

Telefonía Visual

BOSQUEJOS DE CARAS

En escenas típicas



genera imágenes en semitono ("halftoning") haciendo énfasis en la información de la orientación de la imagen. El segundo método genera imágenes en colores, cuantificando los tonos originales mediante el uso de la mínima distancia euclidiana. Previamente se detectan y localizan algunas características faciales como ojos, nariz y boca, para generar mayor detalle en estas regiones de interés (RDI).

PALABRAS CLAVES

Cuantificación de colores, *half-toning*, localización de características faciales, procesamiento de imágenes.

ABSTRACT

This paper proposes two methods to reduce information in video sequences that contain faces to use in applications such transmission of visual telephony through very low capacity channels. The first method generates halftone images with emphasis in the image orientation. The second method generates images in colors,

Omar Ricardo Ayala Peñuela, Ingeniero Electrónico.

Candidato a MSc en Automatización y Robótica. Universidad de Dortmund. Dortmund, Alemania.

Mauricio Díaz Melo, Ingeniero Electrónico.

Candidato a MSc en Ingeniería Electrónica. Pontificia Universidad

Profesor Pedro Raúl Vizcaya Guarín, PhD.

Director de la Maestría en Ingeniería Electrónica. Pontificia Universidad Javeriana.

RESUMEN

Este artículo propone dos métodos para reducir la información en secuencias de

1. INTRODUCCIÓN

La búsqueda de un uso óptimo y eficiente del espectro electromagnético se ha convertido en una de las prioridades de mayor interés en los actuales trabajos de investigación. Dentro de este campo se destaca el surgimiento de tecnologías que involucran el uso de sistemas audiovisuales, tales como la telefonía visual, como medios de comunicación masivos. Estudios realizados acerca de la información proporcionada por los mensajes audiovisuales, concluyen que el video es útil para aumentar la inteligibilidad de señales de voz en valores que van desde uno hasta tres decibeles [1].

Se han desarrollado investigaciones, como la realizada por Peñuela [2] con el objetivo de comprimir imágenes basándose en la detección de bordes. A partir de los resultados obtenidos en este trabajo surgió la idea de obtener un bosquejo del rostro en una secuencia de video que no pierda inteligibilidad para ser utilizado en telefonía visual. Para esto se proponen dos métodos diferentes, el primero de ellos genera imágenes en semitono y el segundo imágenes en colores.

Un paso previo que mejora la inteligibilidad de las imágenes consiste en encontrar unas determinadas regiones de interés (ROI), las cuales son procesadas con parámetros diferentes. Esto con el objetivo de dar una mayor relevancia a las características que hacen identificable a un rostro humano. El proceso de detección y segmentación de las características faciales busca encontrar la región de los ojos y la boca.

El método de imágenes en semitono, antes que buscar representar tonos de grises intermedios, como en los métodos implementados hasta el momento [9], busca resaltar la información proporcionada por la orientación, que posee una imagen [4]. Este método se basa en el uso iterativo de la transformada de Poincaré [5], y la aplicación de una función no lineal a los planos generados por esta transformación.

El método de cuantificación de colores representa los colores de la imagen original sobre un mapa previamente seleccionado a partir de una serie de muestras tomadas y clasificadas mediante el método de k-medias [6]. La cuantificación se realiza usando como parámetro la mínima distancia euclidiana. Para evitar efectos de falsos bordes se realiza una difusión del error de acuerdo a la matriz planteada por Floyd y Steinberg [3].

Estos dos procesos mencionados anteriormente se realizan variando parámetros de acuerdo a las regiones de interés que previamente se han detectado mediante la localización de las características faciales [7], [8].

DESARROLLOS

1.1 Localización de características faciales

La primera etapa consiste en detectar las características faciales de cada una de las imágenes que conforman el video. Las técnicas de detección y localización de características faciales han sido ampliamente estudiadas y desarrolladas en los últimos años debido a la amplia gama de aplicaciones que usan la detección de rostros (y sus características) en sistemas de mayor complejidad.

Yang *et al.* [7] realizaron un estudio a comienzos del nuevo milenio en el cual analizaron los diferentes métodos que se han desarrollado para la detección de caras. Dentro de ese estudio clasificaron las técnicas utilizadas para la detección de rostros dentro de cuatro categorías básicas, aclarando que algunos métodos pueden sobrelapar las fronteras de las otras categorías. Estos métodos son:

- **Métodos basados en el conocimiento.** Este método basado en reglas codifica el conocimiento humano de que constituye una cara típica. Usualmente, las reglas capturan las relaciones entre las características faciales.
- **Aproximación por características invariantes.** Estos algoritmos apuntan a encontrar estructuras características que existen aun cuando la posición, el punto de vista, o las condiciones de iluminación varíen, y después los usan para localizar las caras.
- **Métodos de apareamiento de plantillas.** Patrones estándar de caras son almacenados para describir una cara como un todo, o las características faciales separadamente. La correlación entre una imagen de entrada y los patrones almacenados son procesados para la detección.
- **Métodos basados en la apariencia.** Al contrario del método de apareamiento de plantillas, los modelos son aprendidos de un grupo de imágenes de entrenamiento, las cuales pueden capturar la variabilidad de una apariencia facial. Estos modelos aprendidos son usados para la detección.

Dentro de las técnicas que han sido desarrolladas con mejores resultados, teniendo en cuenta cantidad de aciertos y tiempo de procesamiento, los sistemas que se basan en la detección por color han tenido un relativo éxito. Teniendo en cuenta las condiciones controladas dentro de las cuales se desarrolló este proyecto se optó por hacer uso de las características de color, para hacer una primera aproximación a la detección

de la cara, y posteriormente delimitar la región de búsqueda de las características faciales a la zona encontrada previamente.

Diversos estudios han determinado ciertas regiones en diversos espacios de colores en las cuales se agrupan los píxeles que pertenecen a la cara, sin importar la raza o color de piel. El espacio YCbCr según la investigación realizada por Terrillon *et al.* [8], resulta ser uno de los más óptimos para la detección de las características faciales. Las componentes (planos) Cb y Cr generan máscaras que discriminan la piel del resto de la imagen (figura 2). Para cada uno de estos planos se expande su histograma dentro del rango dinámico para mejorar el contraste, y son procesados de manera independiente.

Los planos son binarizados (blanco y negro) de acuerdo con un umbral adaptativo mediante el algoritmo de Otsu [12]. Este método da por hecho la existencia de dos clases definidas en la imagen y busca el punto que haga mínima la varianza entre estas las dos clases. Para este caso resulta perfecto, ya que se definen estas clases como el fondo (uniforme) y la cara.

Las dos máscaras generadas se procesan mediante la operación lógica AND, para descartar las regiones que no pertenecen a la piel. La máscara resultante de esta operación se aproxima utilizando la elipse que mejor se acomoda a estos puntos. El centro de masa corresponde al centro de la elipse, y a partir de los momentos es posible determinar el ángulo de inclinación (θ), el eje mayor (a) y el eje menor (b) [10]. Estos vienen dados por:

$$\theta = \frac{1}{2} \cdot \arctan \left(\frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}} \right)$$

$$a = \left(\frac{4}{\pi} \right)^{1/4} \left[\frac{(I_{\max})^3}{I_{\min}} \right]^{1/8}$$

$$b = \left(\frac{4}{\pi} \right)^{1/4} \left[\frac{(I_{\min})^3}{I_{\max}} \right]^{1/8}$$

donde $I_{\max}$ e $I_{\min}$ representan el mínimo y máximo momento de inercia de una elipse con orientación θ .

Una vez encontrada la elipse el análisis se limita a esta región. Se selecciona la parte superior de la elipse formando una semielipse con un arco entre 0° y 180° .

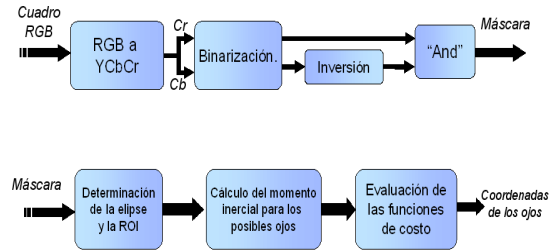


Figura 1. Diagrama de bloques localización de las características faciales.



Figura 2.(a) Planos Cb (b) Plano Cr.

Los “huecos” (regiones de “unos” en un contorno de “ceros”), ubicados dentro de la semielipse forman los posibles candidatos a ojos (figura 3). La selección de los ojos dentro de este grupo de candidatos se realiza utilizando una serie de funciones de costo, que de acuerdo a la posición del centro de masa de cada hueco desecha los puntos que no cumplen con la simetría de un rostro humano [11].

Las funciones de costo que seleccionan los candidatos a ojos son:

- Diferencia de los centros de masa de cada una de las parejas respecto al eje menor de la elipse.
- Simetría respecto al eje Y. Es decir, que la pareja de centros de masa candidata a ojos tenga cada coordenada a un lado del eje mayor de la elipse.
- Menor distancia de las parejas respecto al eje menor de la elipse.



Figura 3. Contornos hallados.

1.2 Método de imágenes en semitono

La generación de imágenes en semitono o *halftoning*, Es una técnica que representa las imágenes, utilizando únicamente un número

reducido de colores. Al disminuirse el número de colores, hay ciertas tonalidades que no pueden ser representadas de forma precisa con los tonos que se tienen. Para evitar que la imagen pierda naturalidad y que se cree un efecto de falsos contornos, se colocan puntos de diferentes colores en esas regiones, con el fin de crear en la imagen el efecto de la existencia de tonos intermedios a los cuantificados. Este proceso aprovecha la propiedad del ojo de realizar la integración de la imagen, percibiendo únicamente la intensidad media del área observada [13]. Esta característica fisiológica permite que se pueda “engañar” al cerebro y así, por ejemplo, hacer que se piense que una imagen de solo puntos blancos y negros parezca que tiene tonos de grises variando la cantidad de puntos existentes en una región.

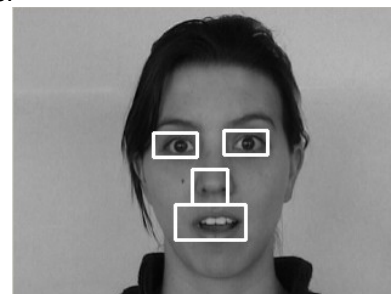
Dentro de los diferentes algoritmos de “*halftoning*” existen unos llamados “*screening*” que básicamente tienen ciertas plantillas o pantallas, que poseen una determinada distribución y densidad de puntos a lo largo de la plantilla, de acuerdo con el valor promedio de grises que una región específica posee. Esa región es reemplazada por la pantalla que más se aproxime. Otros métodos utilizan la orientación de la imagen, que permite transiciones adecuadas en los bordes pero posee problemas en cuanto a granulosidad. En cambio, los métodos de error difusión producen los mejores resultados en lo que a granulosidad se refiere, pero pierden información de bordes y cambios bruscos de tonalidad. Hay algoritmos que son modelos híbridos y mezclan ambas características para obtener las ventajas de ambos métodos [4]. El método desarrollado en este trabajo se diferencia de todos los demás en que debido a la intención de obtener un bosquejo de un rostro, no es de interés el hecho de generar tonos intermedios en regiones como la frente, mejillas, fondo, etc.

Mediante este algoritmo se generan secuencias de video, cuyos cuadros son de tipo binario, a partir de una secuencia de video en escala de grises. Para ello se le da una gran importancia a la orientación de cada uno de los cuadros mediante la transformada de Poincaré.

La transformada de Poincaré [5] es una transformación lineal bidimensional que a partir de 4 píxeles en una vecindad de 2×2 genera 4 valores (DFT-4). Aplicando esta transformación a toda la imagen se obtienen cuatro planos (figura 6), en donde el plano cero almacena la intensidad promedio de la imagen original. Los planos uno y tres son complejos conjugados y poseen información relacionada con la orientación de la imagen. El plano dos representa la coherencia de los planos uno y tres.

Para este procedimiento no es relevante representar todos los niveles de intensidad de gris a lo largo de toda la imagen, sino que lo importante es hacer énfasis en aquellos aspectos o características faciales con los que normalmente un ser humano puede distinguir a una persona de otra. Las regiones a las que se les dio un mayor detalle fueron los ojos y la boca que son determinadas previamente en la etapa de detección y segmentación, dentro del cuadro que se está procesando.

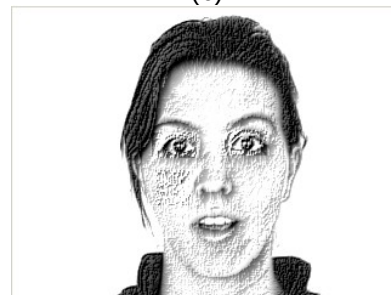
El procedimiento consiste en aplicar la transformación de Poincaré y posteriormente una función no lineal (figura 7), que depende de la región sobre la cual se está trabajando. Este proceso se aplica de forma iterativa, creando sucesivamente imágenes con tonos de grises reducidos, hasta finalmente obtener la imagen binaria. Uno de los puntos importantes que se observaron fue la creación de imágenes intermedias (figura 4), las cuales poseen un efecto artístico tipo “carboncillo” dando la apariencia de las representaciones hechas a mano con este tipo de elementos.



(a)



(b)



(c)

Figura 4. (a) Imagen original. (b) Segunda iteración. (c) Tercera iteración.

Los planos uno y tres, obtenidos con la transformada de Poincaré, son modificados con una función no lineal reduciendo de forma significativa los altos contrastes. El valor del exponente varía según la región que se está procesando. Cuando el detalle dado a las diferentes regiones es muy diferente, se presenta un notorio efecto de ventaneo que debe ser corregido. Este efecto consiste en la aparición de formas rectangulares en las zonas limítrofes de cada una de las regiones internas, debido al cambio brusco que ocurre entre estos grupos de píxeles. El plano dos se mantiene mientras que el plano cero se conforma del promedio de los otros planos.

Para hacer una transición suave entre las regiones de interés y reducir el efecto de ventaneo se realiza una interpolación lineal en una zona de transición de 10 píxeles entre la región más interna y la más externa. Para cada una de las secciones intermedias de la zona de transición se establece una probabilidad mediante la cual se determina la pertenencia de un píxel dado a una de las regiones adyacente.

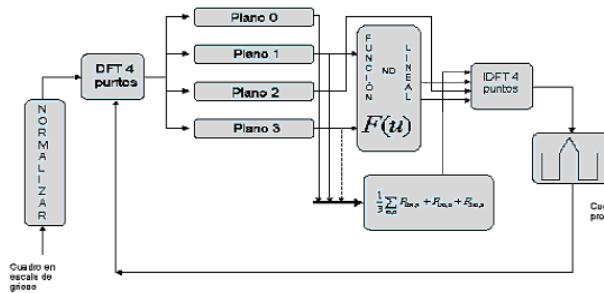


Figura 5. Diagrama de bloques método de imágenes en semitono.

1.3 Método de cuantificación de colores

El método de cuantificación de colores se compone principalmente de dos etapas esenciales: selección de las clases para las diferentes regiones del rostro y la representación de las regiones con las clases cuantificadas.

La necesidad de representar las diferentes regiones del rostro mediante un número limitado de colores o clases es el principio fundamental de operación de este método. Para la selección de estas clases se tomaron una gran cantidad de muestras de una base de datos (imágenes de rostros de diferentes personas), muestras que contienen los colores característicos de cada una

de las regiones de la cara, las cuales son: ojos, boca y piel-pelo. Se utilizó el algoritmo de k-medias [6] capaz de representar una serie de muestras mediante centroides óptimos, en este caso colores óptimos de representación. El número de clases escogidas para cada región se hizo mediante una evaluación subjetiva, teniendo en cuenta que la imagen resultante no puede perder inteligibilidad pero también tiene que ser lo más reducida posible. El resultado fue de 4 clases para la boca, 4 clases para los ojos y 3 clases para la piel y el pelo. Estas clases son las que conforman el mapa de colores para los ojos, la boca, la piel y el pelo, mapas que van a ser utilizados para todos los videos procesados por el método.

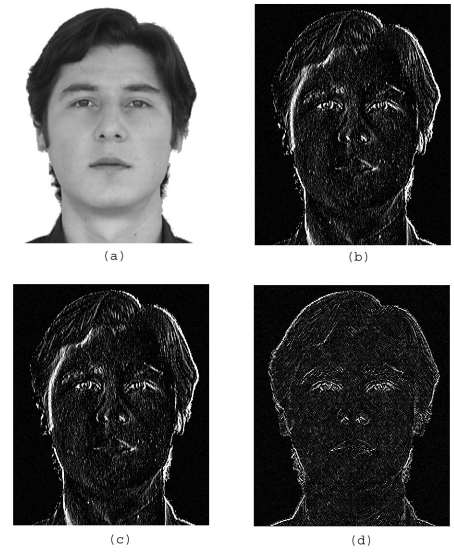


Figura 6. Transformación de Poincaré aplicada a una imagen. (a) Plano cero. (b) Plano uno. (c) Plano dos. (d) Plano tres.

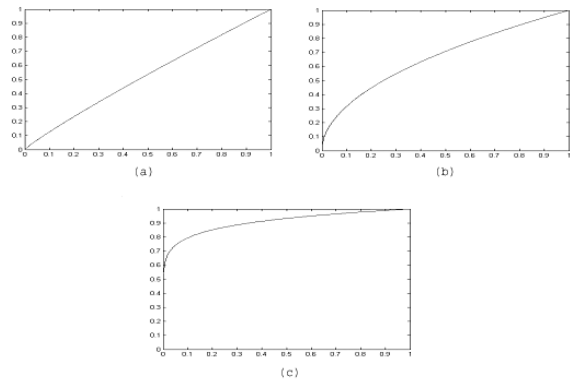


Figura 7. Función aplicada a los planos uno y tres. (a) Exponente 0.9. (b) Exponente 0.5. (c) Exponente 0.1.

Teniendo en cuenta que el sistema ya tiene la localización de las regiones de la cara, se debe proceder a clasificar los colores de cada una de estas con los mapas correspondientes. Este proceso se efectúa mediante la implementación del clasificador de mínima distancia euclidiana, el cual toma el píxel de la imagen de entrada y lo clasifica dentro de las clases correspondientes, asignando una de éstas al píxel de la imagen resultante. Esta asignación la hace escogiendo el color más cercano del mapa de colores de la correspondiente región (mínima distancia euclidiana), al píxel que se está evaluando. La distancia euclidiana está definida de la siguiente forma:

$$d = \sqrt{(R1 - R2)^2 + (G1 - G2)^2 + (B1 - B2)^2} \dots$$

donde d es la distancia entre los píxeles, R representa la componente rojo del píxel, G representa la componente verde del píxel y B representa la componente azul del píxel.

Uno de los problemas presentados fue el ventaneo constante de las regiones de la cara, muy notorio en el video resultante debido a que cada una de las regiones es representada con diferentes colores. Para la solución de este problema se tuvo en cuenta una región de transición entre las ventanas y el fondo, la cual es representada mediante los mapas de las dos regiones para así poder obtener un cambio gradual entre éstas.

El procedimiento implementado para evitar la aparición de falsos contornos en la imagen resultante, fue el de difusión del error de Floyd-Steinberg [3]. Este procedimiento se efectúa después de clasificar cada píxel y consiste en la difusión del error de cuantificación de este en los píxeles vecinos

El error de cuantificación tiene tres componentes, cada una correspondiente a los tres planos del espacio RGB y es distribuido en los píxeles vecinos. La forma de difusión del error se muestra en la figura 8. Las constantes indican la fracción del error que será incrementada en ese píxel antes de ser cuantificado. El resultado del método se observa en la figura 9.

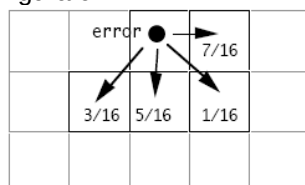


Figura 8. Distribución del error en el método de Floyd-Steinberg.

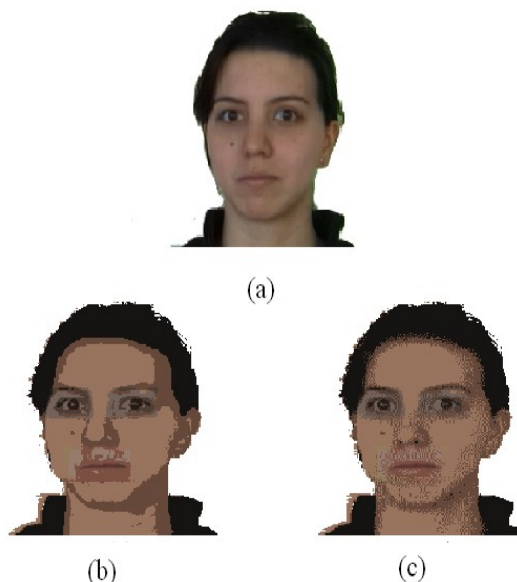


Figura 9. (a) Imagen original (b) Imagen sin difusión de error (c) Imagen con difusión de error.

2. RESULTADOS

Los dos sistemas fueron desarrollados en un computador personal con un procesador Intel Pentium IV de 2.0 GHz. El tiempo de computación utilizado por el método de semitonos fue de 210 ms y el del método de cuantificación de color fue de 72 ms por cuadro. Estos tiempos no incluyen el tiempo requerido para la detección y ubicación de las características faciales, que es de 21 ms por cuadro. El código se desarrolló con la librería de uso libre OpenCV de Intel [16].

La detección de características faciales es exitosa en el 75.5% de los casos analizados. Los errores producidos se deben a factores como la iluminación, posición de la cabeza, obstrucción significativa de la línea de vista hacia los ojos, colores rojizos del fondo, entre otros.

En la evaluación subjetiva del sistema se utilizó una variación del método denominado MOS (*mean opinion score*) [15], mediante el cual se evalúa la inteligibilidad de las imágenes obtenidas por los métodos planteados en este trabajo. Las imágenes seleccionadas para la realización de la prueba fueron escogidas aleatoriamente. La prueba consiste en identificar a una persona dentro de un grupo de 10 personas. Para esto la fotografía de la persona a identificar es procesada y las de otras personas son mostradas en el formato original, o viceversa. Es importante resaltar que los evaluadores pueden o no, conocer previamente a aquellos sujetos que aparecen en los formatos de

prueba con los cuales se realizó el estudio. Igualmente se intentó omitir toda la información, diferente de los rasgos faciales, que permita reconocer a una persona. Se realizaron seis pruebas diferentes sobre una población total de 34 evaluadores.

Los resultados obtenidos fueron clasificados de acuerdo a si el evaluador conocía previamente a alguna de las personas que aparecen en las pruebas.

Las personas que conocen previamente a alguno de los sujetos que aparecen en las pruebas acertaron, en promedio, en un 90,67% de las pruebas presentadas, mientras que aquellas que no conocían con anterioridad a alguna de las personas presentadas en las pruebas, sólo acertaron en promedio un 79,66% de los casos. Estos resultados muestran un alto grado de inteligibilidad de las imágenes procesadas por los dos métodos. Teniendo en cuenta que el sistema se ha desarrollado primordialmente para aplicaciones de telefonía visual estos valores pueden aumentar, ya que una secuencia de video contiene mayor información que un solo cuadro.

3. CONCLUSIONES

- La transformada de Poincaré puede ser utilizada con buenos resultados para la obtención de un bosquejo del rostro.
- Los bosquejos obtenidos se caracterizan por la inteligibilidad de la imagen de la persona que se está mostrando.
- El método de distribución del error de Floyd-Steinberg aporta naturalidad al bosquejo obtenido por medio de la cuantificación de colores.
- Para la obtención de un bosquejo no son críticos los casos fallidos en la detección de características faciales.
- Se pueden lograr niveles de alta compresión sin perder información fundamental para reconocer el bosquejo.
- Los tiempos de procesamiento obtenidos en los métodos pueden llegar a ser apropiados para aplicaciones de video en tiempo real.

4. REFERENCIAS BIBLIOGRÁFICAS

- [1] SUMBY W. and POLLACK I. "Visual contribution to speech intelligibility in noise". *Journal of Acoustic Society of America*, Vol. XXVI. 1954. Págs. 212-215.
- [2] PEÑUELA, E.J.; RESTREPO, A. and VEGA J.A. "Color Edges Maps for Image Compression". Universidad de los Andes. 2000.
- [3] FLOYD, R.W. and STEINBERG, L. "Adaptive Algorithm for Spatial Grey Scale". *SID Int. Sym. Dig. Tech. Papers*. Págs 36-37. 1975.
- [4] SHERSTINSKY, A. and PICARD, R. "M-Lattice: A Novel Non-Linear Dynamical System and its Application to Halftoning". The Media Laboratory, Department of electrical engineering and Computer Science, Massachusetts Institute of Technology. Cambridge MA – U.S.A, 1994, 6p.
- [5] VIZCAYA, P. R. "Transformación De Poincaré Y Detección De Detalles En Imágenes De Huellas Dactilares". Pontificia Universidad Javeriana. 2001.
- [6] LOONEY, C.G. "Pattern Recognition Using Neural Networks". Oxford University Press. Págs 29 - 33. 1997.
- [7] YANG, M.H.; KRIEGMAN, D.J. and AHUJA, N. "Detecting Faces in Images: A Survey". *IEEE Transaction On Pattern Analysis and Machine Intelligence*. Vol. XXIV, No 2. Págs. 34-58. Enero 2002.
- [8] TERRILLON, J.C; SHIRAZI, M.N; FUKAMACHI, H and AKAMATSU, S.. "Comparative Performance of Different Skin Chrominance Spaces for the Automatic Detection of Human Faces in Color Images". *Proc. IEEE Int'l Conf on Face and Gesture Reconignition*, Págs. 54-61, 2000.
- [9] PAPPAS, T.N; ALLEBACH, J.P and NEUHOFF, D.L.. "Model-Based Digital Halftoning". *IEEE Signal Processing Magazine*. Vol XX. Julio 2003.
- [10] SOBOTKA, K. and PITAS, I. "Extractions of Facial Regions and Features Using Color and Shape Information". Department of Informatics. University of Thessaloniki. 2001.
- [11] SABER, E; MURAT-TEKALP, A. "Frontal-view Face Detection and Facial Feature extraction Using Color". Department of Electrical Engineering and center for Electronic Imaging Systems. University of Rochester. Rochester, N.Y.
- [12] OTSU, N. "A Threshold Selection Method from Gray-Level Histograms". *IEEE Transactions On Systems, Man, And Cybernetics*. Vol SMC-9. Enero 1979.
- [13] GONZALES, R and WOODS, R. "Digital Signal Processing". Addison Wesley Publishing Company. 1992.

- [14] WANG, Z and BOVIK, C. "A Universal Image Quality Index". IEEE Signal Processing Letters. Vol XX. Marzo 2002.
- [15] TOBIAS, O. "How To Meet The User Requirement Of Room-Based Videoconferencing". Royal Institute of Technology. 1999.
- [16] <http://developer.intel.com>. Open Source Computer Vision Library. Reference Manual. Intel. 2001.



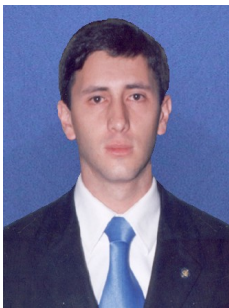
Pedro Raúl Vizcaya Guarín. Ingeniero Electrónico (2004) de la Pontificia Universidad Javeriana, Bogotá (Col). MsC. En Ingeniería, Instituto Politécnico Rensselaer, Troy, Nueva York, EE.UU , 1986. PhD. En Ingeniería, Instituto Politécnico Rensselaer, Troy, Nueva York,

EE.UU, 1997.

Desde 1980, profesor a tiempo completo del Departamento de Electrónica, Pontificia Universidad Javeriana. Actualmente Director de la Maestría en Ingeniería Electrónica en la Pontificia Universidad Javeriana.

pvizcaya@javeriana.edu.co

AUTORES



Omar Ricardo Ayala Peñuela. Ingeniero Electrónico (2004) de la Pontificia Universidad Javeriana, Bogotá (Col). Estudiante de la Maestría en Automatización y Robótica en la Universidad de Dortmund, Dortmund (Alemania).

Correo electrónico:

orap Orion@hotmail.com



Mauricio Díaz Melo. Ingeniero Electrónico (2004) de la Pontificia Universidad Javeriana, Bogotá, Colombia. Estudiante de la Maestría en Ingeniería Electrónica en la misma universidad.

Actualmente se desempeña como auxiliar de investigación en el proyecto "Telefonía

Visual por canales de muy baja capacidad". Pontificia Universidad Javeriana. Cra 7 # 40-62. Bogotá Colombia. Correo electrónico:

mauricio.diaz@javeriana.edu.co



Diego Fernando Macías Sánchez. Ingeniero Electrónico (2004) de la Pontificia Universidad Javeriana, Bogotá, Colombia.

En la actualidad se desempeña como practicante de UCE en Telefónica Data.

Correo electrónico:

dmacias13@cable.net.co