo electrónico: damanzano@icesi.edu.co.
(***) Universidad Icesi, Cali, Colombia. Correo electrónico: lemunera@icesi.edu.co.
Fecha de recepción: 16/10/2012 • Fecha de aceptación: 12/12/2012.

se obtienen la imágenes y la capacidad de procesamiento de los equipo en que se ejecutan las aplicaciones; sin embargo, estos dos últimos aspectos pueden ser menos importantes si se logra identificar aquellas características de los objetos que permiten basar el algoritmo de visión artificial sobre abstracciones de los objetos, en lugar de realizar una comparación exhaustiva de los reales.

Palabras clave: visión artificial, conteo de personas, reconocimiento de objetos.

Abstract

This article introduces different subjects and methodologies used for building pattern recognition artificial vision systems and expose the results found during the study and development of a people counting artificial vision system for indoors. Getting valuable information through a complex image processing is the goal of this kind of software systems. In this process it must be take in account several factors: the method for getting the images, the way to manipulate them and the way to interpret information about those manipulations. Whichever object you follow, the applications accuracy and the quality of the results depend on: the knowledge level about studied object, the devices' resolution and the processing capabilities of the computer that executes the applications, however, these two last points could be less important if it is possible to identify object's features that allow to support the artificial vision algorithm on object's abstractions instead of making and exhaustive comparisons on the real object.

Keywords: artificial vision, people countig, objetcs and pattern recognition.

1. Introducción

El flujo de personas va en aumento en todos los ámbitos de la vida pública. En todas partes (la calle, los centros comerciales, los aeropuertos, etc.) hay grandes cantidades de personas en movimiento aparentemente desordenado. Si se registran sistemáticamente estos movimientos y la cantidad de personas que transitan, los datos obtenidos se pueden utilizar comercialmente. El problema es que, hasta el momento, esas mediciones se realizan de forma manual y son susceptibles al error humano.

Esta información resulta especialmente importante para las empresas, pues les ayuda a tomar decisiones en sectores como: manejo de personal, mercadeo, logística y manejo de la seguridad. Por tal razón, ha surgido en la comunidad dedicada al desarrollo de inteligencia artificial, en especial en la rama de visión artificial, el interés de automatizar el proceso de conteo y seguimiento de los movimientos de las personas.

El área de visión por computador se interesa en la solución de dos problemas fundamentales: la mejora de la calidad de las imágenes para la interpretación humana y el procesamiento de los datos de la escena para la percepción de las máquinas, de forma autónoma.

La primera trata de mejorar las técnicas y tecnologías que permiten obtener una mejor calidad en las imágenes, para que de ellas se pueda extraer más información. Este problema está por fuera del alcance de este proyecto.

La segunda se concentra en otorgar a las máquinas la capacidad para ver el mundo que les rodea, más precisamente para deducir la estructura y las propiedades del mundo tridimensional, a partir de una o más imágenes bidimensionales (González & Woods, 2001). En este tema se ubica el propósito del presente estudio.

El objetivo de este artículo es determinar los procesos y técnicas de procesamiento de imágenes, necesarios para el desarrollo de un sistema de visión artificial destinado al conteo de objetos en movimiento, en especial de personas. Y el diseño de un prototipo que permita mostrar las formas de obtener información valiosa a partir del mismo.

2. Marco teórico

En la actualidad los sistemas de detección y reconocimiento de objetos se pueden clasificar en dos tipos: los que hacen uso de diferentes tipo de sensores (de proximidad, térmicos, infrarrojos, entre otros) y los se basan exclusivamente en la utilización de cámaras, siendo estos últimos sistemas de visión artificial. El sistema de conteo de objetos propuesto en este proyecto realiza su trabajo basado en las teorías de los sistemas de visión artificial.

Un sistema de visión artificial busca reconocer, analizar y entender una escena y sus componentes, partiendo de una o más imágenes bidimensionales. Entre los objetivos más comunes de estos sistemas están: la detección, segmentación, localización y reconocimiento de ciertos objetos en las imágenes; en el caso particular de este proyecto, el objetivo se centra en reconocer y contar el número de personas en movimiento.

A. Proceso de visión artificial

Todo sistema de visión artificial sigue el proceso general representado, de manera sintética, con el esquema de la Figura 1. En la imagen, las cajas representan datos y las burbujas procesos. Se parte de la escena tridimensional y se termina con la aplicación de interés.

Para poder realizar ese proceso, un sistema de visión artificial requiere de dos factores imprescindibles:

- Un sensor óptico para captar la imagen: una cámara de vídeo, una cámara fotográfica, una cámara digital, un escáner, etc., uniéndole un conversor analógico-digital cuando sea preciso (Vélez, Moreno, Sánchez & Sánchez Marín, 2003).
- Un computador que almacene las imágenes y que ejecute los algoritmos de preprocesado, segmentación y reconocimiento de la misma (Vélez et al., 2003).

En la visión por computador, la escena tridimensional es vista por una o más cámaras para producir imágenes monocromáticas o en color, que luego serán procesadas para obtener información relevante.

Figura 1. Esquema del sistema de visión Artificial

Fuente: elaboración propia.

B. Captura de las imágenes

Las imágenes de entrada del proceso pueden ser de diferentes tipos: imágenes de intensidad. Estas se encuentran ligadas estrechamente con el concepto de luminosidad; son aquellas imágenes que se adquieren a través de dispositivos ópticos basados en la captura de luz, como cámaras fotográficas y de video (González & Woods, 2001).

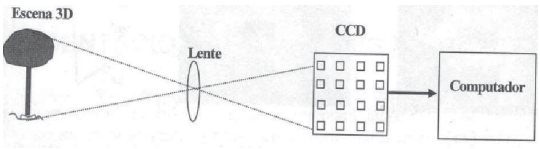
Imágenes de alcance (de profundidad o perfiles de superficie). Tienen su fundamento en los sensores de alcance óptico que utilizan algún fenómeno físico diferente a la luz para adquirir la imagen (por ejemplo el radar, el sonar, el láser) (González & Woods, 2001).

Se podrían contemplar otros tipos de imágenes, como por ejemplo las que son producidas por las cámaras térmicas. Pero, sin importar el medio de obtención de las imágenes, después del proceso de captura se obtiene una matriz 2D de valores, que se conoce comúnmente como una imagen digital (González & Woods, 2001). El sistema de conteo de objetos desarrollado en este proyecto se basa en el procesamiento de imágenes de intensidad. Los valores de la matriz 2D de una imagen de intensidad son valores de intensidad. En el caso de de una imagen de alcance, serían valores de profundidad; y en el de una imagen térmica, valores de temperatura.

En la Figura 2 se muestra el proceso de captura de una imagen de intensidad. Al respecto cabe realizar las siguientes anotaciones:

- El elemento que se muestra como lente se refiere al sistema óptico necesario en los dispositivos de captura.

Figura 2. Captura de una imagen digital de intensidad



Fuente: (González & Woods, 2001).

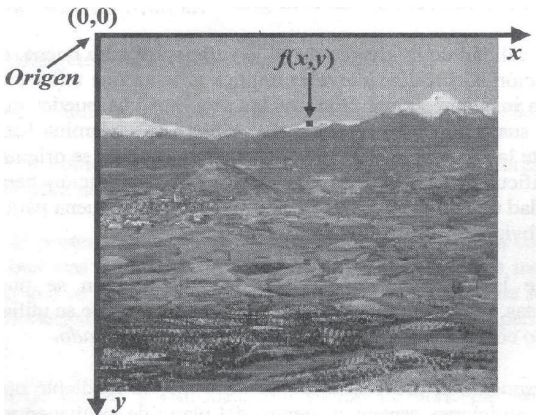
- El valor de intensidad que se obtiene para cada punto de la matriz 2D es el producto de un componente de iluminación y un componente de reflectancia. La primera es consecuencia de las fuentes de iluminación existentes en la escena en el momento de la captura. La segunda está asociada a las

propiedades intrínsecas del objeto. Desde el punto de vista de la teoría de señales, la iluminación corresponde con las bajas frecuencias de la imagen, mientras que la reflectancia está relacionada con las altas frecuencias y, por tanto, con los detalles de la imagen. Existen diferentes mecanismos para separar las componentes de iluminación y reflectancia (Pajares & De la Cruz, 2008). Entre ellos el filtrado homomórfico, que consiste en eliminar las bajas frecuencias de la imagen para quedarse solo con las altas y, en consecuencia, con los detalles (Pajares & De la Cruz, 2008).

C. Representación de imágenes digitales

El término imagen se refiere a una función de intensidad bidimensional, que se representa dependiendo de la literatura y el contexto como f(x, y), f(i, j), I(x, y), I(i, j), E(x, y), g(x, y), g(i, j), etc., donde x e y o bien i y j son las coordenadas espaciales y el valor de f, I, E o g en cualquier punto (x, y) o (i, j) es proporcional a la intensidad o nivel de gris de la imagen en ese punto.

Figura 3. Convención de ejes utilizada para la representación de imágenes digitales



Fuente: (González & Woods, 2001).

D. Segmentación

La segmentación es un proceso que consiste en dividir una imagen digital en regiones homogéneas con respecto a una o más características (como por ejemplo el brillo o el color), con el fin de facilitar un posterior análisis o reconocimiento automático (Vélez et al., 2003). Se basa en dos principios fundamentales: discontinuidad y similitud. Tienen también dos categorías: orientada a bordes (discontinuidad) y orientada a regiones (similitud) (González & Woods, 2001). En términos generales, una región es un área de la imagen en la que sus píxeles poseen propiedades similares (de intensidad, color, etc.), mientras que un borde es una línea que separa dos regiones de diferentes propiedades.

Cualquiera que sea la técnica de segmentación usada, el proceso implica manipular y transformar la imagen original en una que pueda procesarse fácilmente.

• Detección de bordes

Basándose en el hecho de la discontinuidad, se proponen los siguientes tipos de operadores para la detección de bordes: operadores primera derivada, operadores segunda derivada y operadores morfológicos.

• Detección de regiones

Para la detección de regiones se utilizan técnicas basadas en el hecho de la similitud: binarización, apoyada en el uso de umbrales; crecimiento de regiones, mediante la adición de píxeles; división de regiones y similitud de textura, color o niveles de gris.

En general, el proceso de la segmentación suele resultar complejo, debido, por un lado, a que no se tiene una información adecuada de los objetos a extraer y, por otro, a que en la escena a segmentar aparece normalmente ruido. Es por esto que el uso de conocimiento sobre el tipo de imagen a segmentar o alguna otra información de alto nivel puede resultar muy útil para conseguir la segmentación de la imagen (Vélez et al., 2003).

Cuando la calidad de la imagen no es buena y no es posible extraer la información de forma adecuada, es necesario aplicar técnicas de mejora en la calidad de la imagen original (González & Woods, 2001).

E. Descripción

Una vez se han destacado los bordes o las regiones como elementos de interés, el proceso de descripción consiste en extraer propiedades o atributos para su uso en las aplicaciones. En líneas generales, se trata de reconocer e identificar, de forma inequívoca, las diferentes estructuras de la imagen (González & Woods, 2001). Estructu-

ras que el observador considera necesarias para entender y comunicar su significado (Cantoni & Levialdi, 1996).

En el proceso de descripción, se identifican patrones con base en un significado funcional (Cantoni & Levialdi, 1996), es decir, patrones que comparten características similares y permiten identificarse de acuerdo con lo que representan en su forma geométrica, topológica, cromática y morfológica (Cantoni & Levialdi, 1996).

F. Técnicas de detección y seguimiento del movimiento

En particular, son tareas del sistema propuesto: diferenciar el fondo de la escena de los objetos que se encuentran en movimiento; distinguir cuáles de los objetos en movimiento detectados corresponden a una persona o al tipo de forma que se quiere reconocer; y hacer el seguimiento, dentro de la escena, a todos aquellos elementos que hayan sido reconocidos.

A continuación se describen las técnicas utilizadas para satisfacer las necesidades del sistema.

Substracción del fondo *(background subtraction).* La substracción de fondo es una técnica utilizada para la detección de los objetos en movimiento, cuya aplicación consiste en comparar cada uno de los *frames* (cada una de las imágenes que componen un video) de una secuencia de imágenes con el *frame* inicial u otro que se escoja a conveniencia, de tal forma que en el video resultante, los elementos que permanezcan constantes se vean de color negro, y los que han cambiado, de blanco o viceversa. Este proceso corresponde a la etapa de segmentación y está basado en técnicas de detección de regiones que hacen uso de la binarización basada en umbrales (Vélez et al., 2003) y funciones estadísticas de densidad para establecer el modelo de fondo. El resultado (manchas) de este proceso determina cuáles son los objetos de interés en la imagen.

Análisis de manchas *(blob analysis).* El análisis de las manchas primero requiere la separación del objeto del fondo. Usando una imagen digital binarizada (solo píxeles totalmente blancos o totalmente negros), se agrupan los píxeles del objeto para formar un patrón. La geometría de este patrón se utiliza, entonces, para identificar el objeto, localizarlo y examinarlo. Este es un método simple, rápido y capaz de manejar cambios de rotación y tamaño. Este proceso corresponde a la etapa de descripción, ya que permite determinar cuáles de los objetos de la escena corresponden al objetivo de estudio y cuáles pueden ser descartados (González & Woods, 2001).

Seguimiento a las manchas. Los objetos físicos generalmente presentan un movimiento suave a lo largo del tiempo, como consecuencia de la inercia. El seguimiento visual de objetos se basa fundamentalmente en esta consideración. Un factor clave para conseguir que un computador pueda seguir la posición de los objetos en movimiento, es la posibilidad de anticipar este movimiento a partir de su caracterización previa. Esta información a priori se usa para que los algoritmos atiendan al objeto de interés (por ejemplo, una cabeza en una aplicación de videoconferencia, un intruso en un sistema de vigilancia) y no se distraigan con otros objetos presentes en la imagen. Junto a un modelo dinámico que describa la evolución del movimiento a lo largo del tiempo, es fundamental disponer de un conjunto de observaciones sobre el desplazamiento sufrido por los elementos de la imagen en varios cuadros, para plantear así el seguimiento. Los parámetros del algoritmo pueden estar referidos a su localización real en la escena (seguimiento 3D) o bien al movimiento proyectado sobre el espacio de imagen (seguimiento 2D) (Pajares & De la Cruz, 2008).

Regiones y puntos de medición. Gracias al seguimiento de manchas es posible conocer la posición de cada objeto de interés dentro de la escena y el flujo que lleva desde su ingreso. Si se usa esta información y se comparan esos datos con los de cada objeto que ha ingresado en la escena, se pueden obtener y demarcar áreas de especial interés, por ejemplo, áreas de mayor afluencia o áreas de tránsito entre un área y otra.

Para determinar esas áreas, es importante definir puntos de medición que permitan conocer los puntos de entrada y de salida de los objetos de la escena; para mayor utilidad se pueden definir puntos de entrada y salida de cada área definida (Björgvinsson, 2006).

G. Aplicación

Las aplicaciones de visión artificial se clasifican, generalmente, en cuatro grandes áreas: obtención de la distancia de los objetos en la

escena y estructura tridimensional; detección de objetos en movimiento; reconocimiento de patrones y formas; y reconocimiento de objetos tridimensionales. El sistema de este proyecto se puede enmarcar en las tres últimas, pues la solución del problema requiere detectar los objetos, analizarlos y determinar si se trata de personas u otro tipo de objetos.

Es posible detectar la presencia de objetos en movimiento mediante la comparación de una secuencia de imágenes obtenidas de la escena tridimensional en diferentes instantes de tiempo, para lo que se pueden utilizar diferentes técnicas de estimación del flujo óptico, tales como: el método local (Lucas-Kanade) y el método global (Gauss-Seidel). El tratamiento de secuencias también se utiliza para la detección de cambios en diferentes ambientes.

El reconocimiento de formas y patrones se centra en caracterizar los objetos por la forma determinada de su contorno, la cual detalla la región que delimitan y sus propiedades subyacentes.

Las técnicas de obtención de las formas, a partir de sombras (shape form shading), también se encaminan a la obtención de la estructura tridimensional mediante el conocimiento de la iluminación de la escena y la determinación del punto de observación de la misma.

3. Metodología

Los métodos usados para dar cumplimento a cada uno de los objetivos específicos planteados son:

• Implementar un sistema de conteo de personas de bajo costo, utilizando los mínimos recursos de *hardware*.

Se analizaron las características de diferentes dispositivos de captura de imágenes y video, entre los que se tuvo en cuenta: cámaras IP, normalmente utilizadas en circuitos de vigilancia; cámaras web, utilizadas comúnmente en ordenadores portátiles y de escritorio; y cámaras análogas de circuitos cerrados de televisión, las más usadas en las aplicaciones actuales de reconocimiento de personas.

Entre las características que se analizaron están el costo monetario de los equipos, el aporte que induce el dispositivo al desempeño de la aplicación y la calidad de las imágenes proporcionadas por el aparato de captura.

• Establecer cuál es la posición más apropiada de la cámara en la habitación para obtener los mejores datos de la escena. Para esto se deben realizar tomas con la cámara ubicada en diferentes posiciones, a fin de determinar cuáles permiten establecer una geometría estándar aplicable a la mayoría de personas.

• Diferenciar el fondo de la escena de los objetos que se encuentran en movimiento. Para esto se evaluaron diferentes técnicas de segmentación de imágenes, haciendo especial énfasis en la sustracción del fondo y técnicas de binarización basadas en umbrales. También se tuvo en cuenta los antecedentes para determinar cuáles de ellas son más usadas y favorecen a su vez el proceso de detección de movimiento.

• Diferenciar las formas para determinar cuál de ellas corresponde a una persona. Para cumplirlo se utilizaron los resultados obtenidos en los numerales 2 y 3; adicionalmente se debe escoger un algoritmo de análisis de manchas que sea adecuado para dicha situación.

• Realizar seguimiento de cada una de las personas y establecer cuál es el flujo que toman. Para esto se utilizaron los resultados obtenidos en los numerales 2 y 3 de esta sección, y se estudiaron diferentes técnicas de seguimiento de manchas que permitieran hacer el rastreo de los objetos identificados como personas dentro de la escena.

• Mostrar la utilidad que puede tener el sistema de conteo en diferentes campos de acción, y probarlo en diferentes espacios y condiciones.

Una vez implementado un prototipo del sistema, se realizaron pruebas haciendo simulaciones con videograbaciones para mostrar su funcionamiento en diferentes condiciones. Adicionalmente, se propusieron diferentes áreas para su aplicación, explicando el propósito que el sistema cumple en cada una de ellas (por ejemplo, mercadeo, distribución de espacios, manejo de tráfico).

4. Desarrollo y resultados

A continuación se plasman algunos problemas y resultados obtenidos en el desarrollo.

H. Selección de los dispositivos de captura de imagen

Los dispositivos objeto de comparación fueron: cámaras IP, cámaras web y cámaras análogas CCTV. Después de una comparación exhaustiva se decidió usar cámaras IP o cámaras web en lugar de cámaras análogas CCTV. Aunque hay varios factores importantes para esta decisión, el de mayor peso es el factor costo. Las cámaras análogas son más costosas que las digitales. Las cámaras digitales tienen ventajas en su instalación, debido a su facilidad de conexión a través de una red cableada Ethernet, una red inalámbrica o a través de puertos USB. Otra ventaja es que las cámaras IP y web entregan la información directamente en forma digital; se evita el inconveniente de tener que usar una tarjeta PCI especial o implementar un paso más en el proceso de captura para convertir la señal análoga en digital.

Por otra parte, el uso de las cámaras digitales favorece la escalabilidad del sistema, dado que permite tener el control de varias cámaras en diferentes áreas desde una sola unidad central de consolidación de datos (Björgvinsson, 2006), lo cual es útil si se desea realizar una implementación donde se necesite tomar información de diferentes lugares al mismo tiempo.

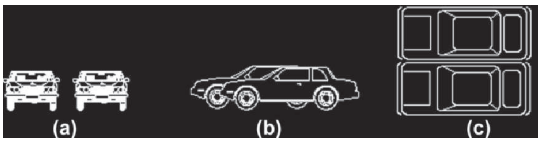
I. Selección de la posición de la cámara

La selección de la posición en que se debe ubicar la cámara no es una tarea sencilla. De ella y del tipo de objeto que se quiera identificar depende la complejidad de la selección o descarte de los objetos de interés (Collins, Lipton & Kanade, 1999). Por este motivo se deben analizar factores como: cuál de las posiciones en las que se puede ubicar la cámara facilita el análisis de la imagen; y cuál de ellas ofrece la menor variabilidad en el tamaño del objeto observado.

En cuanto a la posición, es necesario considerar que las escenas 3D pueden ofrecer mayor o menor cantidad de información dependiendo del ángulo en que se les mire, como se puede ver en la Figura 4, que muestra dos vehículos desde diferentes ángulos.

En esta imagen todas las vistas proporcionan la misma información para el ojo humano. Sin embargo, para una computadora los ángulos sugeridos en las secciones (a) y (c) brindan una mejor manera de procesar la información, pues permiten diferenciar con facilidad cada uno de los objetos por separado, mientras que la imagen en la sección (b) hace necesaria la implementación de un tipo de reconocimiento que pueda intuir la presencia de un objeto a partir de partes de su estructura, lo que implicaría un mayor consumo de recursos y el uso de técnicas de inteligencia artificial mucho más avanzadas.

Figura 4. Información obtenida de una escena desde diferentes

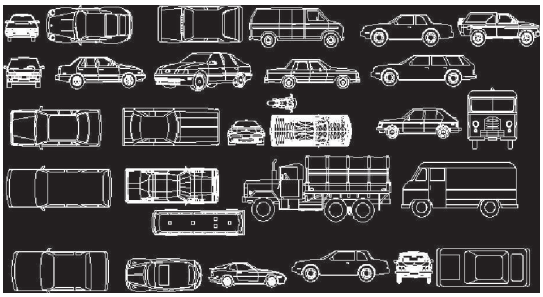


Fuente: elaboración propia.

Es preciso validar si la posición seleccionada introduce algún tipo de deformación a los objetos de interés de la escena. Los objetos que se quiere analizar dentro de la imagen generan diferentes siluetas, dependiendo de la posición de la cámara. La Figura 5 permite comparar nuevamente la forma en que se ve un auto desde diferentes vistas.

El objetivo es el de estandarizar el área que ocupan los objetos de interés dentro de la imagen y facilitar su reconocimiento

Figura 5. Ejemplo de variación del tamaño de los objetos dependiendo del ángulo en el que se observen



Fuente: elaboración propia.

El análisis descrito requiere un conocimiento profundo de la morfología de los objetos de interés, que permita realizar abstracciones que faciliten su identificación. En este proyecto los objetos de interés son las personas que, además de tener características fisiológicas desiguales, pueden adoptar múltiples posiciones como, sentarse, acostarse, estar de pie o incluso inclinarse. Debido a esta gran variabilidad, se debe restringir el dominio de objetos a identificar. Como el objetivo es el conteo de personas que transitan a través de un espacio

cerrado, esta condición disminuye el número de objetos a identificar, pues las personas se desplazan generalmente de pie. Teniendo esto en cuenta, se procede a evaluar los aspectos necesarios para seleccionar la posición de la cámara, tal como se hizo en el ejemplo de los automóviles. En el caso de las personas se decidió que la mejor forma de ubicar la cámara sería una posición cenital, toda vez que desde esta perspectiva las siluetas de las personas pueden abstraerse a una forma de rectángulo (ver Figura 6).

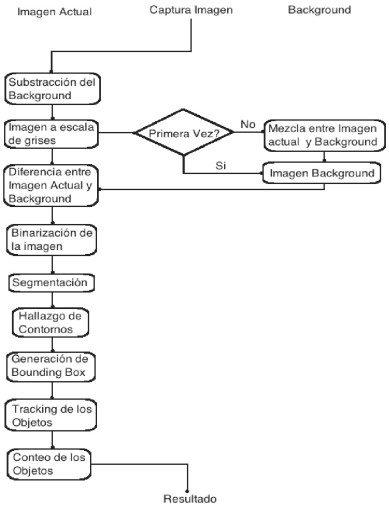
En esta posición, además, la variabilidad de la forma de los individuos es menor que si se observaran de frente, ya que las personas son más diversas en estatura (niños, adultos, ancianos, personas en sillas de ruedas, etc.) que en el ancho de su espalda o la longitud que existe entre la parte frontal y trasera del cuerpo. Para poder realizar un reconocimiento preciso es necesario establecer las medidas estándar de las personas en esta vista, tema que se desarrolla en la sección C.3.

Figura 6. Abstracción de las personas en la imagen

Fuente: elaboración propia.

J. Modelo del sistema propuesto

Figura 7. Diagrama de actividades del sistema propuesto



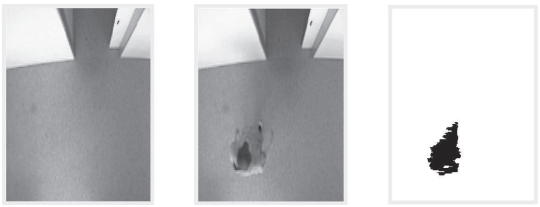
Fuente: elaboración propia.

3. Sustracción del fondo (background substraction)

Al retirar el fondo, la imagen resultante mostrara únicamente el primer plano en una representación binaria (Collins et al., 1999). La Figura 8 muestra la imagen de salida cuando el fondo (primer cuadro) se ha restado del segundo cuadro. En el tercer cuadro se muestran los objetos en primer plano, que a menudo se conocen como manchas.

Algunas de las técnicas disponibles para este procesamiento son: funciones Gaussianas, Histogramas, Kernel Estimation Density (KDE), etc. Igualmente existen diferentes algoritmos de substracción de fondo (Björgvinsson, 2006).

Figura 8. Sustracción del fondo, imagen a procesar



Fuente: (Björgvinsson, 2006).

Sin importar la técnica usada, lo primordial es obtener las manchas del primer plano, con una consistencia que permita identificar fácilmente el tamaño y zona ocupada.

Es necesario, además, que el método de substracción sea muy rápido, pues de no serlo, el resultado del proceso será una imagen totalmente en blanco o totalmente en negro, dependiendo de la lógica usada en la representación de los objetos del primer plano (Viola & Jones, 2001). Por tal razón, en la aplicación creada, se agregó en este proceso una variable más, que consiste en una tasa de refresco modificable, de forma que el proceso de actualización del fondo de la imagen sea más rápido o más lento.

4. Análisis de manchas

El análisis de manchas implica el examen de la imagen binarizada para diferenciar en ella todos los objetos en primer plano individual (manchas). En algunos casos esta tarea puede ser bastante difícil, incluso para el ojo humano.

La Figura 9 muestra dos manchas fácilmente identificables. Cuando las personas que se desplazan visten colores similares a las del fondo, las

manchas resultantes después de eliminar el fondo pueden quedar distorsionadas o con agujeros en el centro, de forma similar a lo que se puede ver en la Figura 10.

Figura 9. Manchas fácilmente diferenciables

Fuente: (Björgvinsson, 2006).

Figura 10. Manchas distorsionados

Fuente: (Björgvinsson, 2006).

Estas distorsiones pueden ocasionar que el sistema interprete una mancha de mayor tamaño, como un conjunto de manchas más pequeñas (ver Figura 11).

Figura 11. Mancha fragmentado

Fuente: (Björgvinsson, 2006).

En algunos casos, los bordes de la mancha se convierten en partes poco claras, y ciertas partes a menudo aparecen como pequeños rasguños, considerados regularmente como ruido. Una mancha con bordes distorsionados se muestra en la Figura 12.

Figura 12. Manchas afectadas por el ruido

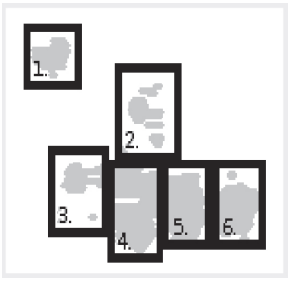
Fuente: (Björgvinsson, 2006).

La gente que permanece en grupos es difícil de manejar, incluso para el ojo humano. Cuando están juntos forman una gran mancha (ver Figura 13). Es difícil determinar con precisión cuántas manchas existen realmente en la imagen. La Figura 14 muestra cómo los objetos en primer plano se pueden dividir en seis manchas.

Figura 13. Mancha generada por la presencia de varios objetos juntos

Fuente: (Björgvinsson, 2006).

Figura 14. Subdivisión de manchas muy grandes



Fuente: (Björgvinsson, 2006).

Puede parecer sencillo dividir las manchas grandes en piezas más pequeñas, aprovechando propiedades como la altura y el ancho, pero este proceso requiere de varios cálculos en cuanto a áreas, bordes y proximidad entre las mismas. Además, existe la posibilidad de que no todas las manchas grandes sean personas; por ejemplo en un supermercado, los clientes pueden empujar carros de compras delante de ellos. Estos carros pueden ser del mismo tamaño de un ser humano, como se ve en la Figura 15.

Figura 15. Mancha de una persona junto con un carrito

Fuente: (Björgvinsson, 2006).

Parte de los problemas descritos son resueltos por el proceso de sustracción del fondo y a través de la aplicación de filtros que mejoran la calidad de las imágenes.

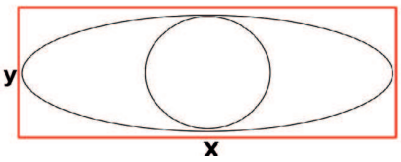
5. Determinación del espacio real ocupado por un objeto

Al observar las personas desde una perspectiva cenital, los datos obtenidos corresponden al ancho y alto del rectángulo envolvente (*bounding box*) que representa una persona en la imagen. Para determinar las dimensiones de ese rectángulo es necesario establecer la medida promedio de las personas, como se muestra en la Figura 19.

Estas medidas X y Y corresponden al ancho de la espalda de una persona (X) y a la distancia que existe entre el pecho y la espalda (Y) (ver Figura 20). Estas medidas varían mucho de una persona a otra, lo que hace difícil la identificación de los objetos como personas, pues habría que tener registradas todas la posibles combinaciones de X y Y en el mundo. En lugar de eso, se estableció un rango de medidas promedio con base en estudios antropométricos.

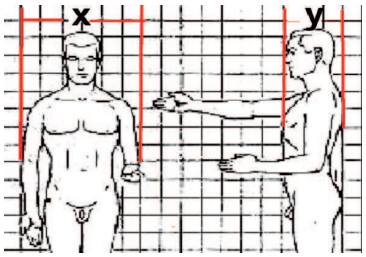
Los resultados de estos estudios se muestran en la Tabla 1. Cabe aclarar que las magnitudes expuestas en esta tabla corresponden a lo encontrado en literatura y estudios previos, y no a un análisis profundo de las medidas del cuerpo humano. Además, corresponden a investigaciones realizadas en Colombia, por lo que estas medidas pueden variar de un país a otro.

Figura 16. Dimensiones del rectángulo envolvente que representan una persona en el sistemapuesto



Fuente: elaboración propia.

Figura 17. Medidas antropométricas equivalentes a las dimensiones del rectángulo envolvente



Fuente: elaboración propia.

Tabla 1. Promedio de las medidas de X y Y en hombres y mujeres de Colombia

Medida	X(mm)	Y(mm)	SD X(mm)	SD Y(mm)
Hombres	480	250	50	100
Mujeres	400	300	40	60

Fuente: elaboración propia.

Determinar este tipo de medidas corresponde al área de la antropometría y requiere de un trabajo estadístico arduo que está por fuera del alcance de este documento. Sin embargo, es bueno recordar que la parametrización del *software* con datos de mayor veracidad, llevará a una mayor precisión en la identificación de los objetos de interés.

6. Determinación del tamaño de un objeto en la imagen

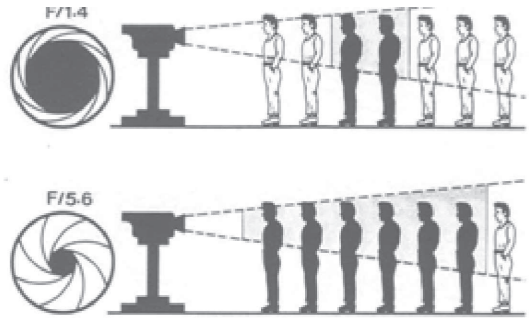
En una imagen digital el ancho (X) y alto (Y) de una persona se mide en píxeles. Esto implica determinar un factor de conversión. No obstante, para poder establecer el tamaño de un objeto en píxeles a partir de una imagen digital, se debe tener en cuenta que no todas las cámaras tienen las mismas especificaciones, en resolución, tamaño del lente, ángulo de apertura, etc., y, dependiendo de la configuración del dispositivo, la foto resultante presentará variaciones. Por lo tanto, es necesario conocer algunos conceptos de fotografía:

Profundidad de campo. Se denomina así a todo el espacio en una fotografía que se encuentra enfocado o, al menos, a toda la zona de la fotografía en la cual el foco es suficientemente bueno como para distinguir los objetos de una manera nítida. Dependiendo de la profundidad de campo, existen distintos planos de enfoque. Hay una zona donde se encuentra el mayor grado de enfoque o *focus*. Por delante y por detrás de este foco, aparecen diferentes planos de enfoque que se encuentran más o menos enfocados.

En el caso del sistema de conteo de personas, se necesita que la imagen tenga una profundidad de campo que abarque todo el espacio o escena. Para lograr esto, el diafragma de la cámara se debe configurar con un número F (grado en que se mide la apertura del diafragma de la cámara) alto, es decir, con una apertura pequeña. Por lo

general, en las cámaras *web* la apertura del diafragma no se puede cambiar y, por defecto, es una apertura pequeña que permite enfocar toda la escena. Por esto, para las pruebas realizadas en este proyecto no representa mayor problema.

Figura 18. Relación entre la apertura del diafragma y la profundidad de campo



Fuente: http://2.bp.blogspot.com/-6A1LlUOJ338/ULWuk15XR8I/AAAAAAAAAI0/-T9WnXEOAYY/s1600/profun+campo.jpg.

Perspectiva. Es básicamente la ilusión visual que, percibida por el observador, ayuda a determinar la profundidad y situación de objetos a distintas distancias. Gracias a ella se puede distinguir cuáles objetos están más cerca y cuáles más lejos. En fotografía y video hay que tener en cuenta este concepto, ya que, en un plano horizontal, y dependiendo de la distancia del objeto al lente, su tamaño va a variar en la imagen, haciéndose más pequeño a mayor distancia.

Esta situación dificulta significativamente establecer medidas estándar para una persona y aumenta los cálculos a realizar. Este problema desaparece al utilizar la cámara en posición cenital, apuntando hacia el suelo en un plano vertical, donde los objetivos están sobre un límite: el piso. De esta manera, las personas permanecen a una misma distancia de la cámara.

Figura 19. Ejemplo de perspectiva (a) y (b)

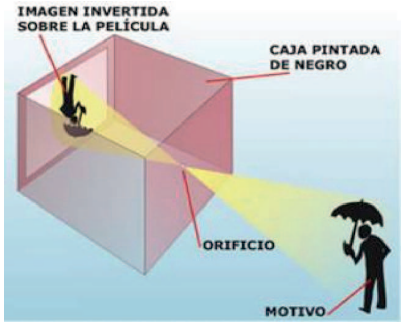
Fuente: http://www.funny22.com/wp-content/uploads/2012/06/family-illusion-4.jpg.

Fuente: http://static0.blip.pl/user_generated/update_pictures/2161107.jpg.

7. Cálculo del tamaño de un objeto en píxeles

Cuando se toma una imagen con una cámara, se utiliza el principio de la cámara estenopeica. Esta es, básicamente, una caja con un agujero de unos 0.5mm, por donde entra la luz, y una película fotográfica en el lado opuesto al agujero. Entonces, si se enfoca un objeto delante de la cara que tiene el orificio, por ejemplo un árbol, dentro de la caja, en la cara contraria se formará una imagen invertida del árbol, ya que la luz se propaga en línea recta, lo que hace que los rayos incidentes de la copa del árbol se proyecten en la parte inferior de la caja.

Figura 20. Principio de la cámara estenopeica



Fuente: http://blog.educastur.es/luciaag/files/2007/10/estenopeica.jpg.

En realidad, las cámaras actuales ya no funcionan con el mismo mecanismo, sino que reemplazan el agujero con el uso de lentes convergentes, que hacen que los rayos de luz se concentren y converjan proyectando la imagen a un sensor. El mecanismo es diferente, pero el principio se conserva.

Utilizando este principio se obtiene la siguiente ecuación (E1):

$$\frac{\text{Altura Objeto}}{\text{distancia Objeto Lente}} = \frac{\text{Altura Imagen}}{\text{distancia Focal}} \quad \text{(E1)}$$

Donde:

Altura Objeto. Altura en (mm) del objeto.

Distancia Objeto Lente. Distancia del Objeto enfocado al lente de la cámara.

Altura Imagen. Altura de la imagen en el sensor.

Distancia Focal. Es la distancia entre el centro óptico de la lente o plano nodal posterior y el foco.

Por otro lado tenemos que:

$$\frac{\text{Altura Objeto}}{\text{distancia Objeto Lente}} = \frac{\text{Altura Imagen}}{\text{distancia Focal}} \quad \text{(E1)}$$

Donde:

#píxeles Imagen. Es el número de píxeles del objeto en la fotografía.

Tamaño del píxel del sensor. Es el tamaño del píxel definido en el sensor.

La altura del objeto y la distancia del objeto al lente se pueden medir fácilmente y la distancia focal se obtiene de las especificaciones de la cámara. La altura de la imagen es el único dato indeterminado, dado que no se conoce el número de píxeles que ocupa en la imagen (dato a hallar), pero sí se puede conocer el tamaño del píxel en el sensor, el cual también debe aparecer en las especificaciones de la cámara.

Por lo cual, a partir de (E1) y (E2) se obtiene

$$\text{\#pixeles Imagen} = \frac{\text{Altura Objeto} * \text{distancia Focal}}{\text{distancia Objeto} * \text{tamaño Píxel Sensor}}$$

(E3)

Aplicando esta ecuación para el alto y el ancho se puede hallar la medida de un objeto o una persona en una imagen.

En caso de no contar con todos los factores para realizar los cálculos de la anterior ecuación, lo más fácil es medir en la escena dos objetos de los cuales se conoce su tamaño, a una misma distancia y a través de comparación y una regla de tres deducir el dato.

Debido a la variedad de dispositivos disponibles, lo ideal en el *software* es incluir controles que permitan calibrar estos datos en tiempo real, de forma que este se adapte mejor a las condiciones del entorno.

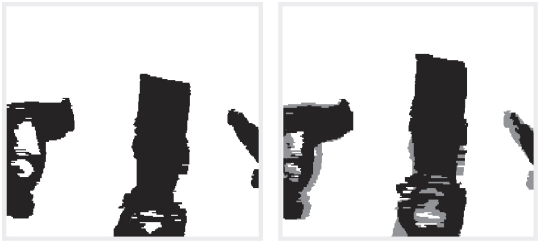
8. Seguimiento de manchas

El módulo de seguimiento de manchas permite mantener identificados a los objetos, con el objetivo de no contar como nuevos aquellos dentro de la escena que ya han sido detectados y que solo se han desplazado en la imagen.

El seguimiento de manchas etiqueta cada uno de ellos con un identificador específico, que se adjuntará a la mancha similar o igual en el cuadro siguiente (Doshi, 2005). ¿Qué define una mancha como algo similar en dos imágenes? Depende del algoritmo empleado. En el caso que se presenta, se tienen en cuenta dos factores, el tamaño del área ocupada, y el centroide de la figura.

La primera comparación que se hace es la del fondo substraído versus la imagen en movimiento; a partir de allí se identifican las primeras manchas. Luego, *frame* por *frame*, se comparan las imágenes y de acuerdo con la proximidad de los centroides de las manchas, es posible mantener el rastro de cada uno. La Figura 21 muestra tres objetos detectados. Estos se mueven en direcciones diferentes. El cuadro siguiente muestra cómo las manchas han modificado ligeramente sus posiciones. Las manchas desvanecidas representan el cuadro anterior, mientras que las manchas negras muestran la nueva posición del objeto.

Figura 21. Seguimiento de manchas, primera imagen izquierda y segunda imagen derecha



Fuente: (Björgvinsson, 2006).

K. Implementación técnica

9. Producto final

Tecnología usada. El proyecto se desarrolló en el lenguaje de programación c++, debido a su nivel medio-bajo, que permite un mayor control sobre el uso de la memoria.

Proceso de ejecución del programa. El programa consta básicamente de tres procesos que

realizan toda la operación del sistema: inicialización de las variables, actualización de la escena y tratamiento de las imágenes (implementación del sistema de visón artificial planteado).

A continuación se indican paso a paso las actividades que realiza el programa cuando se ejecuta:

Inicialización de las variables. En esta etapa se parametriza el sistema, dándole valor inicial a aquellas variables que el programa utiliza para realizar su tarea: tamaño de las personas en la escena, resolución de la cámara, establecimiento del fondo de la escena, etc. Segundo, se define la tasa de refresco del fondo de la imagen, debido a que la imagen y los elementos del fondo pueden ir cambiando con el tiempo.

La pantalla principal muestra la visualización de la cámara y el video convertido a escala de grises. En esta pantalla están las opciones de configuración de los filtros y aplicación de transformaciones.

Tratamiento de las imágenes. Primero se obtiene el fondo de la escena, que es la primera imagen que toma la cámara. Luego, se obtienen los píxeles de las siguientes imágenes, para convertirlos en un objeto de tipo imagen, el cual puede variar dependiendo de la librería de programación que se utilice.

Después de obtener la instancia de la imagen, esta se convierte a escala de grises (es más fácil trabajar un solo canal en vez de tres, como lo es RGB). Luego de obtener la siguiente imagen, se realiza la diferencia entre el fondo y la imagen que se está capturando en el momento. Posteriormente, se aplica una serie de filtros que segmentan la imagen destacando los objetos de interés. Algunos de los filtros aplicados son:

Filtro de erosión *(Erosion filter)*. Corresponde a un filtro morfológico que cambia la forma de los objetos en una imagen, por la erosión (reducción) de los límites de los objetos brillantes y la ampliación de los límites de los oscuros. Se utiliza para reducir, o eliminar, los pequeños objetos brillantes.

Suavizado *(Smooth)*. Filtros de suavizado, también llamados filtros de paso bajo, ya que conservan componentes de baja frecuencia y reducen los componentes de alta frecuencia.

Filtro de paso alto *(HighPass)*. Un filtro de paso alto o HPF es un filtro que ignora las frecuencias altas, pero atenúa (es decir, reduce la amplitud de) las frecuencias más bajas. La cantidad real de atenuación para cada frecuencia es un parámetro de diseño del filtro.

Blurhigh Pass o Gaussian Blur. También conocido como suavizado de Gauss, es el resultado de difuminar la imagen por una función gaussiana. Es un efecto ampliamente utilizado en el software de gráficos, por lo general, para reducir el ruido de la imagen y reducir el detalle. El efecto visual de esta técnica es un desenfoque suave parecido al que resulta de ver la imagen a través de una pantalla translúcida.

Umbral *(Threshold)*. Durante el proceso de umbral, los píxeles en una imagen se marcan como objeto si su valor es mayor que cierto valor umbral (suponiendo que un objeto sea más brillante que el fondo) o como fondo en caso contrario (su valor de intensidad es menor que el umbral). Por lo general, a un píxel objeto se le asigna un valor de "1", mientras que un píxel del fondo tiene un valor de "0". De esta manera, se obtienen una imagen binaria compuesta por cada píxel de color blanco o negro.

L. Pruebas

La Tabla 2 muestra los resultados de las pruebas realizadas con el programa.

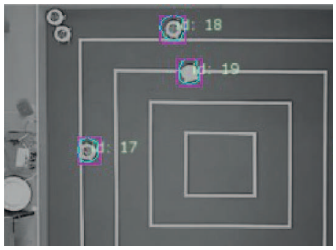
Tabla 2. Configuración de pruebas

Prueba / Configuración	1. Video de robots	2. Video malabaristas 1	2. Video malabaristas 2
Área Mínima	55	638	110
Área Máxima	444	2980	2700
Ancho Mínimo	2	2	16
Ancho Máximo	25	83	200
Altura Mínima	2	2	11
Altura Máxima	25	83	202
Velocidad de aprendizaje	30	0*	338
Threshold	11	53	138
Smooth	1	8	14
Amplify	10	151	160
High Blur	40	27	67
Pass Noise	6	14	14

*Se utiliza una imagen estática en el fondo.

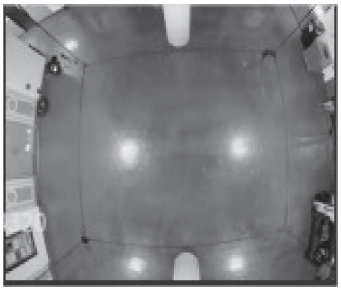
Fuente: elaboración propia.

Figura 22. Prueba 1. Objetos robots



Fuente: elaboración propia.

Figura 23. Prueba 2. Fondo

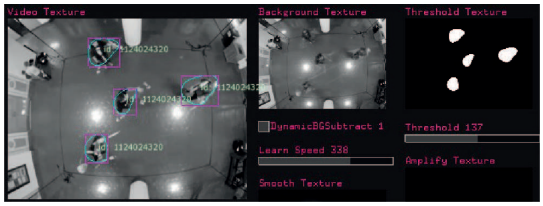


Fuente: elaboración propia.

Figura 24. Prueba 2. Reconocimiento

Fuente: elaboración propia.

Figura 25. Prueba 3. Reconocimiento (b)



Fuente: elaboración propia.

Tabla 3. Resultados

	Prueba 1. Video de robots	Prueba 2. Video malabaristas 1	Prueba 3. Video malabaristas 2
Observaciones por parte del equipo de desarrollo	Con esta configuración se captura fácilmente los objetos individuales cuando están separados. Los objetos son muy uniformes, por lo cual es fácil decidir el rango de tamaños en el programa.	Los movimientos de los bolos y los brazos de los malabaristas modifican la forma de la persona constantemente, lo que dificulta establecer los rangos de tamaño de las mismas. Para la configuración del fondo, fue más adecuado establecer uno estático, ya que todos los objetos en la escena se mueven y este siempre es el mismo.	A diferencia del video anterior, en este son menos personas, por lo que hay más distancia entre ellos y existe menor cantidad de bolos que hagan interferencia.
Problemas	Cuando se juntan los objetos se forma una mancha más grande de lo normal, que se sale de los rangos definidos, por lo cual se pierde la referencia de estos, y cuando se separan de nuevo se vuelven a contar. Debido a que se utiliza el aprendizaje dinámico del fondo, los objetos que dejan de moverse dejan de contarse.	Los bolos que lanzan los malabaristas también están en movimiento, lo cual genera mediciones que deben ser ignoradas validando su área. Al igual que en la prueba anterior, cuando un objeto se junta con otro, provoca la pérdida en el seguimiento de las manchas y hace que se cuenten nuevamente en el total.	Uno de los personajes tiene un color de ropa parecido al del fondo de la imagen y cuando se queda quieto por un rato tiende a desaparecer.
Tiempo de conteo	20 segundos	15 segundos	20 segundos
Número	Robots moviéndose: 4 Robots contados: 10	Malabaristas: varía entre 6 y 8 Malabaristas contados: 36	Malabaristas: varía entre 4 y 6 Malabaristas contados: 15

Fuente: elaboración propia.

Se observa que cuando dos individuos se juntan, las manchas se traslapan y el rastro se pierde. Es necesario dividir esa forma y mantener el rastro. Para ello y de acuerdo con los resultados, se observa que en la unión de manchas se genera una forma con ángulos cóncavos, los cuales pueden usarse para determinar el punto de separación. Esto será tenido en cuenta para próximas versiones del prototipo.

Un problema que queda por resolver son las variaciones de luz dentro de la escena, pues es un elemento del ambiente que muchas veces no puede ser controlado y que cambia de manera drástica la forma en que debe ser parametrizado el sistema.

Conclusiones

En definitiva, es posible crear un sistema de visión artificial y de conteo de objetos, de forma básica, con el uso de los elementos imprescindibles descritos en el marco teórico: un computador, un sensor y un algoritmo de procesamiento, como el descrito en este documento.

El sistema de visión artificial para el conteo de objetos, podría parametrizarse, de diferentes maneras, para que en la etapa de descripción se reconozca un objeto en especial. En este documento, las pruebas se realizaron con personas, pero como se describió con el ejemplo de los automóviles, las posibilidades son muy variadas en cuanto a los campos de acción y los diferentes ambientes en que se use.

Por último, se puede concluir que, habiendo identificado correctamente los parámetros y la información a obtener de la imagen, se pueden lograr resultados de conteo muy certeros. Igualmente, encontramos que el proceso para el sistema propuesto puede llegar a tener mejoras sustanciales y resultados más precisos, si se incluyen o se modifican las técnicas y dispositivos aquí mencionados.

Referencias

Björgvinsson, Tryggvi. (2006). *Peocounter: People Counting Software*. Gothenburg: Chalmers University of Technology.

Cantoni, Virginio, Levialdi, Stefano & Vito, Roberto. (1996). *Artificial Vision: Image Description, Recognition, and Communication (Signal Processing and its Applications)*. Academic Press.

Collins, R., Lipton, A. & Kanade, T. (1999). *A System for Video Surveillance and Monitoring*. American Nuclear Soc. 8th Int. Topical Meeting on Robotics and Remote Systems.

Conrad, Gary & Johnsonbaugh, Richard. (1994). *A real-time people counter*. SAC '94: Proceedings of the 1994 ACM symposium on Applied computing. New York.

Doshi, Anup. (2005). *CSE 252C: People Counting and Tracking for Surveillance*.

González, Rafael C. & Woods, Richard E. (2001). *Digital Image Processing*. Prentice Hall.

Haritaoglu, I., Harwood, D. & Davis, L. S. (1998). *W4S: A Real-Time System for Detecting and Tracking People in 2 1/2 D. ECCV*.

Jain, Anil K. (1989). *Fundamentals of Digital Image Processing*. Prentice Hall.

Kim, Jae-Won, Choi, Kang-Sun, Choi, Byeong-Doo & Ko, Sung-Jea. (2002). *Real-time Vision-based People Counting System for the Security Door*. Seoul: Korea University, Anam-dong, Sungbuk-ku.

Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M. & Shafer, S. (2000). *Multi-camera multi-person tracking for easyliving*. VS'00: Proceedings of the Third IEEE International Workshop on Visual Surveillance (VS'2000). Washington, D. C.

Lipton, A., Fujiyoshi, H. & Patil, R. (1998). *Moving target classification and tracking from real-time video*. Proc. of the Workshop on Application of Computer Vision. IEEE.

Pajares Martinsans, Gonzalo & De la Cruz García, Jesús M. (2008). *Visión por computador: Imágenes digitales y aplicaciones*. 2 ed. Madrid: RA-Ma – Alfaomega.

Siebel, N. & Maybank, S. (2002). *Fusion of Multiple Tracking Algorithms for Robust People Tracking, ECCV 2002*.

Sigvaldason, Einar. (2002). *People Flow Solutions. Fact sheet*.

Sourabh, Daptardar & Makarand, Gawade. (2007). *CS676: People counting*.

Teixeira, Thiago & Savvides, Andreas. (2007). *Lightweight people counting and localizing in indoor spaces using camera sensor nodes. Yale University.*

Viola, Paul & Jones, Michael (2001). *Robust Real-time Object Detection*. International Journal of Computer Vision.

Yang, Danny B., González Baños, Héctor H. & Guibas, Leonidas J. (2003). *Counting People in Crowds with a Real-Time Network of Simple Image Sensors*. Proceedings of Ninth IEEE International Conference on Computer Vision.

Vélez Serrano, José F., Moreno Díaz, Ana B., Sánchez Calle, Ángel & Sánchez Marín, José L. E. (2003). *Visión por computador*. 2 ed. Universidad Rey Juan Carlos.