

# Prototipo de un dispensador automático de bebidas basado en visión artificial para estudio de experiencia de usuario

## Computer vision-based automatic beverage dispenser prototype for user experience studies

Fernando Merchán<sup>1\*</sup>, Elba Valderrama<sup>2</sup> Martín Poveda<sup>3</sup>

<sup>1,3</sup>Facultad de Ingeniería Eléctrica, Universidad Tecnológica de Panamá, <sup>2</sup>Facultad de Ingeniería en Sistemas Computacionales, Universidad Tecnológica de Panamá

<sup>1</sup>fernando.merchan@utp.ac.pa, <sup>2</sup>elba.valderrama@utp.ac.pa <sup>3</sup>martin.poveda@utp.ac.pa

**Resumen**– El presente trabajo presenta diversos aspectos de la implementación de un prototipo de sistema de venta de bebidas automático basado en visión artificial. Este sistema incorpora tecnologías de tipo sin contacto incluyendo reconocimiento facial para la identificación de usuario y reconocimiento de gestos para la selección de la bebida. Este prototipo se presenta como una plataforma de pruebas para explorar la aceptación de estas tecnologías en los usuarios y para compararla con otras tecnologías como las pantallas táctiles. Además de presentar los aspectos técnicos del dispositivo, presentamos observaciones iniciales del estudio de interacción hombre máquina. Estas observaciones de interacción y de tipo técnica nos permiten proponer perspectivas de mejoras para esta plataforma y de las implementaciones comerciales de este tipo de dispositivos.

**Palabras claves**– máquina de ventas, reconocimiento facial, reconocimiento de gesto de manos, interacción hombre máquina.

**Abstract**– This paper presents several aspects of the implementation of a prototype of automatic beverage dispenser with computer vision functionalities. The system presents touchless technologies including face recognition for user identification and hand gesture recognition for beverage selection. This prototype is a test platform to explore the acceptance of these technologies by consumers and to compare it with other technologies such as touch screens. We present both the technical aspects of the device and some observations of human-machine interaction. The perspectives gained may be useful in the future for developing a commercial implementation.

**Keywords**– automatic beverage dispenser, face recognition, hand gesture recognition, human-machine interaction.

**Tipo de Artículo:** Original

**Fecha de Recepción:** 27 de febrero de 2017

**Fecha de Aceptación:** 12 de abril de 2017

### 1. Introducción

Ha habido recientemente mucho interés en el uso de tecnologías de visión por computadora e investigación en aspectos de interacción hombre máquina para crear interfaces fáciles de usar para varias aplicaciones.

Por un lado, los sistemas de reconocimiento de rostros han sido introducidos en muchos contextos incluyendo: la videovigilancia, el control de acceso y la identificación automática de usuarios [1-3]. Muchos enfoques han sido presentados para tratar el reconocimiento de rostros y tareas relacionadas [4-8].

Algunos ejemplos de sistemas de reconocimiento facial en tiempo real son presentados en [9-11]. En [9],

Fei Zuo et al. desarrollan un sistema automático de reconocimiento de usuario en ambiente de casa inteligente interconectada. El mismo alcanza 95% de precisión con una velocidad de 4 fps (fps, fotogramas por segundo). Otros sistemas han sido desarrollados sobre FPGA [12,13]. Por ejemplo, en [12], el sistema fue implementado usando tecnología VHDL, usando una tarjeta Virtex-5 FPGA. El mismo usa el algoritmo de Viola & Jones para la detección de rostros y el método de *eigenfaces* para el reconocimiento de rostros.

El reconocimiento de gestos corporales también ha sido introducido inclusive en los hogares en el contexto de dispositivos de videojuegos tales como el Kinect [14]. También ha habido un gran interés en el desarrollo

de algoritmos de reconocimientos de gestos de manos por visión por computadora en el contexto de la interacción hombre-máquina [15,16].

Los servicios basados en tecnologías de información han atraído la atención de la industria de alimentos. Por ejemplo, las *tablets* son usadas como menú en un restaurante y la ubicación de la persona es utilizada para dar acceso a información de descuento a clientes [17].

Desde hace décadas, las máquinas expendedoras han sido un canal de ventas importantes para la industria de comidas, bebidas y otros productos. Sin embargo, se ha demostrado que los dispensadores automáticos de bebidas y comidas pueden tener bacterias peligrosas [17]. Los investigadores no están seguros cómo las bacterias ingresan en las dispensadoras. Se sospecha que el ingreso es debido a la utilización de las máquinas con las manos contaminadas o en el momento de la limpieza de las mismas con paños húmedos. Por esta razón, surge un interés en incorporar tecnología sin tacto en estos sistemas. Un ejemplo, de un sistema de venta de bebidas sin contacto fue patentado por Barton & Barton [18]. Por ejemplo, se han presentado patentes para la utilización de reconocimiento de gestos para la visualización y selección de productos [19].

Algunos sistemas inteligentes con funcionalidades sin contacto en dispensadores automáticos han sido propuestos. En [20], se presenta una implementación de un sistema de cobro o registro de consumo automático basado en reconocimiento facial en un dispensador de cervezas. La tarea de reconocimiento de rostro es implementada en una Raspberry Pi. El sistema reconoce el rostro de la persona que se sirve la bebida y realiza el cobro en función de la bebida y el volumen servido. Este prototipo es de construcción casera y funciona de manera “*stand alone*”.

Antes de implementar este tipo de sistemas de venta inteligentes es pertinente evaluar la aceptación por parte de los usuarios de estas nuevas tecnologías y detectar las posibles dificultades de utilización.

En este trabajo se presenta el desarrollo de un prototipo de sistema inteligente de venta de bebidas utilizando herramientas de visión computacional. Este prototipo incluye capacidades de:

- a) Reconocimiento facial para la identificación de usuario.
- b) Reconocimiento de gestos de manos para la selección de bebidas.

También incluye el uso de una pantalla táctil con el fin de poder presentar varios escenarios de prueba combinando tecnologías táctiles y sin contacto para la identificación del usuario y la selección de la bebida. El objetivo de este prototipo es servir como plataforma de pruebas de experiencias de interacción de hombre-máquina a fin de explorar la aceptación de los usuarios a comprar comidas y bebidas con la tecnología sin contacto.

La contribución de este trabajo se puede sintetizar en los siguientes puntos:

- Presentar los aspectos de implementación de este sistema en términos de componentes físicos, plataforma de *software* y algoritmos.
- Presentar las conclusiones obtenidas de las pruebas iniciales.

El resto del artículo está organizado como sigue: la sección 2 presenta de forma general el sistema. En la sección 3 presenta los componentes físicos del sistema. La sección 4 presenta las funcionalidades de visión computacional del sistema, respectivamente. Los protocolos de prueba diseñados para el sistema se presentan en la sección 5. La sección 6 presenta las observaciones iniciales. La sección 7 presenta las perspectivas de este trabajo. La sección 8 concluye el artículo.

## 2. Descripción general del sistema

El sistema propuesto consiste en una plataforma que permite realizar pruebas con diferentes funcionalidades de un sistema dispensador de bebidas. Estas funcionalidades incluyen el ingreso de los datos vía teclado, la selección de opciones vía pantalla táctil, identificación de usuario por reconocimiento facial, selección de bebidas por movimientos de la mano. El sistema plantea un conjunto de tres posibles escenarios de prueba que consisten en una combinación de los elementos mencionados.

El sistema está desarrollado sobre el programa Matlab. El mismo es ejecutado en una computadora personal. El sistema se presenta mediante una interfaz gráfica que permite tener acceso a los escenarios de prueba y en cada uno de ellos a las funcionalidades descritas. El sistema registra para cada prueba realizada informaciones tales como: nombre de usuario, datos personales de la cuenta, hora y fecha de la prueba, escenario seleccionado, opciones de compra seleccionada y tiempo de duración de cada una de las

etapas de los escenarios de pruebas. Estos registros están a la disposición de los especialistas que realizan el estudio de experiencia de usuario.

Para el aspecto de dispensado de bebida propiamente dicho, la plataforma cuenta con un sistema electromecánico controlado por micro-controlador Arduino Nano. Este sistema se conecta vía puerto USB a la computadora.

A fin de entender cómo los usuarios se sienten e interactúan con una máquina de venta de bebidas sin contacto y con reconocimiento facial, hemos considerado tres escenarios de interacción. El escenario a utilizar se selecciona en la primera pantalla presentada al usuario.

Cada escenario presentará tres acciones: registro de usuario, identificación de usuario (*log in*) y selección de bebida. En cada escenario se presentan diferentes

combinaciones de tecnologías para la ejecución de las mismas.

A continuación, presentamos una descripción de los escenarios:

**Escenario 1:** el usuario se registra y se identifica usando un nombre de usuario y palabra clave introducidos por teclado. Las bebidas son seleccionadas mediante “botones” presentes en la pantalla táctil.

**Escenario 2:** el usuario se registra y se identifica usando reconocimiento facial. La selección de bebidas se realiza mediante la pantalla táctil.

**Escenario 3:** el usuario se registra y se identifica mediante reconocimiento facial. Las bebidas son seleccionadas usando reconocimiento de gestos de mano.

El diagrama de flujo general del prototipo es presentado en la figura 1.

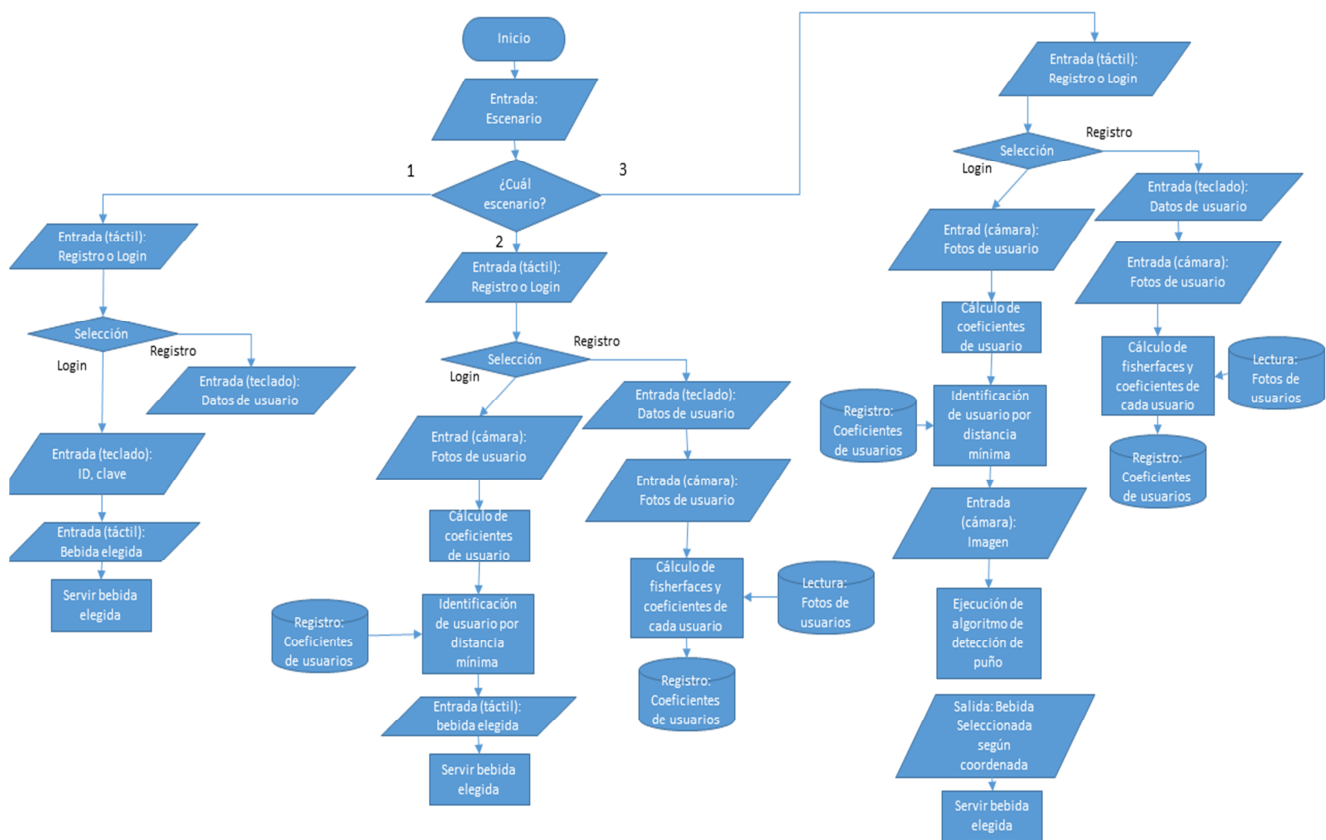


Figura 1. Diagrama de flujo del prototipo.

### 3. Componentes físicos del prototipo

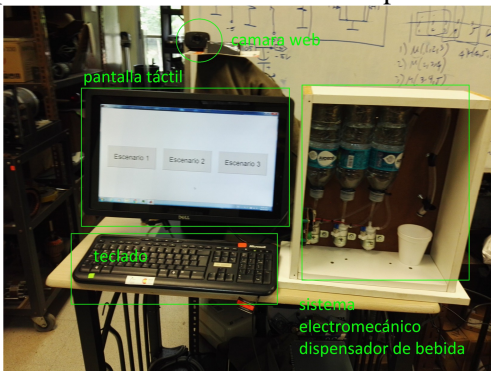
El sistema propuesto está compuesto por una computadora personal, un sistema electromecánico

dispensador de bebida controlado por Arduino Nano, una cámara web, una pantalla táctil y un teclado (ver la figura 2). La computadora personal utilizada tiene un

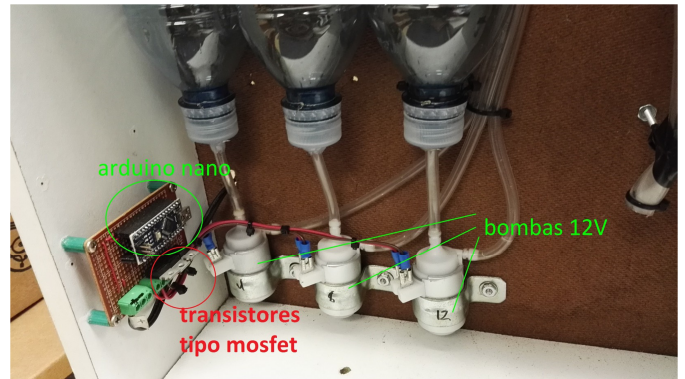
procesador Core i7 de 2.5GHz y 8 GB de memoria RAM.

El sistema dispensador de bebida consiste en un conjunto de 3 bombas eléctricas de 12 voltios alimentadas por un regulador de voltaje con una salida de 12 voltios y 3 amperios. Las mismas bombean el líquido asociado a tres reservorios de tres bebidas diferentes. Las bombas están controladas por una plataforma electrónica basada en micro-controlador Arduino Nano. Tres pines de salida del Arduino activan la alimentación de las bombas mediante circuitos basados en transistores Mosfet tipo N, modelo IRFZ44N. Puede ver detalles de este sistema en la figura 3. El esquemático del circuito accionador se muestra en la figura 4. El microcontrolador Arduino está programado para activar cada una de las bombas en función de la señal recibida del puerto USB conectada a la PC. El tiempo de activación de la bomba ha sido configurado en función del volumen de agua para el llenado de los vasos usado en el prototipo.

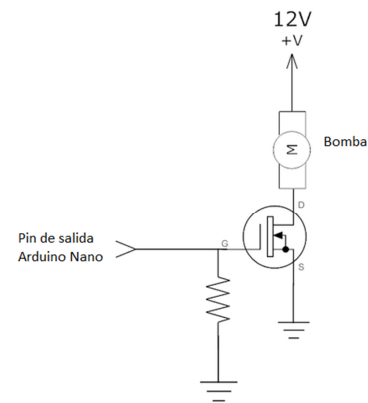
La computadora personal ejecuta la interfaz gráfica desarrollada en Matlab. La cámara web es utilizada para las funciones basadas en visión artificial del equipo. El teclado permite ingresar datos para los escenarios de prueba con contacto. El monitor táctil permite visualizar las diferentes opciones y permite la operación táctil de la plataforma en el escenario correspondiente.



**Figura 2.** Plataforma de pruebas venta de bebidas (la figura presenta el sistema electromecánico dispensador de bebidas, pantalla táctil, teclado y cámara web).



**Figura 3.** Sistema electromecánico dispensador bebidas.



**Figura 4.** Esquemático de circuito accionador de las bombas.

## 4. Implementación de funciones de visión por computadora

Las funciones que distinguen al sistema propuesto son la identificación de usuario por reconocimiento facial y la selección de bebidas mediante el reconocimiento de gestos de la mano. A continuación, presentaremos aspectos de la implementación de estas funciones.

### 4.1 Reconocimiento facial para identificación de usuario

Para el reconocimiento facial se ha implementado un algoritmo basado en el Análisis Lineal Discriminante (LDA, por sus siglas en inglés) que también es conocido en la literatura como el método de *fisherfaces* [21,22]. Este método está a su vez relacionado con el método de *eigenfaces*.

El método de *eigenfaces* está basado en el Análisis de Componentes Principales (PCA, por sus siglas en inglés). El mismo es ampliamente utilizado en el área



del reconocimiento facial por su simplicidad como por los buenos resultados que proporciona en ambientes controlados [23]. La idea principal de este método es aplicar una transformación ortogonal para convertir un conjunto de vectores de entrenamiento posiblemente correlacionados en un conjunto de vectores linealmente no correlacionados. Se encuentran así los mejores vectores que representan a la distribución de las imágenes. Estos vectores definen un subespacio correspondiente a las imágenes de los rostros de entrenamiento.

A continuación, daremos detalles de este método.

Se parte de un conjunto de imágenes de entrenamiento que corresponden a los usuarios del sistema. Estas imágenes, representadas en matrices  $N \times N$ , son organizadas en vectores de dimensiones  $N^2 \times 1$  bajo la notación  $\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_{M-1}, \Gamma_M$ . En este conjunto de imágenes puede haber una o más imágenes por persona. El promedio  $\Psi$  de los vectores es calculado como sigue:

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad (1)$$

Al conjunto de vectores  $\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_{M-1}, \Gamma_M$  se le resta la imagen promedio como sigue:

$$\Phi_i = \Gamma_i - \Psi \quad (2)$$

Estos vectores se agrupan en la matriz  $A$  de tamaño  $N^2 \times M$  como sigue:

$$A = [\Phi_1, \Phi_2, \Phi_3, \dots, \Phi_{M-1}, \Phi_M] \quad (3)$$

La matriz de covarianza  $C$  de la matriz  $A$  se calcula como sigue:

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = AA^T \quad (4)$$

Para el cálculo del PCA es necesario el cálculo de valores y vectores propios de la matriz  $C$ . En este punto notamos que las dimensiones de esta matriz, que es  $N^2 \times N^2$ , es órdenes muy grandes. Esto implicará un alto costo computacional y altos requerimientos de memoria. Se puede demostrar que podemos obtener un conjunto relacionado de vectores propios de la matriz  $M \times M$  definida como sigue

$$L = A^T A \quad (5)$$

En efecto, dado los vectores propios  $v_k (k = 1, 2, 3, \dots, M)$  de la matriz  $L$ , se pueden obtener los primeros  $M$  vectores propios  $u_k (k = 1, 2, 3, \dots, M)$  de la matriz de covarianza  $C$ , mediante una combinación

lineal de los vectores diferencia  $\Phi_i$  y los vectores propios de la matriz  $L$ . Este cálculo se expresa como sigue:

$$U = [u_1, u_2, \dots, u_M] \\ = [\Phi_1, \Phi_2, \dots, \Phi_M][v_1, v_2, \dots, v_M] = AV \quad (6)$$

En la práctica no es necesario utilizar todos los  $M$  vectores para reconstruir los rostros, esto permite reducir el tamaño de los vectores que representan a cada rostro en el nuevo subespacio PCA. Esto se logra eligiendo los vectores propios asociados con los valores propios de mayor valor. Cabe destacar es que si estos vectores se redimensionan al tamaño original de la imagen  $N \times N$  se obtienen los *eigenfaces*, a partir de los cuales, haciendo una combinación ponderada de ellos, se pueden obtener las imágenes aproximadas de todos rostros almacenados en la base de datos.

Una nueva imagen de rostro de entrada  $\Gamma$  es transformada en sus componentes de eigenfaces (es decir, proyectado en el *facespace*) mediante la operación siguiente:

$$\omega_k = u_k^T [\Gamma - \Psi] \quad \text{for } k=1, 2, 3, \dots, M' \quad (7)$$

donde el valor  $M' < M$  se considera suficiente para la adecuada identificación.

Los pesos obtenidos forman el vector  $\Omega^T = [\omega_1, \omega_2, \omega_3, \dots, \omega_M, \omega_{M'}]$  que describe la contribución de cada eigenface para representar la imagen de rostro de entrada. El vector puede ser usado en un algoritmo estándar de reconocimiento de patrones para encontrar a cuál clase de rostros predefinido describe mejor la imagen de entrada. El método más simple para determinar cuál clase provee la mejor descripción de una imagen de entrada es encontrar la clase de rostro  $k$  que minimice la distancia euclidiana definida como sigue:

$$E_k = \|\Omega - \Omega_k\| \quad (8)$$

donde el vector  $\Omega_k$  es un vector que describe a la clase  $k$  de rostros.

El objetivo de LDA es encontrar los vectores en el espacio subyacente que mejor discriminen entre clases. Para ello LDA proyecta los vectores obtenidos en PCA en un nuevo subespacio vectorial de manera que la razón entre la distancia entre clases y la distancia dentro

de la clase se maximice. De este modo, se garantiza una máxima discriminación entre las clases.

De este modo el LDA utiliza al PCA para encontrar los vectores que mejor definan las características de las imágenes dentro de todo el espacio de la imagen.

Para todas las imágenes de todas las clases se definen dos tipos de matrices de dispersión, una de ellas es representada como “entre clases”,  $S_B$ , y la otra como “dentro de las clases”,  $S_W$ . Las cuales son calculadas como sigue:

$$S_B = \sum_c N_c \cdot (\mu - \bar{x}) \cdot (\mu - \bar{x})^T \quad (9)$$

$$S_W = \sum_c \sum_{i \in c} N_c \cdot (x_i - \mu_c) \cdot (x_i - \mu_c)^T \quad (10)$$

$$\bar{x} = \frac{1}{N} \sum_i x_i = \frac{1}{N} \sum_c N_c \cdot \mu_c \quad (11)$$

donde  $\bar{x}$  es el promedio de todos los vectores proyectados en el espacio de *eigenfaces*. El vector  $\mu_c$  es el promedio de los vectores proyectados en el espacio *eigenfaces* de cada clase y  $N_c$  es el número de muestra por clase  $c$ , el número de muestras de rostros por usuario.

Para maximizar la razón entre la distancia entre clases y la distancia dentro de la clase se deberá maximizar la siguiente expresión

$$J(W) = \frac{W^T \cdot S_B \cdot W}{W^T \cdot S_W \cdot W} \quad (12)$$

Donde las columnas de la matriz  $W$  es el conjunto de los vectores propios generalizados de  $S_B$  y  $S_W$ . A partir de estos vectores propios generalizados podemos obtener un nuevo espacio vectorial para la representación de las imágenes o espacio de *fisherfaces*. Para la identificación de usuario seguimos un proceso similar al utilizado en el método de *eigenfaces* basado en la distancia mínima euclidiana de los coeficientes en este espacio vectorial.

Cabe destacar que el método de *fisherfaces* es un método más robusto a condiciones de diferente iluminación con respecto al método de *eigenfaces* [22].

#### 4.2 Preprocesamiento de la imagen

Ambos métodos de *eigenfaces* y *fisherfaces* requieren que el rostro bajo análisis esté perfectamente alineado para un óptimo reconocimiento. Por lo mismo se ha implementado 4 etapas de preprocesamiento propuestas por dos de los autores y pueden ser consultadas en [24]. Estas etapas buscan que las

imágenes adquiridas estén alineadas y centradas tomando como referencia la posición de los ojos. Además, se elimina aspectos adicionales al rostro como cabello y cuello. Las etapas de preprocesamiento son las siguientes:

##### A. Detección de rostro por algoritmo de Viola & Jones [25, 26].

Esta etapa tiene por objetivo determinar las coordenadas del rostro analizar (ver figura 5).



Figura 5. Rostro detectado.

##### B. Detector de falsos positivos de rostros.

Este detector de falsos positivos consiste en un filtro detector de color de piel y un detector de ojos (ver figura 6 y figura 7). El detector de ojos está basado en el algoritmo de Viola & Jones entrenado para este propósito [25, 26]. Se considera que el rostro ha sido detectado correctamente si el área correspondiente al mismo cumple con estas dos condiciones:

- Presenta un porcentaje de más de 70% de superficie de piel.
- Presenta dos ojos detectados.



Figura 6. Filtro detector de color de piel.



Figura 7. Ojos detectados.

### C. Ajuste de Rotación, Escalamiento y Centrado

Esta etapa se aplica debido a que los rostros que se detectan no siempre cuentan con la misma posición, inclinación, escalamiento y centrado que se tienen en el conjunto de imágenes de muestra de usuarios. Todos estos pasos dependen de las coordenadas (x, y) de las pupilas, las cuales servirán de referencia para corregir estas diferencias.

### D. Aplicación de máscara

Para que el reconocimiento no sea afectado por el fondo que se toma como parte del rostro, todas las imágenes, incluyendo las de la base de datos se pasan a través de una plantilla, la cual recorta algunas secciones de la imagen de menor interés para el análisis (ver figura 8).



Figura 8. Máscara aplicada a la imagen.

El método de *fisherfaces* requiere para cada usuario más de una foto. En la tarea de registro de usuarios de la plataforma, el sistema realiza siete capturas de imagen consecutivas. De cada imagen se obtiene un rostro y de esta manera se cumple con el requisito del método.

En la tarea de identificación de usuario, el sistema captura 20 imágenes consecutivas. Para cada una de estas fotos se realiza una identificación de usuario. El usuario seleccionado será aquel que sea identificado en mayor número de imágenes, en el caso de que en

algunas de las imágenes el sistema identifique usuarios diferentes debido a errores en el reconocimiento. Estos errores pueden deberse a imágenes en movimiento, mal enfocadas o pobre iluminación.

Tanto en la etapa de registro como en la etapa de identificación, el usuario puede ver en pantalla la imagen capturada por la cámara y un círculo de puntos amarillos alrededor del rostro detectado por el algoritmo de Viola & Jones. Esto se puede ver en la tarea de registro de usuario, presentada en la figura 9 y en la tarea de identificación de usuario, presentada en la figura 10. En caso de que haya más de un rostro en la imagen capturada, el sistema solo toma en cuenta aquel que ocupe una mayor superficie de la imagen, asumiendo que será el del usuario más cercano a la cámara (ver figura 11).

El tiempo que toma el sistema para identificar el usuario, incluyendo la activación de la cámara, la captura de las imágenes y el reconocimiento facial, es de aproximadamente 15 segundos en una computadora con procesador Core i7 de 2.5GHz y 8 GB de memoria RAM.

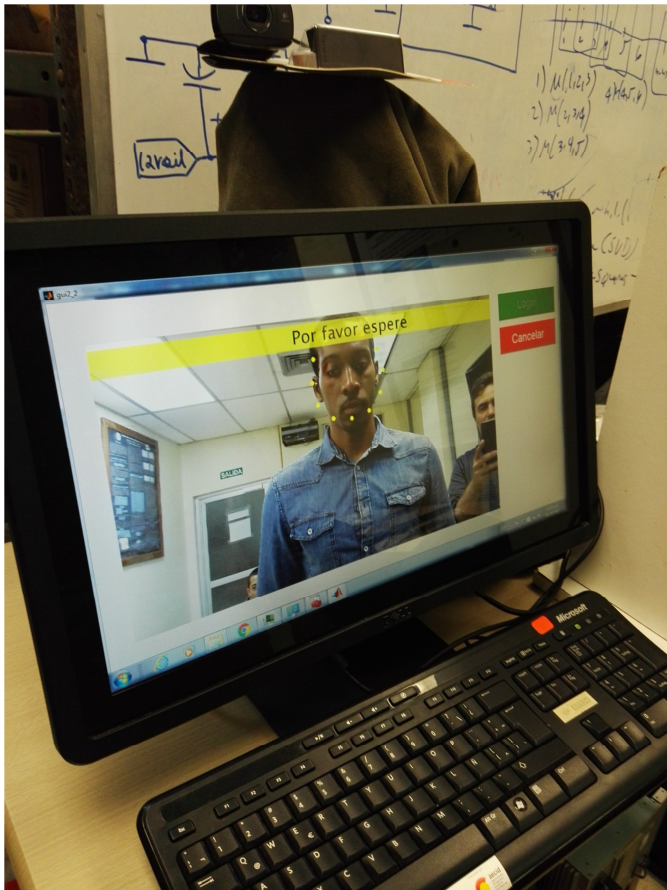


Figura 9. Ejemplo de etapa de registro de usuario.



Figura 10. Ejemplo de etapa de identificación de usuario.





**Figura 11.** Identificación de un usuario con la presencia de más de un rostro en la imagen.

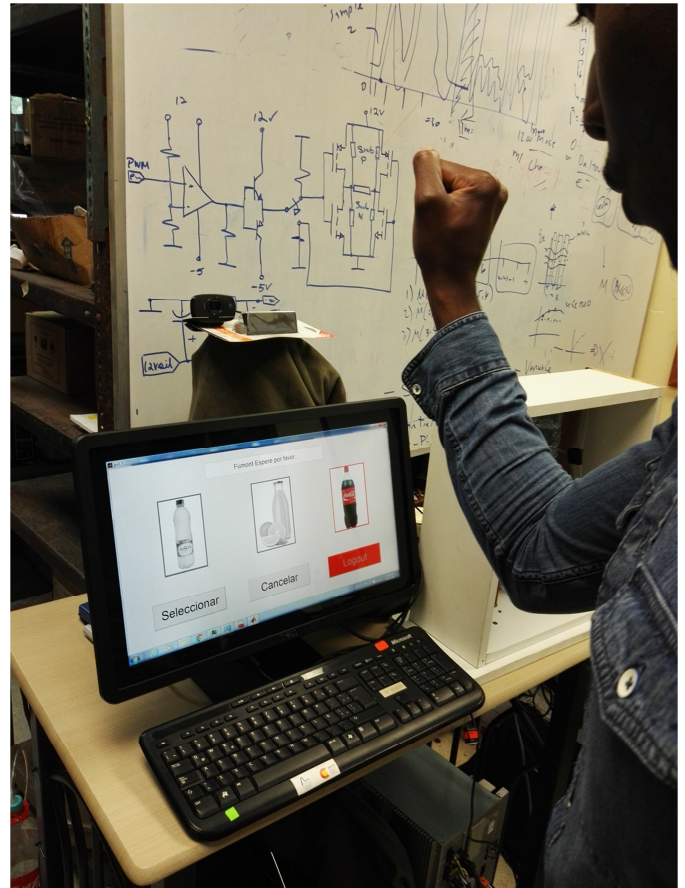
#### 4.1 Reconocimiento de gestos de mano para selección de bebidas

Para el reconocimiento de gestos se utilizó un detector entrenado para el reconocimiento de la mano en posición de puño en posición vertical. Se puede apreciar en la figura 11 el gesto de mano mencionado. El detector está basado en el uso de filtros de Haar entrenados en cascada usando el método de Viola & Jones [25,26].

En el prototipo, el usuario tiene la opción de seleccionar entre tres tipos de bebida. En la etapa de selección de bebida el algoritmo identifica en coordenadas de la imagen se encuentra el puño. La bebida será seleccionada en función de la posición de estas coordenadas dependiendo si las coordenadas se encuentran en el tercio izquierdo, central o derecho de la imagen. Por ejemplo, si el gesto de puño vertical es detectado en el tercio derecho del de la imagen captada por la cámara (es decir, según la perspectiva del

usuario), la bebida seleccionada será aquella correspondiente al lado derecho. En la interfaz gráfica, la imagen de la bebida pasará de ser en tono de grises a estar a todo color (ver la figura 12).

El detector de gestos opera en tiempo real a una tasa de 30 imágenes por segundo.



**Figura 12.** Etapa de selección de bebida por reconocimiento de gesto.

### 5. Protocolo de pruebas de experiencia de usuario

Estos tres escenarios que tiene el prototipo permiten realizar pruebas centradas en usuario para estudiar la experiencia de los mismos al utilizar el sistema. Para evitar sesgos se intercalan los escenarios de prueba para los diferentes participantes de la prueba. Mediante técnicas tales como encuestas, entrevistas, observación, cuestionarios previos y posteriores a la prueba y registro de tiempo de las etapas se puede determinar cómo interactúan y se comportan los usuarios con el prototipo.



El estudio está en progreso con participantes de diferente género, edades y nivel de formación.

Entre los resultados del estudio se espera obtener:

- Promedio de tiempo requerido en cada etapa y cada escenario en función del género, edad y antecedente del usuario.
- Escenarios preferidos según la categoría de usuario.
- Ventajas y dificultades del uso de reconocimiento facial y reconocimiento de gestos en sistemas de este tipo.

Los resultados de este estudio serán presentados en una futura publicación.

## 6. Observaciones de pruebas iniciales de la plataforma de pruebas

Una vez que el prototipo estuvo en operación se realizaron pruebas con 15 participantes a fin de identificar problemas técnicos y de experiencia de usuario de este prototipo. Algunas de estas observaciones ya han sido tomadas en cuenta en ciertos aspectos del sistema. Las observaciones de estas pruebas son las siguientes:

- La cámara debe ser ubicada de tal manera que los usuarios de diferentes estaturas entren en el campo de captura de la misma. En algunas de las pruebas, no se capturó el rostro de participantes altos y ubicados muy cerca de la cámara.
- En escenarios reales pueden haber más de una persona en la pantalla. Es necesario que el sistema pueda manejar esta situación seleccionando al usuario más cercano a la cámara.
- Presentar al usuario elementos que indiquen que las funciones de visión artificial están activas favorece las interacciones con la plataforma. Ejemplo de elementos indicativos son: el círculo de puntos amarillos alrededor del rostro detectado en la etapa de registro de usuario y en la etapa de identificación de usuario.
- Es conveniente presentar mensajes al usuario con indicaciones de acciones o de espera para facilitar la interacción.
- Para aumentar la fiabilidad del sistema a la hora de registrar o identificar los usuarios es preferible trabajar con varias imágenes capturadas. Esto permite minimizar errores de reconocimiento por problemas de imágenes en movimiento.

## 7. Perspectivas del sistema

Este prototipo puede ser mejorado en varios aspectos. Por un lado, se plantea implementar el sistema en una computadora más compacta tal como la Raspberry Pi. Se tiene planificado que la interfaz gráfica y los algoritmos sean implementados utilizando otros lenguajes y plataformas de desarrollo abiertas, pensando en una posible implementación comercial.

Con respecto a los algoritmos de visión, se plantea estudiar la pertinencia de otros métodos de reconocimiento de rostros y gestos basados en redes neuronales convolucionales (CNN, por sus siglas en inglés) [27], considerando el contexto de aplicación. En lo que a la selección de bebidas se refiere podrían explorarse otros enfoques tales como el uso de diferentes tipos de gestos de manos, análisis de orientación de un dedo que señale el producto deseado o el uso de seguimiento de la vista.

### 7.1 Arquitectura distribuida propuesta

En paralelo con la implementación de este sistema se realizó una reflexión sobre el tipo de arquitectura que podría tener un sistema de reconocimiento de usuario en la plataforma de venta automática operando en diferentes tiendas de una franquicia.

Podemos notar que, en la mayoría de las aplicaciones de reconocimiento facial, por ejemplo, en sistemas de videovigilancia, los algoritmos son ejecutados en un servidor, mientras que el dispositivo local solo dispone del sistema de adquisición de imágenes. También, en muchos casos, las bases de datos están preestablecidas

Si consideramos el contexto de múltiples dispositivos dispensadores en varios lugares, proponemos el uso de una arquitectura distribuida cliente-servidor para el sistema de reconocimiento de rostro. En esta arquitectura, el sistema cliente ejecuta una rutina de identificación de bajo costo computacional usando parámetros o detectores entrenados previamente provistos por el servidor. Por otra parte, el servidor será el responsable de la etapa de entrenamiento del algoritmo. En el caso de la llegada de un nuevo usuario, el cliente proporcionará al servidor de un grupo de imágenes del usuario. Luego el servidor procede a la etapa de entrenamiento del algoritmo y finalmente proporciona a los sistemas clientes de los parámetros o detectores entrenados actualizados para la tarea de identificación.

## 7.2 Posibles aplicaciones y ventajas del sistema propuesto

Este tipo de sistema de venta de productos inteligente podría implementarse en otros contextos. Por ejemplo, en las escuelas podría permitir un monitoreo y control del consumo de comidas dispensadas en máquinas de ventas a cada niño. Incluso podría permitir el establecimiento de políticas de control de venta en función de la edad o problemas de salud, incluyendo alergias a productos.

También en el ámbito escolar, un sistema similar se podría establecer a nivel de las cajas de cafeterías para un cobro directo a la cuenta de los clientes por reconocimiento facial.

## 8. Conclusiones

Este artículo presenta un prototipo de un sistema de venta de bebidas automático con funcionalidades basadas de visión por computadora para la identificación de usuarios y selección de productos. El sistema se ha diseñado como una plataforma para realizar estudios de experiencia de usuarios bajo diferentes escenarios que permiten comparar la aceptación de las funcionalidades mencionadas con respecto a las tecnologías de tipo táctil.

Se han propuesto perspectivas al sistema con el fin de hacerlo más adaptado a condiciones comerciales de operación.

## 9. Agradecimientos

Agradecemos a Víctor López y Jorge Lezcano de la Facultad de Ingeniería en Sistemas Computacionales de la Universidad Tecnológica de Panamá, por los intercambios en torno a la arquitectura distribuida propuesta para el sistema.

Este trabajo es parcialmente financiado por el Sistema Nacional de Investigación (SNI) de Panamá bajo el contrato No.124-2015 correspondiente al primer autor.

## 10. Referencias

[1] V. Gouaillier and A. Fleurant, "Intelligent video surveillance: Promises and challenges," Technological and commercial intelligence report, CRIM and Technôpole Defence and Security, pp. 456-468, 2009.

[2] H. K. Ekenel and B. Sankur, "Feature selection in the independent component subspace for face recognition,"

Pattern Recognition Letters, vol. 25, no. 12, pp. 1377-1388, 2004.

[3] A. K. Jain and A. Kumar, "Biometrics of next generation: An overview," Second Generation Biometrics, 2010.

[4] X. Wang, Q. Ruan, and Y. Ming, "3D Face recognition using Corresponding Point Direction Measure and depth local features," Signal Processing (ICSP), 2010 IEEE 10th International Conference on. IEEE, pp. 86-89, 2010.

[5] V. Zeljkovic, D. Zhang, V. Valev, Z. Zhang and J. Li, "Personal access control system using moving object detection and face recognition," High Performance Computing & Simulation (HPCS), 2014 International Conference on. IEEE, pp. 662-669, 2014.

[6] Q. Al-Shebani, P. Premarante, and P. Vial, "Embedded door access control systems based on face recognition: a survey," Signal Processing and Communication Systems (ICSPCS), 2013 7th International Conference on , vol., no., pp.1,7, 16-18 Dec. 2013.

[7] J. Harguess and J. K. Aggarwal, "A case for the average-half-face in 2D and 3D for face recognition," Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on. IEEE, pp. 7-12, 2009.

[8] F. Merchán, S. Galeano and H. Poveda, Mejoras en el Entrenamiento de Esquemas de Detección de Sonrisas basados en AdaBoost, Revista de I+D Tecnológico, Vol. 10, No.2, pp. 17-30, Dec. 2014.

[9] F. Zuo, and P. H. N. de With, "Real-time embedded face recognition for smart home," Consumer Electronics, IEEE Transactions on, vol. 51, no. 1 pp. 183-190, 2005.

[10] R. Meng, Z. Shengbing, L. Yi and Z. Meng, "CUDA-based real-time face recognition system," Digital Information and Communication Technology and its Applications (DICTAP), 2014 Fourth International Conference on. IEEE, pp. 237-241, 2014.

[11] S. Lin, "The study and implementation of real-time face recognition and tracking system," 2010 International Conference on Machine Learning and Cybernetics, Vol. 6, pp. 3050-3055, 2010.

[12] M. Janarbek, A. Irturk, and R. Kastner, "Design and implementation of an fpga-based real-time face recognition system," Field-Programmable Custom Computing Machines (FCCM), 2011 IEEE 19th Annual International Symposium on. IEEE, pp. 97-100, 2011.

[13] S. Sardar, and K. Babu, "Hardware Implementation of Real-Time, High Performance, RCE-NN Based Face Recognition System," VLSI Design and 2014 13th International Conference on Embedded Systems, 2014 27th International Conference on, IEEE, pp. 174-179, 2014.

[14] World News Report, June 14, 2010, "Microsoft fully unveils kinect for xbox 360 controller-free game device" [http://world.einnews.com/pr\\_news/56709028/microsoft-fully-unveils-kinect-for-xbox-360-controller-free-game-device](http://world.einnews.com/pr_news/56709028/microsoft-fully-unveils-kinect-for-xbox-360-controller-free-game-device)

[15] H. Jalab and H. Omer, "Human computer interface using hand gesture recognition based on neural network," in Information

- Technology: Towards New Smart World (NSITNSW), 2015  
5th National Symposium on , vol., no., pp.1-6, 17-19 Feb.  
2015
- [16] J. Nagi, F. Ducatelle, G.A. Di Caro, D. Ciresan, and U. Meier, A. Giusti, F. Nagi and J. Schmidhuber and L.M. Gambardella, "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in Signal and Image Processing Applications (ICSIPA), 2011 IEEE International Conference on , vol., no., pp.342-347, 16-18 Nov. 2011.
- [17] A. White, R. Godard, C. Belling, V. Kasza, R. L. Beach, Beverages obtained from soda fountain machines in the U.S. contain microorganisms, including coliform bacteria, International Journal of Food Microbiology, Volume 137, Issue 1, 31 January 2010
- [18] Barton, J. C., & Barton, N. J. (2004). U.S. Patent No. 6,688,134. Washington, DC: U.S. Patent and Trademark Office.
- [19] C. Smolen (2013). U.S. Patent No. US 8417376 B1. U.S. Patent and Trademark Office
- [20] K. Panos, "Smart kegerator bills based on beer consumption", hackaday.com, March 11, 2014, <http://hackaday.com/2014/03/11/smart-kegerator-bills-based-on-beer-consumption/>
- [21] J. Ye, R. Janardan, Q. Li, "Two-Dimensional Linear Discriminant Analysis," Advances in neural information processing systems, pp. 1569-1576, 2004.
- [22] P. Belhumeur, J. Hespanha and David Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 19, no. 7, pp. 711-720, 1997.
- [23] M. Turk, A. Pentland, Eigenfaces for Recognition, Journal of Cognitive Neuroscience, Vol. 3, No. 1, 1991, pp. 71-86.
- [24] M. Poveda, F. Merchán, Implementación de un sistema de control de acceso basado en reconocimiento facial, Revista Prisma Tecnológico, ISSN 2076-8133, Vol. 6, No.1, 2015.
- [25] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137-154, 2004.
- [26] P. Viola and M. J. Jones, "Fast Multi-view Face Detection," Mitsubishi Electric Research Laboratories, TR2003-096, vol. 3, pp. 14, 2003.
- [27] H. Li, Z. Lin, X. Shen, J. Brandt and G. Hua, "A convolutional neural network cascade for face detection," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 5325-5334.