

adecuados para realizar la debida separación de las señales fetales y maternas, aplicando el algoritmo JADEop.

E. Detección de onda R

Permite detectar los picos R del ECG necesarios para los análisis de variabilidad de la rata cardiaca y extracción de la actividad auricular, mediante la sustracción del complejo promedio o mediante la cancelación espaciotemporal del complejo QRST.

F. Análisis espectral

Aplica la transformada de Fourier mediante el periodograma de Welch, y presenta en pantalla el valor en frecuencia de los tres picos principales, útil en el análisis de arritmias auriculares.

G. Medidas en tiempo y en amplitud

Si bien el software del ECG-PS puede realizar estas funciones, este módulo también permite activar un par de cursores horizontales o verticales. Se puede ubicar con ayuda del *mouse* los cursores en los puntos de interés obteniendo como resultado la diferencia entre las magnitudes de las variables que se estén midiendo, bien sea tiempo, amplitud o frecuencia. Esta opción es útil para medir magnitudes de las ondas del ECG, obtener la frecuencia cardiaca o determinar el ancho de banda de un pico en el espectro.

H. Análisis tiempo-frecuencia (TF)

Inicia un menú en el que se puede seleccionar diversas transformadas TF entre las que se incluye las transformadas de Wigner Ville y Choi Williams.

I. Aplicación del algoritmo JADEop

Esta opción ejecuta el algoritmo JADEop que es aplicado a las señales fuente para la obtención de las componentes FECCG y MECCG separadas, permitiendo la posterior selección de la señal FECCG.

J. Salvar y exportar

Luego de cualquier procedimiento de análisis, es posible guardar los resultados obtenidos hasta ese momento. También es posible exportar los registros procesados al software del ECG-PS.

IV. Implementación del Software

Para la selección del algoritmo JADEop, se utilizaron los registros de la base de datos DaISy [17] compuesta por ocho derivaciones, obtenidas ubicando 5 electrodos en la región abdominal y 3 en la región torácica, a la cual se aplicaron todos los algoritmos contenidos en una toolbox de MatLab llamada Icalab, que contiene diferentes algoritmos basados en

ICA; de todos los algoritmos aplicados, JADEop entregó los mejores resultados [13]; lo que determinó la aplicación directa de este algoritmo a una base de datos propia.

El procedimiento se aplicó posteriormente a una base de datos propia adquirida con un equipo de electrocardiografía estándar computarizada ECG-PS, marca Galix Biomedical Instrumentation, con el cual se adquirieron señales compuestas por tres registros tomados de las extremidades de la madre y seis registros abdominales. La frecuencia de muestreo fue de 1000Hz por canal, la duración 10 segundos y la resolución de 14 bits. La respuesta en frecuencia del equipo es de 0.05Hz a 500Hz @ -3dB. Para la toma de los registros abdominales se emplearon electrodos de succión.

V. Resultados

La base de datos propia contiene las señales adquiridas en uno de los estudios tomados con el Electrocardiógrafo ECG-PS que son mostradas en la Figura 2, los tres primeros registros contienen las derivaciones tomadas de las extremidades de la madre y los restantes contienen los seis registros obtenidos por los electrodos de succión sobre la región abdominal. Mediante el software del equipo, en el momento de la adquisición se aplicó un filtro pasa-bajos FIR con frecuencia de corte en 150Hz.

Los datos fueron exportados desde el programa del Electrocardiógrafo ECG-PS al software para el análisis de señales ECG para procesarlos con la metodología propuesta.

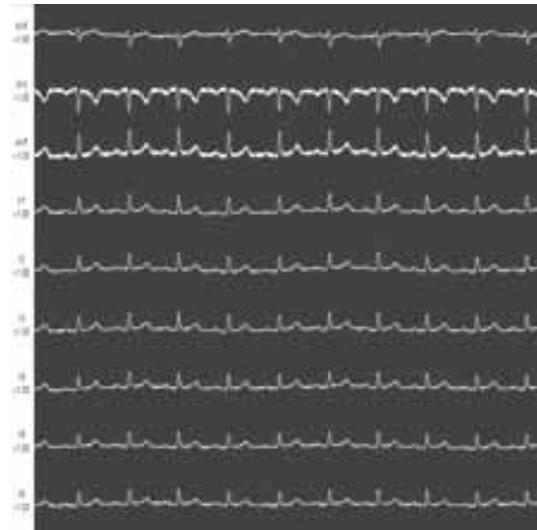


Fig 2. Estudio realizado con el ECG-PS. Las primeras tres señales son tomadas en las extremidades de la madre, las restantes son los registros abdominales.

A las señales importadas en el software, se les aplicó un filtro elimina-banda de 60Hz para remover el ruido causado por la red eléctrica y un filtro pasa-altos de 0.5Hz de frecuencia de corte para remover la línea de base. Luego del preprocesado se ejecuta el algoritmo JADEop para obtener las

fuentes estimadas. En la Figura 3 se muestran los resultados luego de aplicar nuevamente el procedimiento para cancelación de la línea de base.

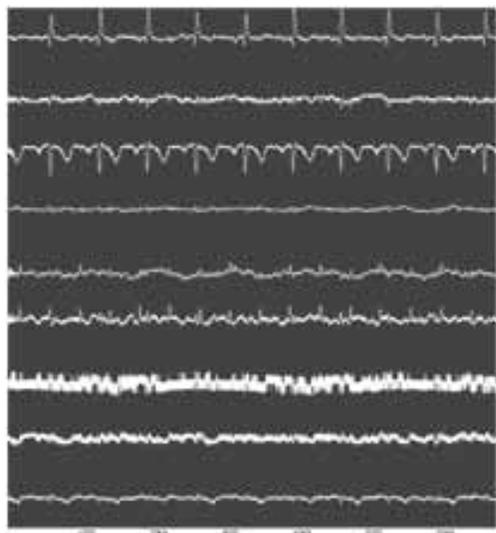


Fig 3. Fuentes estimadas con el algoritmo JADEop y cancelación de la línea de base.

Ya con las fuentes procesadas y separadas, se selecciona la componente fetal de acuerdo a los siguientes dos criterios: primero la distancia entre las ondas R consecutivas de las fuentes estimadas para el FECG debe ser mayor a 100 pulsaciones por minuto (bpm); el segundo criterio es la selección de la señal con menos perturbaciones ocasionadas por ruido.

De arriba hacia abajo, la tercera fuente contiene el MECG y la sexta fuente contiene el FECG. Las tasas cardíacas materna y fetal son 91bpm y 151bpm respectivamente.

La Figura 4 muestra un segmento del resultado de un monitoreo fetal realizado a la paciente. Se observa que la tasa cardíaca fetal oscila entre 140bpm y 160bpm con un valor medio de 150bpm, con lo cual se validan los resultados obtenidos de la aplicación del algoritmo JADEop planteado en la metodología propuesta.



Fig 4. Detalle del monitoreo fetal, mostrando una tasa cardíaca fetal de aproximadamente 150bpm

VI CONCLUSIONES

El desarrollo de una interfaz gráfica de usuario que permite analizar señales electrocardiográficas, es la aplicación de tratamiento de señales biológicas que nos ha permitido la implementación de diversos algoritmos y la visualización de señales, para ofrecer una herramienta más completa a la hora de entregar un diagnóstico médico, sobre la actividad cardíaca fetal.

La aplicación del algoritmo JADEop basado en ICA ha entregado resultados certeros para la identificación del FECG. El procesamiento de los datos de las fuentes estimadas obtenidas por el algoritmo JADEop permitió determinar la FHR, que de otra manera ha sido validada por los resultados de un monitoreo fetal realizado bajo las condiciones médicas y que es el mecanismo actualmente utilizado por los especialistas en el área de medicina para medir la FHR, y entregar los diagnósticos respectivos para la aplicación de los tratamientos.

VII REFERENCIAS

- [1] Deam, "The use of fetal electrocardiogram," *J. Obstet. Gynecol.*, pp. 101:9-17, Jan. 1994.
- [2] Suzanna M M Martens, Chiara Rabotti, Massimo Mischi and Rob J Sluijter, "A robust fetal ECG detection method for abdominal recordings", *Physiological Measurement*, March 2007.
- [3] V. Vigneron, A. Paraschiv-ionescu, A. Azancot, O. Sibony, C. Jutten, "Fetal Electrocardiogram extraction based on Non-stationary ICA and wavelet denoising", *IEEE* 2003.
- [4] E. R. Ferrara and B. Widrow, "Fetal electrocardiogram enhancement by time-sequenced adaptive filtering," *IEEE Trans. Biomed. Eng.*, vol. BME-29, pp. 458-460, 1982.
- [5] S. Abboud, A. Alaluf, S. Einav, and D. Sadeh, "Real time abdominal fetal ECG recording using hardware correlator," *Comput. Biol. Med.*, vol. 22, pp. 32-335, 1992.
- [6] D. Callaerts, B. D. Moor, J. Vandewalle, and W. Sansen, "Comparison of SVD methods to extract the fetal electrocardiogram from coetaneous electrode signals," *Med. Biol. Eng. Comput.*, vol. 28, pp. 217-224, 1990.
- [7] G. Camps, M. Martínez, and E. Soria, "Fetal ECG Extraction using an FIR Neural Network," in *Computers in Cardiology*, 2001, pp. 249-252.
- [8] G. Camps-Valls, M. Martínez-Sober, E. Soria-Olivas, R. Magdalena-Benedito, J. Calpe-Maravilla, and J. Guerrero-Martínez, "Foetal ECG recovery using dynamic neural networks," *Artif. Intell. Med.*, vol. 31, no. 3, pp. 197-209, 2004.
- [9] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "Fetal electrocardiogram extraction by source subspace separation," in *Proc. IEEE SP/ATHOS Workshop HOS*, Girona, Spain, June 1995, pp. 134-138.

- [10] V. Zarzoso, A. K. Nandi, and E. Bacharakis, "Maternal and fetal ECG separation using blind source separation methods," *MA J. Math. App. Med. Biol.*, vol. 14(, no. 3), 1997.
- [11] L. Tong, S. Yu, Y. Inouye and R. Liu, "A Necessary and Sufficient Condition of Blind Identification", International Signal Processing Workshop on Higher-Order Statistics, Chamrousse, France, July 1991, pp. 261-264.
- [12] J. Vanderschoot, D. Callaerts, W. Sansen, J. Vandewalle, G. Vantrappen and J. Janssens, "Two methods for optimal MECG elimination and FECG detection from skin electrode signals", *IEEE Transactions on Biomedical Engineering*, Vol. BME-34, N° 3, March 1987, pp 233-243.
- [13] Irina Vélez López, Jhoan Samuel Romero Peña, Msc. Luis Omar Sarmiento Alvarez, Dr. Juan Carlos Otero Pinto, Dr. Francisco Manrique, "Extracción del Electrocardiograma Fetal a partir del ECG superficial empleando Separación Ciega de Fuentes", Agosto 2007.
- [14] K.V. K. Ananthanag, J. S. Sahambi, "Investigation of Blind Source Separation Methods for extraction of fetal ECG," Dept. of Electronics & Comm. Engg. 1. I.T. Guwahati, Assam, India, IEEE 2003.
- [15] P. Comon, "Independent Component Analysis, A new concept?", *Signal Processing*, Vol. 36, N° 3, April 1994, pp 287-314.
- [16] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," in *Pmc. IEE - F*, vol. 140, pp. 362-370, Dec. 1993.
- [17] De Moor B *DaISy: Database for the Identification of Systems* (Belgium: Department of Electrical Engineering, ESAT/SISTA, K U Leuven) <http://www.esat.kuleuven.ac.be/sista/daisy>.

Análisis espectral para detección de fallas en motores diesel serie ISC (Tracto camiones)

J. M. Salazar**, M.S. L. O. Sarmiento*, M.S.P. Guzmán* M.S.H. González**
*Unidad de Investigación en Telecomunicaciones UNITEL **Grupo de Investigación AMSCP

Abstract— Este artículo presenta los resultados obtenidos en el análisis espectral de señales sonoras en Tracto-Camiones con motores Diesel para la detección de fallos que pueden afectar el correcto funcionamiento, aun así este análisis se extiende para que funcione en cualquier tipo de motor para diferentes vehículos.

I. INTRODUCCION

Anteriormente se implementaron métodos para la detección de fallos donde el mecánico por medio de su oído interpretaba el problema (este método aun se emplea en la actualidad), con el paso de los años la tecnología evolucionó y se comenzaron a usar sistemas electrónicos que permiten al usuario por medio de un tablero identificar fallos mínimos como temperatura del motor, combustible y otros, pero aun no se desarrollaban sistemas que permitieran al usuario indicar problemas directamente del motor, hasta que se empezaron a utilizar las computadoras abordo y sistemas electrónicos (sensórica encargada de detectar y enviar datos al computador abordo), donde las computadoras son las encargadas de leer la información de cada sensor, calculando presiones, temperaturas y así indicando en un tablero de control un numero de errores que para el usuario es complicado interpretar, por lo que es necesario diseñar un software encargado de interpretar estos errores, para esto se requiere conectar un computador externo al computador del vehículo que indica de una forma visual la pieza o el sitio donde se encuentra el problema, pero este método no siempre es eficiente, existen problemas y fallas que no siempre el computador interpreta.

El análisis espectral de señales sonoras se ha llevado a cabo desde las primeras etapas de desarrollo de los motores, en 1931 se encontró una descripción que relacionaba el aumento de presión producida por el combustible y el ruido producido [1, 2], con el tiempo se desarrollaron ecuaciones que permitían analizar las señales sonoras como la transformada de Fourier y el periodograma de WELCH que permiten identificar en una señal sonora las frecuencias fundamentales, eliminando parte del ruido y tomando las frecuencias más importantes de la señal. Estas frecuencias fundamentales representan el funcionamiento del motor, que de acuerdo al

presento trabajo, debe variar dependiendo de cómo se encuentre el estado del mismo, figura 1.

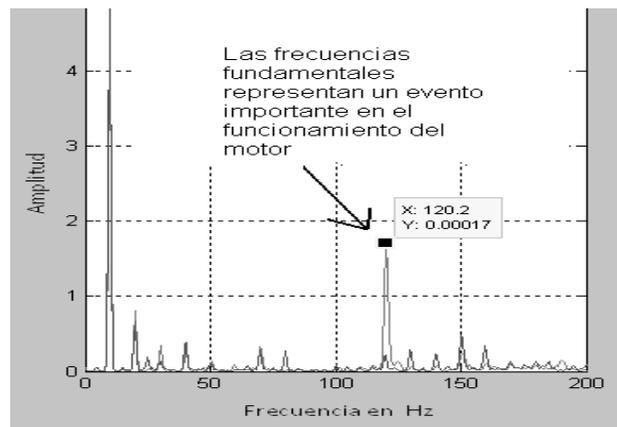


Figura. 1 Señal sonora donde se define un evento en una diferente frecuencia empleando uno de las diferentes técnicas para analizar espectros en frecuencia.

II. DESCRIPCIÓN DE LAS FUENTES DE RUIDO DE LOS MOTORES

Es bien aceptado que las principales fuentes de ruido en los motores de combustión interna incluyen las siguientes categorías: combustión (relacionada con el proceso de movimientos mecánicos), admisión, escape, e impactos [1, 3, 4,5].

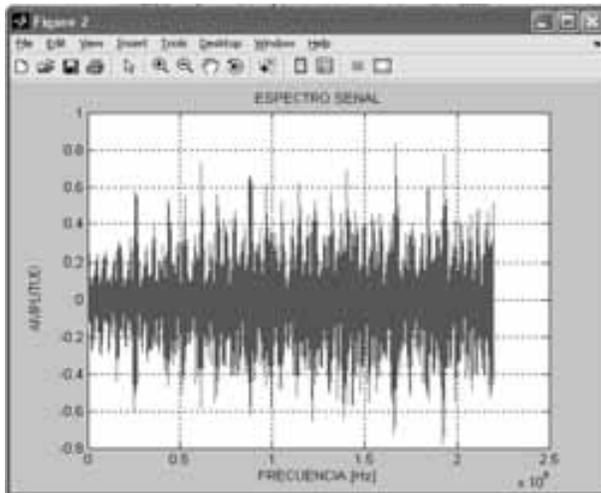


Figura. 2 Espectro de una señal sonora con ruido adquirida del motor

Así mismo, los motores diesel de 6 pistones, presentan esas mismas fuentes de radiación de ruido: la combustión, válvulas, pistones e impactos. Estas fuentes son de corta duración y se producen en relación con la posición angular del cigüeñal, figura. 2. Esta señal adquirida no es útil al momento de analizar y comparar con otras señales puesto que tiene altos contenidos de ruido, pero sí es posible filtrar, y procesar espectralmente para obtener un mejor resultado y así poder comparar con los demás datos almacenados. Estas señales acústicas se miden a partir de las pruebas realizadas al motor.

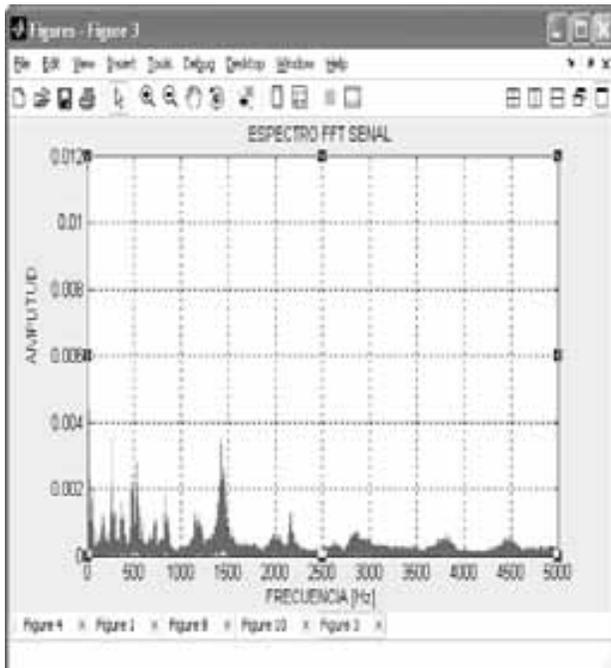


Figura. 3 Tratamiento de la señal adquirida con la transformada de Fourier FFT

La Transformada de Fourier es la representación en frecuencia de la señal adquirida, básicamente se encarga de transformar una señal del dominio del tiempo, al dominio de la frecuencia. El trabajo con la señal en frecuencia, no solo sirve como información, sino que se puede modificar, de forma que es ampliamente utilizada en filtros, procesamiento de imágenes y sonido [6, 7]. La figura 3, muestra la representación espectral de la señal de la figura 2. Para evitar efecto del ruido en la representación espectral de la señal, se emplea el periodograma de Welch.

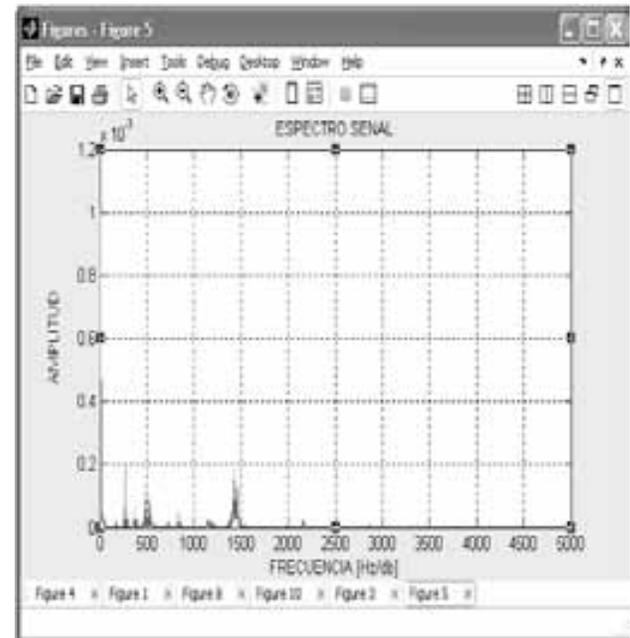


Figura. 4 Tratamiento de la señal adquirida con PWELCH

Esta técnica permite disminuir la varianza propia de las señales con ruido [7, 8, 9], se obtiene un mejor resultado en el manejo de las señales, puesto que la función permite implementar los diferentes tipos de ventana (hamming, hanning), también tiene las opciones de indicar el número total de secciones en las que se dividirá la señal a analizar. La figura 4 muestra el periodograma de la señal de la figura 2.

III. ANALISIS EN MOTORES DIESEL DE 6 PISTONES SERIE ISC.

- A. *Adquisición de datos.* Para la adquisición de se empleó la tarjeta multimedia de un computador portátil HP Pavilion ZV5200 con un frecuencia de muestreo de 44100 Hz. Esta tarjeta aplica el filtro antialiasing respectivo.
- B. *Características del automotor.*

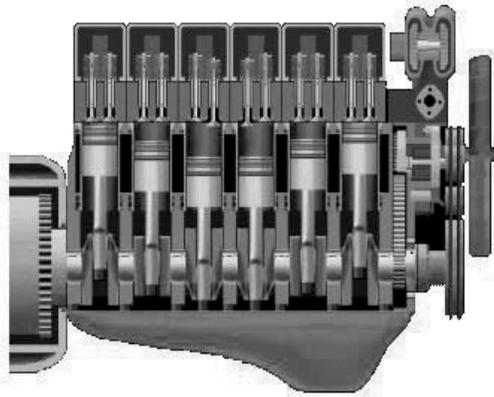


Figura. 5 Motor diesel de 6 pistones serie ISC. Tomada de <http://img233.imageshack.us/img233/8540/cummins5fj0.jpg>

Para definir los diferentes problemas en un motor de combustión interna se ha especificado el funcionamiento y análisis de un motor DIESEL Cummings de la serie 9400i, las muestras se tomaron a velocidad constante durante 5 segundos, en este momento el motor internamente gira a 600 rpm. Figura .5.

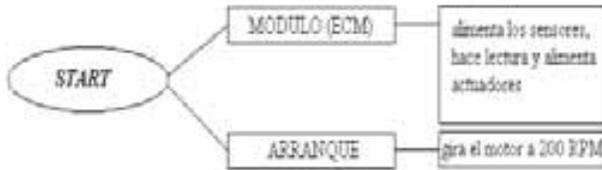


Figura. 6 Esquema del encendido del motor.

El mecanismo de combustión de un motor DIESEL comienza por darle *start* al motor e inmediatamente este envía una señal al módulo o computador del vehículo y esta a su vez alimenta los sensores, hace lectura de lo que ocurre en el motor y alimenta los actuadores, e inmediatamente que alimenta el módulo enciende el arranque y el motor gira ente 200 y 600 rpm. El módulo de control electrónico (ECM) se muestra Fig.6.

Los motores probados fueron analizados previamente por el computador de abordaje el cual no detectó ningún problema, pero en la cabina del conductor se sentían vibraciones de origen desconocido y/o problemas de encendido. Tres de estos casos se reportan en este estudio.

Es de aclarar que el computador de abordaje requiere a su vez de un computador externo con un hardware y software especial disponible únicamente para la casa matriz. Esta posibilidad de análisis es empleada por los usuarios dentro del periodo de garantía, típicamente de 2 años, pero generalmente después de terminada la garantía los

conductores no recurren a este método por los altos costos que implica.

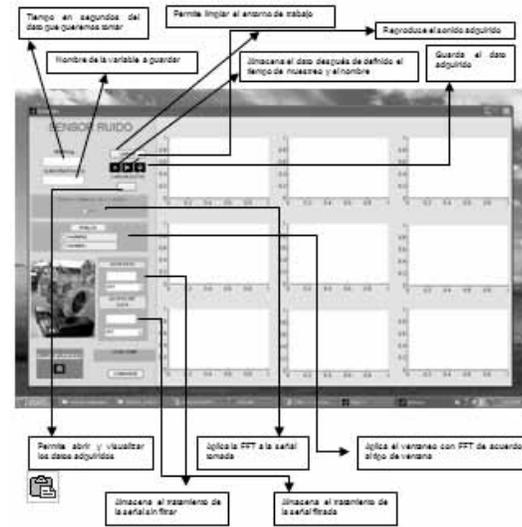


Figura. 7 Interfaz grafica para lectura de sonido del motor.

C. Interfaz gráfica.

En el análisis y adquisición de señales sonoras se diseñó una interfaz gráfica que permite adquirir, comparar, y filtrar las diferentes señales tomadas como se muestra en la figura 7.

Etapa de Adquisición

El proceso de adquisición inicia instalando un micrófono de la multimedia en el orificio de ingreso del aceite al motor para minimizar el ruido externo.

Los datos fueron obtenidos de la siguiente forma: Se analizaron en total 3 motores de la misma serie, bajo las mismas condiciones de operación. Dos de ellos en su estado inicial donde presentaban una falla en la bomba de combustión (Motor 1 y Motor 2) y posteriormente una vez realizada su reparación y un tercer motor (Motor 3) en buen estado de funcionamiento el cual es comparado con los que presentaban la falla. Se debe tener en cuenta que para que exista un patrón los motores o vehículos solo pueden tener un tipo de daño).

IV. ANÁLISIS DE RESULTADOS

Para el análisis de los resultados se plantea la siguiente tabla de comparaciones

TABLA I
COMPARACIÓN DE ESPECTROS DE LAS SEÑALES ADQUIRIDAS EN FRECUENCIA

Señal Roja	Señal Azul
Motor 1 con falla	Motor 1 Reparado
Motos 2 con falla	Motor 2 Reparado
Motor 2 con fallo	Motor 3 Reparado

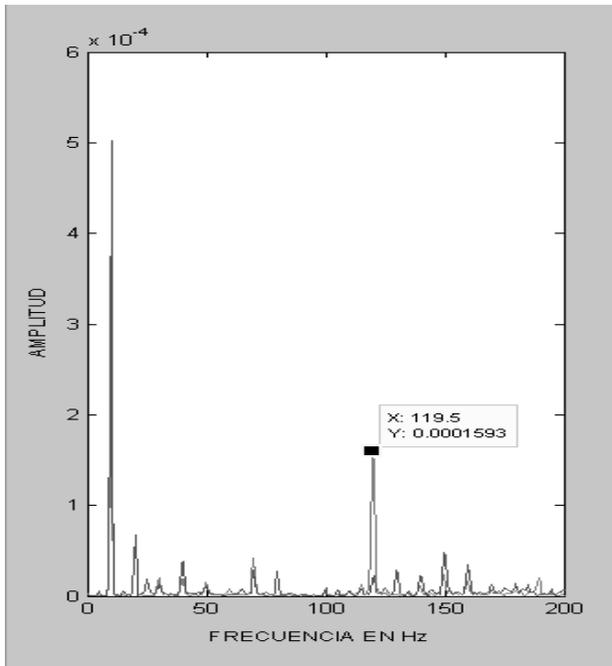


Figura 8. Comparación del espectro cuando el motor 1 esta reparado y Cuando el motor 1 tiene el problema.

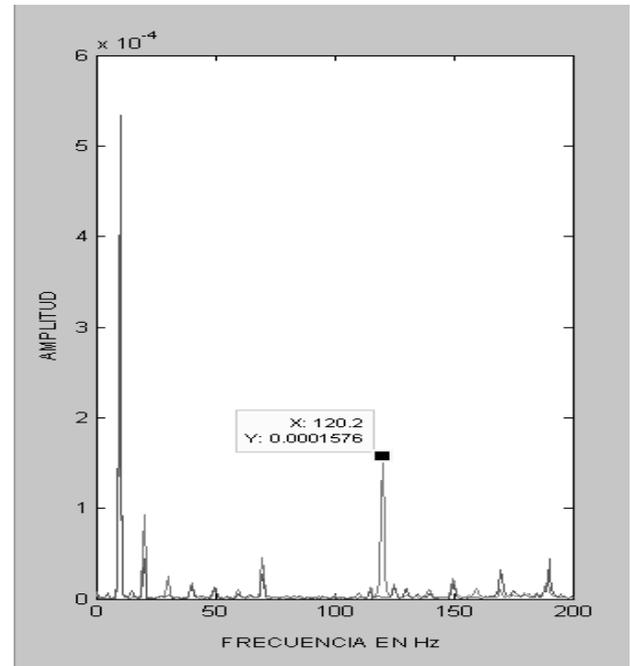


Figura 10. Comparación del espectro cuando el motor 3 esta reparado y Cuando el motor 2 tiene el problema.

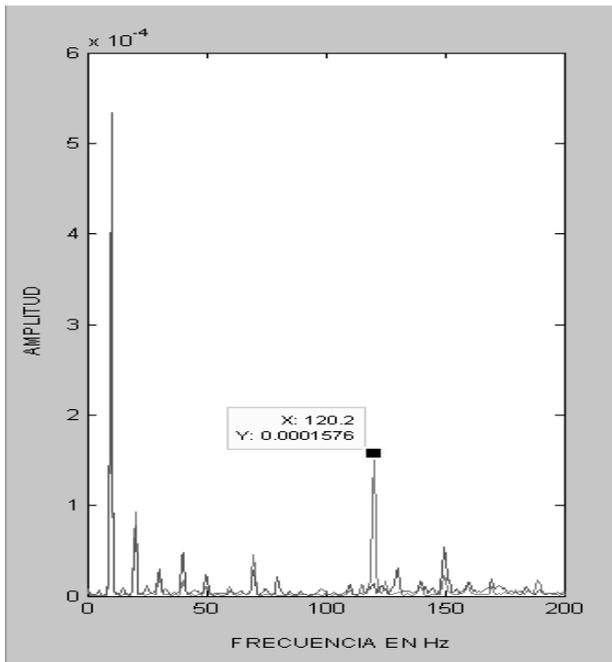


Figura 9. Comparación del espectro cuando el motor 2 esta reparado y Cuando el motor 2 tiene el problema.

Para la interpretación de los resultados tendremos en cuenta que el color azul representa la señal producida por el motor reparado y el color rojo la señal producida por el motor con daño que para este caso es la fuga de combustible.

Como interpretación de las gráficas en figuras (8, 9, 10), se puede concluir que cuando ocurrió el problema se sitúa un pico de magnitud [0.0001576, 0.0001600] a una frecuencia entre [118, a 122] Hz y cuando se repara (cambio de la bomba de combustible) este desaparece, en este caso el problema era una fuga de combustible por la bomba eléctrica y siempre se caracteriza como una vibración y sonidos diferentes en el motor que reflejan el mal funcionamiento del mismo.

En los 3 diferentes espectros podemos concluir que es posible que ocurra un patrón en estos motores, siempre y cuando la diferencia de cada uno de los vehículos por uso, no sea mayor a 2 años. Ya que según las pruebas realizadas con algunos motores de mayor antigüedad, se observa una variación en el patrón que puede ser atribuida al desgaste o sedimento de las piezas.

V.CONCLUSIONES

Muchas veces los computadores a bordo del vehículo no detectan fallas que involucren los sistemas mecánicos o variaciones en los cambios internos del motor, que se pueden manifestar con sonidos o vibraciones. Por esta razón la persona encargada de la reparación debe recurrir a una inspección sensorial y a la experiencia previa, con esta

propuesta se muestra que si se implementa un método como el análisis espectral de sonido es posible tener un patrón y resolver el problema, teniendo una base de datos de estas señales sonoras y almacenándolas para luego ser comparados con otros posibles adquisiciones.

Si bien el uso de las dos técnicas FFT y PWELCH proporciona buenos resultados, el método de WELCH como una aplicación de FFT a diferentes ventanas que permite la selección de las amplitudes más importantes únicamente, suprimiendo las señales de ruido que se producen durante la adquisición (micrófono y tarjeta de adquisición) o propias del ambiente de trabajo, presenta mejores resultados que la FFT por sí sola.

Como parte del trabajo se implementó el uso de un filtro digital para el preprocesamiento de los datos, sin embargo, el uso de este tipo de filtros puede implicar pérdida de información importante de la señal por lo que se recomienda, el uso de una etapa de amplificación y filtrado analógico.

Este método funciona mejor en vehículos de menor cilindraje puesto que las perturbaciones no son tan notorias como en los motores de 6 pistones.

Durante la adquisición es importante tener en cuenta y mantener constante el ambiente donde se realizaran las pruebas, ya sea un espacio abierto o cerrado, por que las señales sonoras adquiridas pueden verse afectadas por el ruido del entorno.

Al procesar las señales con el método WELCH se debe tener en cuenta la utilización de las ventanas, se recomienda seleccionar y emplear un solo tipo de ventana ya que las características de cada una pueden llevar a variaciones en los datos.

Agradecimientos.

Los autores de este trabajo agradecen a la empresa TRYENERGI, encargada del mantenimiento y reparación a nivel nacional de motores Cummins, por permitir el acceso a los motores y permitir la toma de datos. .

BIBLIOGRAFIA

[1]. T. Priede 1980 *sae* 800534, 2039}2069." In search of origins of engine noise*an historical review".

[2]. P. W. Schaberg, t. Priede and r. K. Dutkiewicz 1990 *sae* 900013, 1}13. "Effects of rapid pressure rise on engine vibration and noise".

[3]. B. J. Challen and d. M. Croker 1982 *sae* 820517, 1}32. "A review of recent progress in diesel engine noise reduction."

[4]. G. Gelle, m. Colas, g. Delaunay, "Blind source separation applied to rotating machines monitoring by acoustical and vibration analysis, *journal of mechanical systems and signal processing*" 14 (3) (2000) 427-442.

[5].x. Liu, R.B. Randall. "Blind source separation of internal combustion engine piston slap from other measured vibration signals. *mechanical systems and signal processing*" 19 (2005) 1196-1208.

[6].[HTTP://IEEEXPLORE.IEEE.ORG/XPL/FREEABS_ALL.JSP?TP=&ARNUMBER=681704&ISNUMBER=14979](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=681704&isnumber=14979)

[7].THE SCIENTIST AND ENGINEER'S GUIDE TO DIGITAL SIGNAL PROCESSING BY STEVEN W. SMITH, PH.D. [HTTP://WWW.DSPGUIDE.COM/CH12/2.HTM](http://www.dspguide.com/ch12/2.htm)

[7] Hayes, M., "Statistical digital signal processing and modeling", John Wiley & Sons, 1996.

[8] Stoica, p., and R.L. Moses, "Introduction to spectral analysis", Prentice-Hall, Englewood cliffs, nj, 1997, pp. 52-54.

[9] Welch, P.D, "The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE® trans. audio electroacoustics*, vol. au-15 (june 1967), pp.70-73.

Autores

José Miguel Salazar estudiante noveno semestre Facultad de Ingeniería Mecatrónica Universidad Santo Tomás, sus áreas de interés son la automatización y control de procesos industriales y agroindustriales.

M.S Luís Omar Sarmiento Álvarez, Ingeniero Electricista. Magíster en Potencia Eléctrica Universidad Industrial de Santander, sus áreas de interés son el procesamiento digital de señales, aplicaciones biomédicas.

M.S Paola Guzmán Castillo. Ingeniero Electrónico. Magíster en Ingeniería Electrónica Universidad Industrial de Santander, su interés investigativo se orienta a las áreas de procesamiento digital de imágenes, gestión de redes de comunicaciones y desarrollo de aplicaciones de telecomunicaciones.

M.S Hernando González. Ingeniero Electrónico. Magíster en Ingeniería Electrónica Universidad Industrial de Santander, interés investigativo en las áreas de control, robótica y energías alternativas.

Sistema para la recuperación de imágenes usando descriptores de forma

Gloria E. Jaramillo, John W. Branch

Abstract -Con la masificación de las Tecnologías de la Información y la Comunicación -TIC- se ha hecho evidente el incremento de la información digital. Dicha información multimedial (texto, imágenes y sonido) es necesaria no sólo almacenarla, sino también acceder a ella de una manera óptima y eficiente. En este artículo, se propone un sistema de recuperación de imágenes de piezas mecánicas y automotrices basándose en su contenido visual. Específicamente, se evalúan técnicas de recuperación de imágenes por contenido de forma basada en regiones (momentos de Zernike y momentos geométricos) y contorno (Curvature Scale Space -CSS). Para la evaluación de dichos algoritmos, se trabajó con una base de datos compuesta por 902 imágenes, agrupadas semánticamente en 19 categorías. Se presenta la precisión y recall como medidas de desempeño de los algoritmos. Los resultados muestran que la forma basada en regiones presenta un rendimiento mucho mayor que el Curvature Scale Space, llegando a un máximo de 98% de recuperación positiva.

Index Terms- Curvature Scale Space, momentos de Zernike, momentos geométricos, Recuperación de Imágenes por Contenido.

I. INTRODUCCIÓN

El análisis de la forma ha sido motivado principalmente por los estudios adelantados sobre el sistema de visión humano. Entre las teorías neuropsicológicas clásicas de percepción de forma se encuentra el trabajo de Hebb[1] basado en la conexión entre el movimiento ocular y las estructuras neuronales, el cual ha contribuido considerablemente a las bases de las redes neuronales artificiales; y los aportes de Wertheimer [2] en los que presenta una lista de leyes para el reconocimiento de los objetos, entre éstas se encuentran la mayoría de los principios establecidos para los sistemas computacionales actuales de reconocimiento de objetos por medio de la forma: cambios en la orientación, traslación, tamaño y color no afecta la identidad de los objetos. Entre las teorías modernas, se encuentran las investigaciones de Koenderik y Van Doorn [3] basadas en la psicología visual, quienes proponen un nuevo modelo dinámico fundamentado en la percepción en varias escalas de resolución. Marr y Hildreth [4] proponen un paradigma para la detección de bordes basado en la identificación de puntos de

inflexión, el cual a diferencia de las teorías clásicas permitía una implementación computacional.

En este contexto, teniendo en cuenta la importancia de la forma para la búsqueda, reconocimiento y clasificación de objetos, los sistemas de Recuperación de Imágenes por Contenido – CBIR (Content-Based Image Retrieval) han apuntado hacia la extracción de características de la forma de los objetos como una métrica clave que permita su reconocimiento. Los sistemas CBIR se basan en el sistema visual humano para asociar las características visuales con el significado semántico, además de perseguir invariabilidad a rotación, traslación y escalado para la identificación de formas.

La siguiente sección presenta las bases teóricas sobre las cuales se fundamentan las tres técnicas de extracción de características propuestas para nuestro sistema: Curvature Scale Space, momentos geométricos y momentos de Zernike. Posteriormente, se presenta el sistema DORIS (Domain – ORiented Image Searcher), su implementación y los resultados obtenidos. Las conclusiones y trabajo futuro son presentados en la sección 4.

II. DESCRIPTORES DE FORMA

Los descriptores de forma se dividen en dos grandes categorías: basados en contorno y basados en regiones. Los descriptores basados en contorno utilizan métricas extraídas de una curva parametrizada que representa el contorno cerrado del objeto; aunque el contorno puede definir satisfactoriamente un objeto y las implementaciones computacionales derivadas de éste tratan de imitar el sistema de visión humano, su gran debilidad se encuentra en la imposibilidad de extraer características para objetos consistentes en varias regiones disjuntas. A esta categoría pertenecen las descripciones basadas en código cadena, B-spline, polígonos, descriptores de Fourier, Curvature Scale Space, perímetro y descriptores de wavelet, entre otros. La segunda categoría está basada en la extracción de las características tomando todos los píxeles que conforman el objeto, superando el inconveniente que presentan los descriptores de contorno; a esta categoría pertenecen principalmente los momentos geométricos, momentos de Zernike y pseudo-Zernike, momentos de Legendre, área y método de Grid.

A. Curvature Scale Space

El Espacio Escalar, propuesto por Asada y Brady [5], consiste en la representación de la forma de un objeto por

G. E. Jaramillo. Universidad Nacional de Colombia (e-mail: gejarami@unalmed.edu.co).

J. W. Branch (PhD en Ingeniería de Sistemas) Universidad Nacional de Colombia. (e-mail: jwbranch@unalmed.edu.co).

medio de segmentos (primitivas) y de una detección sucesiva de puntos de inflexión, a medida que el contorno se suaviza utilizando un filtro paso bajo Gaussiano. Los puntos de inflexión son luego organizados en un árbol de intervalos, cuya interpretación consiste en la localización de los picos de las ramas del árbol comenzando desde las escalas más altas hasta las más bajas. Mokhtarian y Mackworth [6] proponen una modificación del Espacio Escalar, Curvature Scale Space (CSS), en el cual los picos no son interpretados como primitivas de alto nivel sino que son utilizados directamente para comparar dos representaciones de imágenes.

La creación de una representación CSS comienza con la detección de N puntos equidistantes dentro del contorno, comenzando desde una posición aleatoria. La elección de N no obedece a ningún parámetro, sin embargo para asegurar invariabilidad en la resolución se debe tomar el mismo número de puntos para todas las imágenes. El siguiente paso consiste en la medición de la curvatura en cada punto del segmento. Definiendo la curvatura como la derivada del ángulo de la tangente de la curva y creando una parametrización (Γ), la curvatura en el punto i puede ser expresada matemáticamente como

$$k(u) = \dot{x}_i \ddot{y}_i - \ddot{x}_i \dot{y}_i \quad (1)$$

$$\text{donde } \dot{x}_i = \frac{\partial u}{\partial x}; \ddot{x}_i = \frac{\partial^2 u}{\partial x^2}; \dot{y}_i = \frac{\partial u}{\partial y}; \ddot{y}_i = \frac{\partial^2 u}{\partial y^2}$$

Al tener la curvatura de cada segmento, es posible hallar los puntos de inflexión determinando los puntos en los cuales se pasa de una curvatura negativa a una positiva o viceversa. Posteriormente, se procede a suavizar la imagen por medio de filtros Gaussianos donde cada nueva coordenada puede ser representada por medio de las siguientes ecuaciones

$$x_i' = x_i \otimes g(i, \sigma) \quad (2)$$

$$y_i' = y_i \otimes g(i, \sigma) \quad (3)$$

donde \otimes es el producto convolución y $g(i, \sigma)$ representa el kernel gaussiano. Con éstas ecuaciones, la curvatura para cada iteración i se define como

$$K_i = \frac{\dot{x}_i \ddot{y}_i - \ddot{x}_i \dot{y}_i}{(\dot{x}_i^2 + \dot{y}_i^2)^{3/2}} \quad (4)$$

En cada iteración se encuentran los puntos de inflexión correspondientes a cada desviación del filtro Gaussiano, el suavizado termina cuando no se encuentran puntos de inflexión. Al obtener los puntos de inflexión correspondientes a cada iteración se grafican en el plano (u, σ) formando el llamado mapa CSS o imagen binaria CSS donde la coordenada x representa el número de puntos tomados en la imagen original (N) y la coordenada y el número de iteraciones antes de que ésta se vuelva completamente convexa. El número de ramas del mapa CSS indica el número de curvaturas cóncavas en la imagen original.

El último paso para indexar la imagen por medio de la curvatura de su contorno, consiste en hallar los picos del mapa CSS y normalizarlos, estos picos son capaces de describir la

imagen y servir como índice para el proceso de matching. En esta etapa, los picos mayores a un umbral, son localizados, extraídos y organizados en orden descendiente según la longitud del filtro Gaussiano.

B. Momentos Geométricos

Los momentos geométricos, a diferencia de los momentos ortogonales, definen características físicas interpretables, es decir, es posible saber directamente cuáles características están siendo extraídas de la imagen. Los momentos centrales se definen en (5):

$$\gamma(p, q) = \sum_x \sum_y (x - X_c)^p (y - Y_c)^q f(x, y) \quad (5)$$

Los momentos centrales normalizados permiten invariabilidad de escalado normalizando los momentos centrales con la masa de la imagen (6).

$$\eta(p, q) = \frac{\gamma(p, q)}{\gamma(0, 0)^{1 + ((p+q)/2)}} \quad (6)$$

A partir de los momentos centrales normalizados de orden dos y tres se obtienen los siete momentos geométricos invariantes a rotación, traslación y escalado propuestos por Hu [7] en 1962 aplicados a la extracción de características de imágenes (7) - (13).

$$\phi_1 = \eta_{20} + \eta_{02} \quad (7)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4(\eta_{11})^2 \quad (8)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (9)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (10)$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{03} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})] [3(\eta_{03} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (11)$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (12)$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{03} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (13)$$

Finalmente, el vector compuesto por los siete momentos invariantes sirve como índice de la forma. En la implementación de nuestro sistema, se utiliza la mínima distancia euclidiana entre los vectores característicos de dos posibles imágenes similares como criterio de relevancia en la etapa de matching.

C. Momentos de Zernike

Teague [8] en 1980 propone el uso de polinomios ortogonales, en particular los polinomios de Zernike, para la extracción de características visuales, argumentando que estos polinomios mostraban una mayor efectividad en la recuperación y una mayor eficiencia en cuanto a redundancia de información, frente a métodos como la transformada de Fourier-Mellin propuesto por Li [9] como una generalización de los siete momentos invariantes de Hu.

Los momentos de Zernike forman un conjunto de funciones complejas definidas en un círculo unitario, los cuales aplicados al tratamiento de imágenes digitales representan la longitud de la proyección ortogonal de las funciones base a la imagen. Los momentos de Zernike se basan en los polinomios de Zernike definidos en (14), sujeto a $p-|q|$ un número par y $|q| \leq p$; con q un entero positivo, p un entero positivo o cero y (ρ, θ) las coordenadas polares de $(x, y) \in \mathbb{R}^2$.

$$R_{p,q}(\rho) = \sum_{s=0}^{(p-|q|/2)} \frac{(-1)^s (p-s)!}{s! \left(\frac{p+|q|}{2} - s\right)! \left(\frac{p-|q|}{2} - s\right)!} \rho^{p-2s} \quad (14)$$

Las funciones base son calculadas mediante la expresión:

$$V(x, y) = V_{p,q}(\rho \cos \theta, \rho \sin \theta) = R_{p,q}(\rho) \exp(iq\theta) \quad (15)$$

Finalmente, los momentos de Zernike de orden p quedan definidos por (16).

$$Z_{p,q} = \frac{p+1}{\pi} \sum_{x^2+y^2 \leq 1} \sum V_{p,q}^*(x, y) f(x, y) \quad (16)$$

Donde $V_{p,q}^*$ representa la conjugada de $V_{p,q}$.

Debido a que los momentos de Zernike están definidos sobre el círculo $x^2+y^2 \leq 1$ es necesario realizar una transformación de la imagen a coordenadas polares antes del cálculo de los momentos:

$$\rho = \sqrt{x^2 + y^2} \quad (17)$$

$$\theta = \tan^{-1} \left(\frac{y}{x} \right) \quad (18)$$

Donde ρ representa el radio en el punto (x, y) y θ representa el ángulo formado entre el eje de la abscisa y ρ .

Los momentos definidos en (16) no son invariantes a traslación ni escalado, para cumplir estos requerimientos fundamentales de los descriptores de imágenes, es necesario realizar un pre-procesamiento consistente en mover el centro de masa de la imagen (Z_{00}) al centro del círculo unitario, además todas las imágenes deben ser normalizadas a un radio de número fijo de píxeles. Cumpliendo estos requerimientos y sabiendo que la invariabilidad a la rotación se obtiene utilizando solo las magnitudes de los momentos (una rotación de la imagen original solo produce un cambio de fase) se puede calcular los momentos de Zernike con invariabilidad a rotación, traslación y escalado.

$$N_{p,q} = \frac{Z_{p,q}}{Z_{00}} \quad (19)$$

Finalmente, la magnitud de los momentos de Zernike extraídos de la imagen mediante (16) son normalizados (19) y utilizados como índice de la imagen. En la etapa de matching, la medida de similitud corresponde a la distancia Euclidiana entre cada uno de los componentes de vector compuesto por

los momentos de Zernike de la imagen ingresada por el usuario y los almacenados en la base de datos.

III. SISTEMA DORIS (DOMAIN-ORIENTED IMAGE SEARCHER)

El sistema DORIS (Domain-ORiented Image Searcher) es una aplicación construida bajo el lenguaje Java, utilizando como caso de estudio una base de datos compuesta por 902 imágenes de piezas mecánicas y automotrices agrupadas semánticamente en 19 categorías.

DORIS implementa dos tipos de algoritmos para la recuperación de imágenes por contenido basado en forma: momentos de Zernike y momentos geométricos, y un algoritmo para la extracción basada en contorno: CSS. Debido a que la implementación de los polinomios de Zernike requiere realizar varias veces las mismas operaciones y que el costo computacional aumenta debido a las operaciones factoriales, se construyó una Lookup Table (LUT), la cual almacena los polinomios y es consultada para el cálculo de las funciones base. El sistema cuenta con una interfaz gráfica que permite al usuario ingresar la imagen que servirá de petición y un panel donde se presentan las primeras 20 imágenes con un mayor grado de similitud a la imagen petición. Las Fig. 1 y 2 muestran los resultados arrojados por el sistema para dos categorías distintas utilizando los momentos geométricos y momentos de Zernike, respectivamente.

Para calcular la efectividad de recuperación de DORIS cada una de las imágenes fue ingresada al sistema y se utilizó la precisión y recall como medida de desempeño. Para cada uno de los tres algoritmos evaluados se construyó la gráfica de precisión y recall para las 19 categorías de la colección (Fig. 3, 4 y 5).

A partir del análisis de los resultados es posible comprobar la efectividad de los métodos basados en regiones frente a los basados en contorno. Mientras que en los momentos de Zernike se obtuvo una precisión promedio de 75.5%, obteniendo un máximo de 98% de recuperación positiva, el método de contorno presenta una precisión promedio de 23%.

IV. CONCLUSIONES Y TRABAJO FUTURO

En este artículo, se presentó una propuesta para la recuperación de imágenes de piezas mecánicas y automotrices evaluando tres algoritmos de extracción de información visual basados en la forma. La indexación de las imágenes se realizó por medio del cálculo de momentos para los algoritmos basados en regiones y en los puntos de inflexión para el caso del contorno. Los resultados obtenidos con 902 pruebas muestran una mayor precisión en la recuperación de imágenes utilizando los momentos de Zernike sobre los momentos geométricos y CSS. Estos resultados pueden ser justificados por la similitud en la forma que presentan algunos objetos y que no necesariamente corresponden al mismo significado semántico (ver imagen de la fila 4, columna 1 en la Fig. 1 y fila 3, columna 4 de la Fig. 2). Además, los métodos basados en puntos de inflexión no ofrecen una buena definición para contornos de objetos parcial o completamente circulares como

arandelas, rodamientos o tapas de motor. Sin embargo, esta técnica mostró una buena recuperación en imágenes que contenían considerables regiones cóncavas como arranques y motores.

Los resultados que se presentan en este artículo, se derivan de una primera versión del sistema. Como trabajo futuro, se pretende incorporar una recuperación automática que incluya descriptores de alto nivel que representen el contenido semántico, uniendo la brecha existente entre las descripciones de alto y bajo nivel. Asimismo, se espera que el sistema DORIS sea llevado a una plataforma cliente-servidor que indexe imágenes provenientes de dominios específicos, para recuperar tanto las imágenes como la página web que las contiene, esto con el objetivo de brindar al usuario final un sistema que recupere tanto contenido visual como textual basado en búsqueda por contenido de imagen.

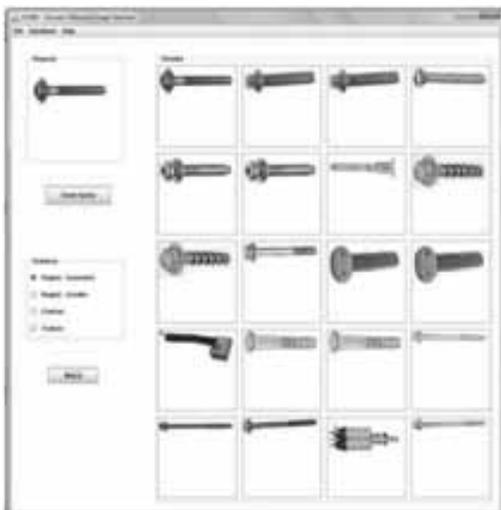


Fig. 1. Resultados utilizando momentos geométricos



Fig. 2. Resultados utilizando momentos de Zernike

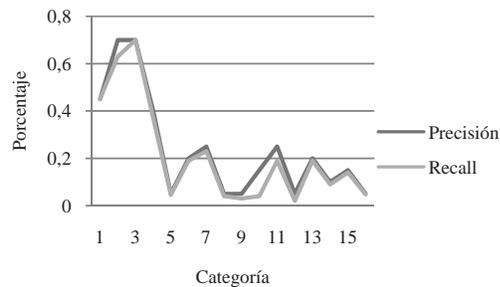


Fig. 3. Precisión y recall obtenido empleando Curvature Scale Space

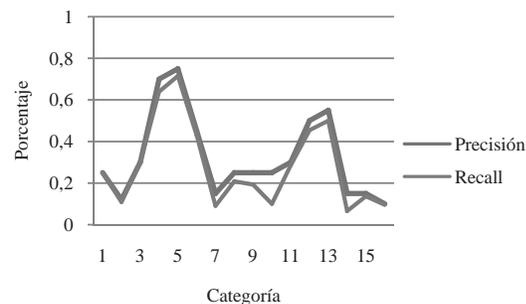


Fig. 4. Precisión y recall obtenido empleando momentos geométricos

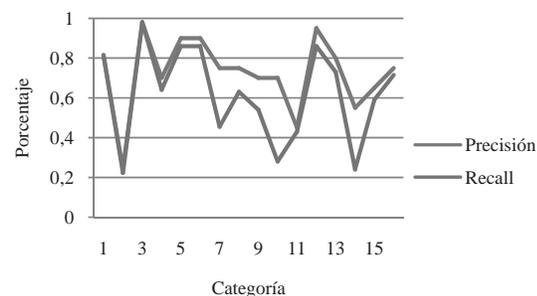


Fig. 5. Precisión y recall obtenido empleando momentos de Zernike

REFERENCIAS

- [1] D. O. Hebb. *The Organization of behavior*. John Wiley, 1949.
- [2] M. Wertheimer. *Laws of Organization in Perceptual Forms*. A Source Book of Gestalt Psychology. Harcourt Brace Jovanovic, 1923.
- [3] J. Koenderink y A. Van Doorn. *Dynamic Shape*. Biological Cybernetics, 53, pp 383-396, 1986.
- [4] D. Marr y E. Hildreth. *Theory of edge detection*. Proceedings of the Royal Society of London, B207, pp 187-217, 1980.
- [5] H. Asada y M. Brady, *The Curvature Primal Sketch*. IEEE Trans. PAMI (6), pp 2-14, 1986.
- [6] F. Mokhtarian y A. Mackworth, *Scale-Based Description and Recognition of Planar Curves and two-dimensional shapes*. IEEE Trans. Pattern Analysis and Machine Intelligence, 8, 1986.
- [7] M. K. Hu, *Visual Pattern Recognition by Moments Invariants*. IEEE Trans. on Information Theory, 8, 1962.

- [8] M. Teague, *Image Analysis via the General Theory of Moments*. J. Opt. Soc. Amer., Vol. 70, No. 8, pp 920-930, 1980.
- [9] Y. Li, *Reforming the Theory of Invariant Moments for Pattern Recognition*. Pattern Recognition Vol. 25, No. 7, pp 723-730, 1992.
- [10] L. J. Latecki, R. Lakämper y U. Eckhardt. *Shape Descriptors for Non-rigid Shapes with a Single Closed Contour*. IEEE Conf. On Computer Vision and Pattern Recognition (CVPR), 424-429, 2000.
- [11] D. Zhang y G. Lu, *A comparative study of curvature scale space and Fourier descriptors for shape based image retrieval*. Journal of Visual Communication & Image Representation. 14, pp 41-60, 2003.
- [12] W.-Y. Kim y Y.-S. Kim, *A region-based shape descriptor using Zernike moments*. Signal Processing Image Communication 16, pp 95-102, 2000.
- [13] J. Peng, W. Yang y Y. Li, *Shape Classification Based on Histogram Representation in Curvature Scale Space*. IEEE Trans. 1-4244-0605 pp 1722-1725, 2006.
- [14] S. Kopf, T. Haenselmann y W. Effelsberg. *Enhancing Curvature Scale Space Features for Robust Shape Classification*. IEEE Trans. 0-7803-9332-5. 2005.
- [15] L. A. Torres-Mendez, J. C. Ruiz-Suárez, L. E. Sucar y g. Gómez. *Translation, Rotation and Scale-Invariant Object Recognition*. IEEE Trans. on Systems, man and Cybernetics, Vol 30, No. 1, pp 125-130, February 2000.
- [16] J. D. Shutler y M. S. Nixon, *Zernike velocity moments for sequence-based description of moving features*. Image and Vision Computing, pp 1-14, 2006.
- [17] S. Loncaric, *A survey of Shapes Analysis Techniques*. Faculty of Electrical Engineering and Computing, Univeristy of Zagreb.
- [18] F. Mokhtarian, S. Abbasi y J. Kittler, *Efficient and Robust Retrieval by Shape Content through Curvature Scale Space*. Department of Electronic and Electrical Engineering, University of Surrey.
- [19] T. Adamek y N. O'Connor, *Efficient Contour-based Shape Representation and Matching*. MIR'03, November 7, 2003, Berkeley, California, USA.
- [20] D. Zhang y G. Lu, *Evaluation of MPEG-7 shape descriptors against other shape descriptors*. Multimedia Systems 9:15-30. Springer-Verlag 2003.

Carpet wear classification using cooccurrence matrices and support vector machines

S. A. Orjuela Vargas, C. Copot, S. Syafie, E. Vansteenkiste, F. Rooms, W.Philips, R. de Keyser, L. Van Langenhove

Abstract— Nowadays, carpet industries have to comply with standard accreditations, requiring labels to certify both the quality and application of their products. Therefore, carpet companies are demanding a reliable, accurate and objective evaluation of the carpet wear level. This paper presents a new approach to analyzing textures on carpet surface data taken using a 3D laser scanner which provides both structural and intensity information. Our aim is to develop and validate a strategy to classify changes in the surface texture on worn carpets. 3D data are resampled on different grid sizes in order to be able to apply 2D image techniques. The features we used are based on Haralick descriptors from the cooccurrence matrix. This paper explains how an optimal experimental design is applied to identify the best combination of sizes and descriptors. Subsequently, the results of a support vector machine classification of the chosen descriptors are shown. With the methodology proposed, we achieve an average of over 90% correct (label) classification.

Index Terms— 3D laser scanner, Wear carpet, Cooccurrence matrix, Haralick's descriptors, Mahalanobis distance, Support vector machine.

I. INTRODUCTION

IN carpet quality evaluation facilities, the conservation of carpet aspects after common wear is one of the most important parameters. To make the appropriate carpet selections, depending of the application, usually labels are assigned using the Carpet and Rug Institute (CRI)'s Performance Standards [1]. This rating only takes into account how the carpet texture changed due to matting and crushing that might occur from walking on it. Years of wear are not associated with the performance ratings. To simulate these wear processes in an accelerated, controlled environment, several laboratory instruments are used, such as "Tetrapod Walker Tester", "Hexapod Drum Tester" and "Vettermann Drum Tester" [1].

There are five standard carpet quality classes, which range from class 1: maximum wear to 5: no wear. The changes in appearance in the carpet due to wear have been classified visually and subjectively by human experts by comparing them to a set of reference standard samples. Usually, at least 3 human experts are involved to classify the carpet level. Each of them may give his/her own grading level depending on a visual inspection and comparison of the examined carpet with the standard one.

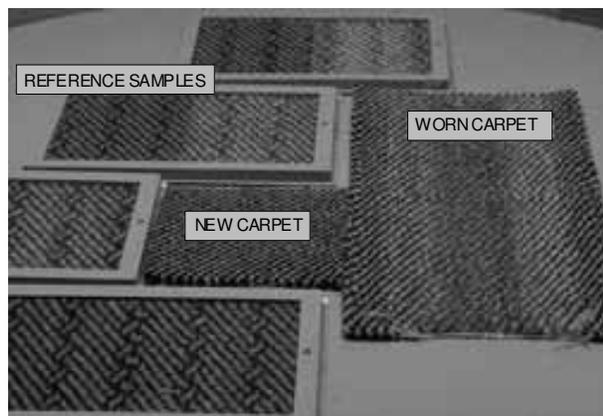


Fig. 1. Test environment used by human experts to classify wear carpets. Source: Textile Department, Ghent University.

Each human expert could give different levels, so the class of the carpet can be in between standard levels. Therefore, the level scales are extended from integer levels to half levels as well. Consequently, human experts frequently disagree in their conclusion as their judgments can be affected by their mood, expertise and other external factors. Fig 1 shows a common line-up used by human experts to classify through visual inspection.

Consequently, carpet companies demand a more reliable, accurate and objective evaluation of the carpet wears level. In the last decades, researches to replace the human expert by automated evaluation methods have been an active topic in academic institutions and industry [2-4]. However, until now, human judgement is still the "best" choice because no qualified automated systems for this task are available yet.

The textile department at Ghent University has been working on a project named COMPAS (Computer-based Assessment of Aspect Change in Carpets due to Wear), in order to develop an

S. A. Orjuela Vargas is supported by a grant of the "LASPAU Academic and Professional programs for the Americas" in agreement with COLCIENCIAS and Universidad Antonio Nariño, Colombia.

S. A. Orjuela Vargas (serealeov@telin.UGent.be), E. Vansteenkiste (ervsteen@telin.UGent.be), F. Rooms (frooms@telin.UGent.be), and W.Philips (Philips@telin.UGent.be) are with the Department of Telecommunications and Information Processing (TELIN), Ghent University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium. Tel: +32 9 264 34 12, Fax: +32 9 264 42 95.

C. Copot (copot@autoctrl.UGent.be), S. Syafie (syafie@autoctrl.UGent.be), and R. de Keyser (rdk@autoctrl.UGent.be), are with the Department of Electrical Energy, Systems and Automation (EESA), Ghent University, Technologiepark 913, 9052 GENT, Belgium.

L. Van Langenhove (Lieve.VanLangenhove@UGent.be) is with de Department of Textiles, Ghent University, Technologiepark 907, 9052 GENT, Belgium.

automated system that allows an objective [5], flexible, accurate, low cost and fast assessment of carpet wear. The goal is to deliver a prototype system that can be developed further for commercial use. Previous research has been focused on either the CCD camera or the 3D laser scanner [5]. Some of the results have been applied successfully to a limited set of carpet samples.

A 3D laser scanner provides both structural (carpet surface) as well as intensity information. Previous research involving this scanner concentrated on depth information only and focused on finding a regression function between the depth and the wear. The validation was performed using nested cross-validation, whereas in our paper the validation is done using the input carpet labels [6].

Moreover, our study presents techniques of automated carpet wear classification of changes in carpet *texture* using the 3D laser scanner, pattern recognition approaches and SVMs. SVMs are a new type of pattern classifier based on a statistical learning technique that has been proposed by Vapnik and his co-workers [7-9].

The paper is organized as follows; a resampling technique from 3D laser data to a regular 2D image grid needed to perform the texture analysis is given in section II. Subsequently, texture feature extraction and carpet rating performances are discussed in section III and IV, respectively. Section V discusses the feature selection of our descriptor set. Finally, a SVM classification approach is presented in Section VI and final remarks are discussed in section VII.

II. 2D IMAGE CONSTRUCTION

During scanning, pieces of wood are used to keep the carpet flat. These pieces are also scanned and need to be removed from the dataset (see Fig. 4 – and comment on the white holes in the carpet there in the caption of the figure). Therefore, 21 random carpets were analysed to define two depth thresholds $\{-112 \mu\text{m}, -99 \mu\text{m}\}$ to separate the pieces of wood from the carpet: scanned carpets are typically between these thresholds, scanned wood is not. The 3D laser scanner produces measurements on a random grid, so these 3D data need to be resampled into a regular grid to construct 2D *images* of them. The way we perform this directly influences the image resolution we will work on.

We empirically found that a sampling of 10×10 cells per square centimetre contains at most two points. In this case, the number of cells containing sample points is on average 30% of the total number of cells. In order to increase this percentage, we also tested different sample schemes (5×5 cells and 2×2 cells per cm^2). Examples of the three schemes are shown in Fig 2.

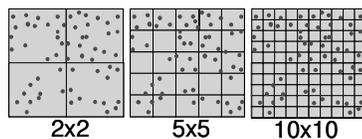


Fig. 2. Example of schemes for resampling 3D laser data.

Next, two resampling methods were used: the first estimates the data in each cell by using the samples in the cells and its 8 neighbours are averaged. In the second, only the sample points in the cell itself are averaged, see Fig 3.

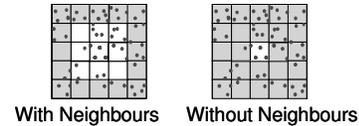


Fig. 3. Filling type methods used for estimate data in cells.

Per data cell, the scanner provides 4 types of measurements: the first two are the mean depth “Z-mean” and median depth “Z-median”. The third one is the median intensity value (named “Q”), and the last one is the density “D”, which represents the number of points either per cell. As a result there are 24 images per carpet to analyse (4 measurement values; each time for both resample techniques and that for 3 different resample sizes: 2×2 , 5×5 and 10×10 cells per cm^2). A set of 10 different carpets types was used in this experiment {A8-501, A8-701, BIG4, 20 KL 803 Beige, LA7, BIG8 Beige, 517, BIG8 Dark, LA9, 20 KL 803 Dark}, one particular example is shown in Fig 4.

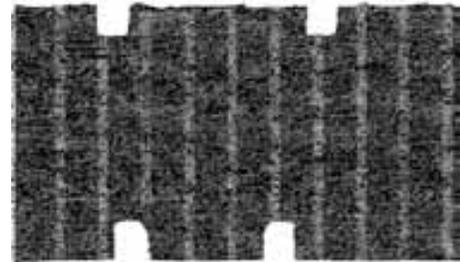


Fig. 4. Representative image of subset {Carpet Type, Measurement, Resample Size, Filling type} = {BIG4, Z-Median, 5×5 , without}.

III. FEATURE DESCRIPTORS

Due to the characteristics of the scanner, the concentration of sample points is higher in the middle of a scan strip than at the borders. Due to the mechanical acceleration wear setup, the wear is concentrated on the middle of the X-axis of the carpet, so the amount of wear typically is only a function of the Y-direction. Therefore, each scan line can be used as a replica for classification with a support vector machine, as will be explained in section VI. How we represent the scan lines, is explained next.

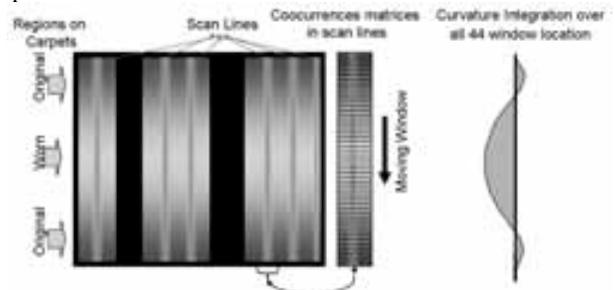


Fig. 5. Computation of Haralick descriptors on scan lines.

First, we extract the regions around the scanned lines (holes are removed with morphological operators). Next, each scan strip is divided into 48 equal regions and cooccurrence matrices [10] are calculated for each (Resample, Filling, Measurement) combination. Subsequently, a window of 5 cooccurrence matrices is defined and moved along the X-axis. Then, 14 Haralick descriptors [11-12] (per combination) are computed for each window location resulting in a set of $24 \times 14 \times 44$ descriptors per scan line. This technique is illustrated in Fig.5

Finally, the curvature obtained from plotting each of the 24×14 descriptors is computed integrating over all 44 window location. This technique is illustrated in Fig 5.

IV. CARPET RATINGS PERFORMANCE (CRP) CRITERION

To unambiguously characterize the status of a given carpet, the estimated labels must be a monotonic function (either increasing or decreasing) of the descriptors. To verify this, the following error criterion is computed.

For each replica (Carpet Type, Measurement, Filling Type, Resample Size), we have a set of 14 descriptors which by Principal Component Analysis (PCA) can be reduced to 3 descriptors (the 3D projection of the 14D space still contains 95% of variation) [13]. Thus, per label, we obtain a 3D cluster of sample points (PC_1, PC_2, PC_3). We then compute the Mahalanobis distance between the current cluster and the total set of all other clusters [14], [15].

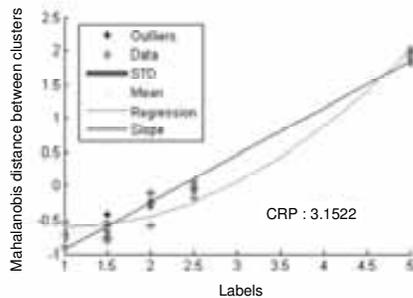


Fig. 6. Carpet Rating Performance (CRP) for subset {Carpet, Measurement, Resample Size, Filling type} = {701, Z-Median, 5x5, without}.

We assume that in the ideal case, the relationship between the labels and the Mahalanobis distances would be linear with a slope of about 45° (i.e. resulting in a linear 1-1 relationship). We first fit a second order regression function (constrained to be either monotonically increasing or decreasing) to the mean values per label over all tested carpets. As a performance measure for our carpet rating method, we computed the ratio between the square error of the regression function of $\sin(2\theta + \pi/4)$, where θ is the slope of the first order regression line (this function is 0 for 0° and 90° and has a maximum at 45°). The lower this ratio, the better the performance of our method is. An example is shown in Fig. 6.

V. FEATURE SELECTION

The performance measure we described in the previous section is used to compare combinations of {Carpets Type, Measurement, Filling Type, Resample Size} [15].

From Fig. 7, we noticed that for a significance level of 0.05, there are significant differences on the responses of the Measurements. So, we focused on a comparison of the measurements.

From Fig 8, it can be seen that the Z-median and the Z-mean scores have means significantly different from Q and D values. Consequently, further comparisons are done using only Z-median.

TABLE I
DESIGN FOR ANALYSIS OF VARIANCE

Carpet Type	Measurement	Filling Type	Resample Size	Replicas	CRP
1	1	1	1	1	X
1	1	1	1	2	X
1	1	1	1	3	X
1	1	1	1	4	X
1	1	1	1	5	X
.
.
.
10	4	2	3	5	X

Source	Sum Sq	d.f.	Mean Sq	F	Prob > F
Resample Size	196.7	2	98.33	0.11	0.8983
Filling Type	423.2	1	423.21	0.7	0.4032
Measurement	17127.6	3	5709.19	2.43	0.0693
Resample Size*Filling Type	894.1	2	447.02	0.55	0.5756
Resample Size*Measurement	9447	6	1574.5	1.88	0.1827
Filling Type*Measurement	1030.7	3	343.55	0.39	0.7627
Error	197065.6	222	887.68		
Total	225674.6	228			

Fig. 7. Result of the analysis of variance for the experiment.

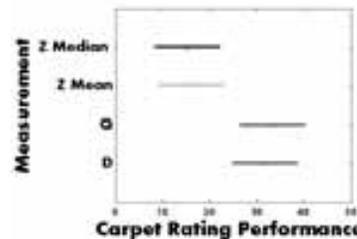


Fig. 8. Comparison of Measurement data.

In Fig. 9, we show there are no significant differences between filling and no filling and neither between the three different resample sizes. Therefore, our choice is based on minimal means factor for Filling Type (Without), and resample size 5x5 because it has minimal changes comparing with the original optimal resample size 10x10.

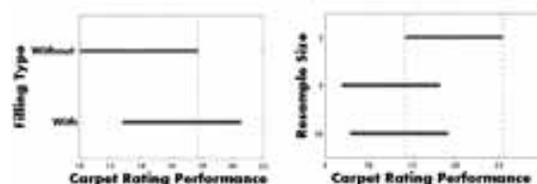


Fig. 9. Comparison of Filling Type and Resample Size.

VI. CLASSIFICATION

In this study two different kernel SVM functions are used to analyze the performance of the classification. For both cases, the classifier is individually trained to 10 types of carpets. Each carpet has a different class distribution. The three first principal components from the chosen combination are used for the training input. Every class of the carpet has 5 replicas, of which 4 replicas are used for training and 1 replica is used for testing.

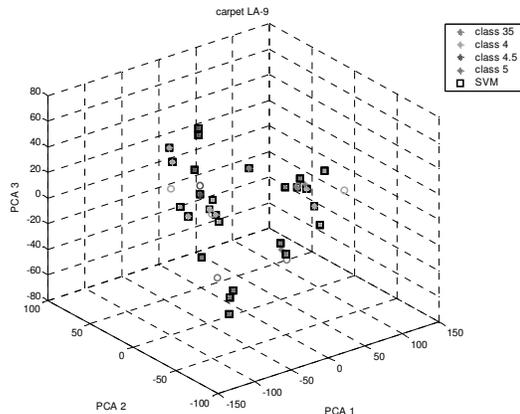


Fig. 10. Training SVM with Gaussian kernel for carpet LA-9.

SVMs aim at minimizing an upper bound of the generalization error through maximizing the margin between the separating hyperplane and the data [16]. SVMs are known to generalize well even in high dimensional spaces under small training sample conditions [17]. Therefore, SVM is applied in this study.

TABLE II
OVERALL CLASSIFICATION USING 2 KERNELS FUNCTIONS

Type of carpet	Polynomial kernel (%)	Gaussian kernel (%)
A8 – 501	84.4	84.4
A8 – 701	84.4	100
Big 4	100	100
20 KL 803 Beige	84.4	84.4
LA 7	84.4	100
Big 8 Beige	100	100
20 KL 517	100	100
Big 8 Dark	86	86
LA 9	72	100
20 KL 803 Dark	100	100
Over all	89.56	95.48

A training SVM is plotted in Fig. 10 for carpet LA-9. This LA-9 carpet has 4 classes plotted in different colors. The SVM classifier is built with these 4 classes of data for training. The classifier is also tested using 1 left replica data for each class of the carpet. Using a polynomial kernel, the machine classified 72% of the given testing data correctly, whilst using a Gaussian kernel, the machine classified 100% of testing data correctly in several cases (see Table II). Overall, the machine classifies 89.56% correctly using a polynomial kernel and 95.48% using a Gaussian kernel. As the previous approach with 3D laser data was studied on a regression fitting between

depth and labels. Therefore, it is not possible to compare our results with that one.

VII. CONCLUSION

In this paper we presented a new methodology for classifying appearance changes on carpet texture from 3D laser scanner data. By using the CRP criterion we proposed in Section IV, we have shown that this methodology is well-suited for dealing with multiple types of carpets. In this paper, we only applied our methodology to texture descriptors based on the cooccurrence matrix, but we believe that our methodology can be used in a more general context to investigate relations between a mathematical description of carpet observation (descriptors based on depth and texture) and carpet quality to achieve a generic automated carpet wear evaluation system.

REFERENCES

- [1] CRI Test Method – 101, Technical Bulletin, The Carpet and Rug Institute, Revision July 2003.
- [2] Wood, E., Hofgson, R., "Carpet texture measurement using image analysis", *Textile Research Journal* 59, 1-12, 1989.
- [3] Poudherimi, B., Xu, B., Nayemouri, A., "Evaluation carpet appearance loss: pile lay orientation", *Textile Research Journal* 64, 130-135, 1994.
- [4] Van Steenlandt, W., Collet, D., Sette, S., Bernam, P., Luning, R., Teze, L., Bohland, H., Schulz, H., "Automatic assesment of carpet wear using image analysis and neural networks", *Textile Research Journal* 66, 55-561, 1996.
- [5] Department of Textiles, Faculty of engineering, Ghent University. Annual Report 2007.
- [6] Waegeman, W., Cottyn, J., Wyns B., Boullart, L., De Baets, B., Van Langenhove, L., Detand, J. "Classifying Carpets based on Laser Scanner data". *Engineering Applications of Artificial Intelligence*, Volume 21, Issue 6, September 2008, Pages 907-918.
- [7] Boser, B., Guyon, I., Vapnik, V. "A training algorithm for optimal margin classifiers". In *Proceedings of Fifth Annual Workshop on Computational Learning Theory*, New York, (1992).
- [8] Cortes, C., Vapnik, V. "Support vector networks, In *Proceedings of Machine Learning*", vol. 20, pp. 273-297, (1995).
- [9] Vapnik, V. "The nature of statistical learning theory", Springer, (1995).
- [10] Haralick, R., Shanmugam, K., Dinstein, Its'Hak. "Textural Features for Image Classification." *IEEE Transactions on Systems, Man, and Cybernetics*. Vol 3 nr 6, Nov. 1973 pp 610-621.
- [11] Haralick R., (May 1979). *Statistical and structural approaches to texture*, *Proceedings of the IEEE*, vol. 67, No.5, pp. 786-804.
- [12] Miyamoto E., T. Merryman, "Fast Calculation of Haralick Texture Features", Technical Report, Carnegie Mellon University, www.ece.cmu.edu/~pueschel/teaching/18-799B-CMU-spring05/material/eizan-tad.pdf, (accessed 2008).
- [13] Jolliffe I. T., *Principal Component analysis*, Second Edition, Springer, Chapters 2,4 , 2002.
- [14] Bar-Hillel Aharon, Tomer Hertz, Noam Shental, Daphna Weinshall "Learning a Mahalanobis Metric from Equivalence Constraints", *Journal of Machine Learning Research* 6, Published June 2005
- [15] Montgomery, Douglas C. *Design and analysis of Experiments*, 5th Edition, John Wiley & Sons, INS, chapters, 1-3.5, 2001.
- [16] Amari, S., Wu, S. "Improving support vector machine classifiers by modifying kernel functions", In *Proceedings of International Conference on Neural Net- works*, 12, pp. 783-789, (1999).
- [17] Jonsson, K., Kittler, J., Matas, Y.P. "Support vector machines for face authentication", *Journal of Image and Vision Computing*, vol. 20, pp. 369-375, (2002).

Numerical reconstruction of spectral reflectance curves for multispectral imaging, combining different techniques and a neural Network

Carlos Andrés Osorio Gómez, Luis Gabriel Valdivieso González, Néstor Alonso Arias Hernández, Jáder Enrique Guerrero Bermúdez.

Abstract— An alternative procedure for the digital reconstruction of spectral reflectance curves of oil painting on canvas, by means multispectral imaging, is presented. The technique is based on a combination of the results obtained by pseudo-inverse, principal component analysis and interpolation, these results are the input to a feed-forward back propagation neural network fitting the values of the curves to a target obtained using a spectrophotometer *Shimadzu UV2401*.

Goodness-of-Fit Coefficient (GFC), absolute mean error (ABE) and spectral Root Mean Squared error (RMS) are the metrics used to evaluate the performance of the procedure proposed.

Index Terms—Artificial neural network, Spectral reflectance curves, Multispectral imaging, Curve fitting.

I. INTRODUCTION

The multi-channel nature of multi-spectral images allows the numerical reconstruction of spectral reflectance curves, which are suitable elements for several tasks, among others, restoration and digital archiving of art painting and historical documents [1], [2]. This reconstruction is equally useful in the objects description in color spaces, mainly that whose study is hard using a spectrophotometer. Obtaining reflectance spectral curves, either through a spectrophotometer or numerical techniques is an essential step towards the full colorimetric characterization of color objects. In addition, the enormous quantity and quality of information minimizes the *metamerism* and reduces ambiguity in the verification of color patterns [3], [4].

Fitting the reflectance of an object using multispectral imaging is essentially a process of sampling and quantization

Manuscript received July 11, 2008.

C. A. Osorio, L. G. Valdivieso and N. A. Arias, post-graduate students, Escuela de Física, Facultad de Ciencias, Universidad Industrial de Santander. (e-mail: andres_uis@yahoo.com)

J. E. Guerrero, professor Escuela de Física, Facultad de Ciencias, Universidad Industrial de Santander. (e-mail: jader@uis.edu.co).

Authors are members of Grupo de Óptica y Tratamiento de Señales, Universidad Industrial de Santander.

in spectral bands [3]. This sampling requires the proper selection of illuminating and optical filters. The information recorded by means imaging devices is the input for different linear numerical procedures, such as, pseudoinverse, Principal Component Analysis (PCA) and cubic spline interpolation. Moreover, non-linear processes such as neural networks and genetic algorithms have been reported [5]. Any procedure for digital reconstruction of spectral reflectance requires the following elements: define the objects of interest, in our case, oil painting on canvas. Generally a database of spectral reflectance curves of the objects of interest is recommended and a system of multispectral imaging according to International Commission on Illumination (ICI).

Recently an algorithm for digital reconstruction of spectral reflectance curves, combining the spectral curves obtained by means several linear techniques has been proposed [6]. This procedure minimizes the standard deviation or variance of the root mean squared spectral error and the colorimetric error using quadratic programming. Based on this idea, we propose to estimate a spectral reflectance curve, combining two or more linear techniques known and a neural network.

This paper is organized as follow: the theoretical framework of the linear procedures above mentioned, is explained in section II. In section III, a feed-forward Back propagation Neural Network (BNN) is described. The characteristics of procedure proposed are given in section IV. Experimental setup is mentioned in section V, and finally the results and validation metrics, Goodness-of-Fit Coefficient (GFC), Absolute Mean Error (ABE) and spectral Root Mean Squared error (RMS) are shown in section VI.

II. THEORETICAL FRAMEWORK

For computation, the spectral quantities are replaced by their sampled version, that is,

$$c_k = \sum_{h=0}^{N-1} r(\lambda_h) W(\lambda_h) \Delta\lambda \quad (1)$$

With $W(\lambda_h) = i(\lambda_h)o(\lambda_h)s(\lambda_h)F_k(\lambda_h)$, which is called spectral sensitivity of k -th channel and where, $i(\lambda_h), o(\lambda_h), s(\lambda_h), F_k(\lambda_h)$ and $r(\lambda_h)$ are respectively the radiant spectral flux of the illuminant, the dispersion of the optical imaging system, the spectral response of the sensor, the spectral transmittance of k -th filter and the reflectance of the sample in the range of wavelengths $[\lambda_0, \lambda_{N-1}]$ [7]. In equation (1), λ_h are uniformly spaced wavelengths covering the visible region of the spectrum, $\lambda_h = \lambda_0 + h\Delta\lambda$, with $\Delta\lambda$ as the wavelength sampling interval. n_k , is the additive noise of k -th channel, with temporal and spatial characteristics (dark current, lack of uniformity in lighting, etc.). Putting $\Delta\lambda = 1$, the equation (1) can be written as a dot product [1],

$$c_k = \mathbf{r}^t \mathbf{W}_k + n_k \quad (2)$$

In equation (2) the superscript t denote transpose. For work conditions with short time acquisition, the temporary noise is diminished by using an average of multiple acquisitions. On the other hand, the spatial noise is attenuated by means of the subtractions of the background of the image. This allows neglecting n_k in the equation (2), becoming

$$c_k = \mathbf{r}^t \mathbf{W}_k \quad (3)$$

Considering k filters, equation (3) can be written in matrix-vector notation,

$$\mathbf{c}_{K \times 1} = \mathbf{\Theta}_{K \times N} \mathbf{r}_{N \times 1} \quad (4)$$

Where $\mathbf{\Theta}_{K \times N}$, is a matrix with rows are formed by \mathbf{W}_k vector.

So that, digital reconstruction of spectral reflectance curves can be seen as determining an operator $\mathbf{Q}_{N \times K}$, which applied on the responses of the imaging sensor produces the reflectance.

$$\mathbf{r}_{N \times 1} = \mathbf{Q}_{N \times K} \mathbf{c}_{K \times 1} \quad (5)$$

There are many techniques to estimate the \mathbf{Q} operator. Calculating the pseudoinverse, $\text{pinv}(\cdot)$ of $\mathbf{C}_{K \times p}$ matrix (equation 6), whose columns are p reflectance spectra, measured previously from patches of oil painting on canvas, with similar characteristics to the spectral reflectance curves which are tried to reconstruct

$$\mathbf{Q}_{N \times K} = \mathbf{R}_{N \times p} \text{pinv}(\mathbf{C})_{p \times K} \quad (6)$$

$\mathbf{R}_{N \times p}$, is a matrix with columns formed by p reflectance

spectra recorded by means a spectrophotometer. Once obtained $\mathbf{Q}_{N \times K}$ operator, $\mathbf{r}_{N \times 1}$ is calculated using equation (5).

Another procedure to reconstruct the spectral reflectance curves requires an orthogonal basis obtained by Principal Component Analysis (PCA) using singular value decomposition (SVD) of data matrix \mathbf{R} . Thus, spectral reflectance can be expressed like a linear combination of $m \leq N$ orthogonal vectors \mathbf{v}_i , of N components [8] – [11].

$$\mathbf{r} = \sum_{i=1}^m \alpha_i \mathbf{v}_i \quad (7)$$

The scalar coefficients α_i in the equation (7) are unknown. They are the components of the vector $\mathbf{r}_{N \times 1}$ in the orthogonal basis $\mathbf{V}_{N \times p}$. In order to determine α_i , it is necessary that the number of channels K and the number of vectors p be the same.

Coefficients for each channel, $\mathbf{a}_{K \times 1} = [\alpha_1, \alpha_2, \dots, \alpha_K]^t$, can be obtained using equation (10):

$$\mathbf{a}_{K \times 1} = \mathbf{V}_{K \times K}^{-1} \mathbf{c}_{K \times 1} \quad (8)$$

Where the vectorial basis $\mathbf{V}_{K \times K}^{-1}$, is formed by sampling \mathbf{v}_i vectors at K wavelengths corresponding to each channel. $\mathbf{c}_{K \times 1}$ is the system reconstruction response respective. So that, spectral reflectance, $\mathbf{r}_{N \times 1}$, is obtained replacing in the equation (7).

Another approach for the reconstruction of spectral reflectance curves is supported on the interpolation techniques [12]. Amongst the most known and employed is cubic-spline interpolation.

III. FEED-FORWARD BACKPROPAGATION NEURAL NETWORK: A BRIEF DESCRIPTION

An Artificial Neural Network (ANN) is a general mathematical computing paradigm that models the operations of biological neural systems [13]. In this work, we use a feed-forward back propagation neural network. This architecture is very known and employed. Developed in 1974 by Werber, Parker and Rumelhart, it exhibit characteristics such as, robustness, easy learning, although the computational cost during its training is relatively high.

This ANN type, applies to models with more than two layers of neurons. So that, it can have one input layer with L neurons, an output layer with M neurons and at least one hidden layer with Q neurons [14].

Learning in a feed-forward *BNN*, has two steps: in the first one, the pattern is introduced to the input layer of *ANN* and it is propagated through each upper layer until an output is generated. By comparison, of the output pattern with the desired output an error signal is computed for each output. In the second step, error signals are transmitted backward from the output layer to each node in the hidden layers that contributes directly to the output. However, each neuron in the hidden layers receive only a portion of the total error signal, based approximately on the relative contribution the neuron made to the original output. Training is done by repeating this process, layer by layer, until each node in the network has received an error signal that describes its relative contribution to the total error. Based on the error signal received, connection weights and bias are then updated by each neuron in order the network to converge toward a state that allows all the training patterns to be encoded [15]-[16].

IV. THE PROPOSED METHOD

Recently has been proposed reconstruction of spectral reflectance curves, by means of a weighted sum of the results from different linear techniques, namely, Wiener estimation, pseudoinverse and *PCA*. The weightings of these techniques are calculated by minimizing the combined standard deviation of both spectral errors and colorimetric errors [6].

Based on in this idea, we proposed training a feed-forward backpropagation neural network. After systematic scan on several architectures, we implement the network with the following characteristics: the input pattern to the *ANN* is a $3 \times N$ matrix, whose rows are the results obtained by *PCA*, pseudoinverse and interpolation, no strict order is required. The network used is 3-85-1, that means, three (3) input neurons with linear transfer function, 85 neurons in the hidden layer using Radial Basis Neuron and one (1) output neuron with linear transfer function.

At once the weights and biases are initialized the training is done. Levenberg-Marquardt algorithm for optimization is used [14]. A total of 113 reflectance curves were used, 98 like a training set and the other ones 15 like a testing set. The training over until average value of the *RMS* between the output desired and output calculated is fulfilled.

V. THE EXPERIMENTAL SETUP

Conforming matrix $\mathbf{R}_{N \times p}$ in the equation (6), p spectral reflectance curves from patches of oil painting on canvas are obtained using the *Shimadzu UV-2401* spectrophotometer with integrating sphere attached. The sample (patch) is compared with a BaSO_4 white standard, and spectrally scanned between 450 and 800 with scanning step 1 ± 0.1 [nm]. Spectrophotometer light source is a halogen lamp of 50 [W].

In order to obtain the multispectral images, the patches of oil painting on canvas were illuminated according to the standard *ICI* viewing geometry, $45^\circ/0^\circ$, using two halogen lamps. The light reflected by the patches is transmitted through seven narrow-band interferential filters with central wavelength at: 480, 515, 550, 580, 600, 636, 650 ± 2 [nm]. The full width half maximum of the filters (FWHM) is 10 ± 2 [nm].

Filters are placed in front of a monochrome CCD camera that records the corresponding response to each filter. The acquired image is digitized by means *Matrox™ Meteor II*, video card. Image size is 640×480 pixels and 8-bit deep resolution. In order to reduce the noise, the record of the camera is the average of 20 captures. An additional processing to improve the uniformity of the illumination was done.

Once $\mathbf{R}_{N \times p}$ and $\mathbf{C}_{K \times p}$ arrays are obtained, the numerical treatment to reconstruct the spectral reflectance curves, is made using software *Matlab™* [17]-[20].

VI. RESULTS

The best results in the reconstruction of spectral reflectance curves according to proposed method, from the testing set of curves, are displayed in figure 1. In table I to IV, performance of the metric are shown.

In order to evaluate the performance of the reconstruction, three metric are used. The root mean square error, defined as:

$$RMS = \frac{1}{N} \sum_{i=1}^N \|r_m(\lambda_i) - r_e(\lambda_i)\|^2 \quad (9)$$

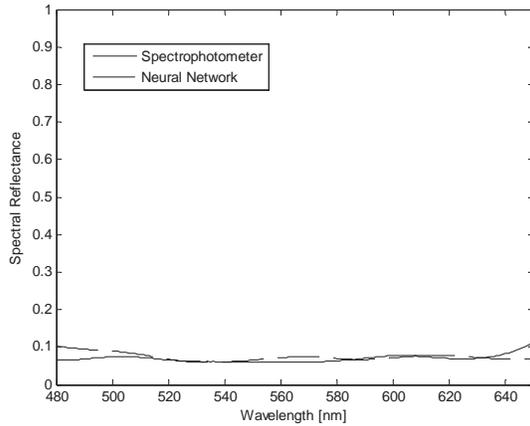
Other metrics are, the absolute mean error,

$$ABE = \frac{1}{N} \sum_{i=1}^N |r_m(\lambda_i) - r_e(\lambda_i)| \quad (10)$$

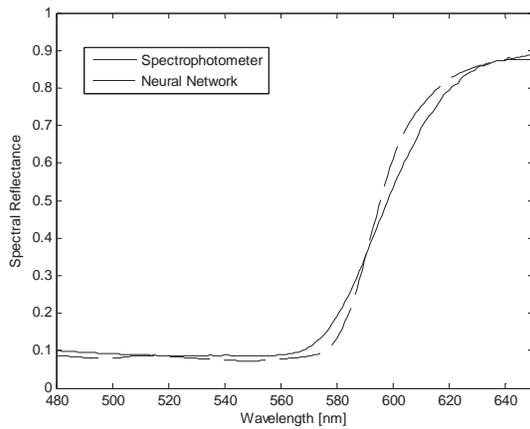
and goodness-of-fit coefficient described by

$$GFC = \frac{\left| \sum_{i=1}^N r_m(\lambda_i) r_e(\lambda_i) \right|}{\sqrt{\sum_{i=1}^N [r_m(\lambda_i)]^2} \sqrt{\sum_{i=1}^N [r_e(\lambda_i)]^2}} \quad (11)$$

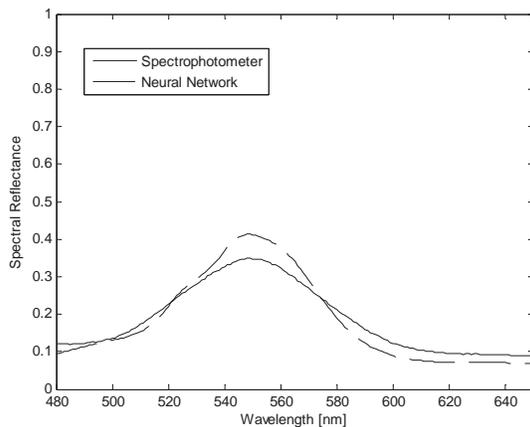
Where in equations (9)-(11), the N -vectors, $r_m(\lambda_i)$ y $r_e(\lambda_i)$ are the spectral reflectance measured by the spectrophotometer and estimated by the neural network (and any other technique) respectively.



(a)



(b)



(c)

Fig. 1. Once trained the neural network, testing set is used to evaluate the performance of the same. (a), (b) and (c) show the best results.

TABLE I
RESULTS OF PERFORMANCE METRICS FOR RECONSTRUCTION BY PROPOSED METHOD

Curve	<i>RMS</i>	<i>ABE</i>	<i>GFC</i>
a	0.0002	0.0097	0.9842
b	0.0009	0.0213	0.9982
c	0.0009	0.0260	0.9909

TABLE II
RESULTS BY PSEUDO-INVERSE

Curve	<i>RMS</i>	<i>ABE</i>	<i>GFC</i>
a	0.0031	0.0483	0.7046
b	0.0024	0.0461	0.9969
c	0.0015	0.0356	0.9820

TABLE III
RESULTS BY INTERPOLATION

Curve	<i>RMS</i>	<i>ABE</i>	<i>GFC</i>
a	0.0068	0.0714	0.9201
b	0.0071	0.0661	0.9877
c	0.0023	0.0403	0.9749

TABLE IV
RESULTS BY PRINCIPAL COMPONENT ANALYSIS

Curve	<i>RMS</i>	<i>ABE</i>	<i>GFC</i>
a	0.0072	0.0682	0.9113
b	0.0082	0.0674	0.9862
c	0.0036	0.0466	0.9638

Table I in comparison with tables II, III and IV, exhibit best performance. In table I, *GFC* results are satisfactory because they exceeded the benchmark of 0.9. However, values more than 0,999 are desirable [1]. In other hand, *RMS* of the proposed method is two order lesser than others metrics.

VII. CONCLUSION

According to results, we find an alternative procedure for digital reconstruction of spectral reflectance curves with satisfactory performance. This procedure is supported by the results obtained using linear techniques very well known. The training of the *ANN* is the key of the method; it requires testing several architectures and evaluates the performance with the appropriate metrics.

In order to improve the results, a supervised learning is recommended. That means, during the training, curves with important spectral contents in different bands could be introduced to the network.

Although the technique was used for digital spectral reconstruction of oil painting on canvas, it can be generalized to any other situation in different fields.

REFERENCES

- [1] A. Ribés, "Multispectral analysis and spectral reflectance reconstruction of art paintings," Ph.D. dissertation, École Nationale Supérieure des Télécommunications", Paris, 2003.
- [2] J. Y. Hardeberg, "Acquisition and reproduction of colour images colorimetric and multispectral approaches," Ph.D. dissertation, École Nationale Supérieure des Télécommunications", Paris, 1999.
- [3] H. J. Trussell, E. Saber and M. Vhrel "Color Image Processing," IEEE Signal Processing Magazine, pp 14-22, (Jan. 2005).
- [4] J.L. Nieves, J.Hernández, E. Valero and J. Romero, " Spectral-reflectance linear models for optical color-pattern recognition," Appl. Opt., vol 43, No. 9, pp 1880-1891, Mar. 2004.
- [5] A. Ribés and F. Schmitt, " A fully automatic method for the reconstruction of spectral reflectance curves by using mixture density networks," Pattern Recognition Letter, vol 24, pp. 1691-1701, 2003.
- [6] H. L. Shen, J. H. Xin and S. J. Shao. (2007, April) "Improved reflectance reconstruction for multispectral imaging by combining different techniques," Optics Express [online], 15(7), pp 5531-5536. Available : <http://www.opticsexpress.org>
- [7] G. Sharma, "Digital Color Imaging Handbook," CRC PRESS Ed., 844 p, New York, 2003.
- [8] H. Haneishi, T. Hasegawa, A. Hosoi, Y. Yokohama, N. Tsumura and Y. Miyake, "System design for accurately estimating the spectral reflectance of art paintings," Appl. Opt., vol 39, No. 35, pp. 6621-6632, Dec. 2000.
- [9] J. Conde, H. Haneishi, M. Yamaguchi, N. Ohya and J. Baez. "Spectral reflectance estimation of ancient mexican codices, multispectral image approach," Rev. Mexicana de Física, vol. 50, No. 5, pp 484-489, Oct. 2004.
- [10] S. Tominaga, "In living color," spie's oe magazine, pp 24-26, Jan. 2003.
- [11] M. Vilaseca, J. Pujol and M. Arjona "Spectral reflectance reconstruction in the near-infrared region by use of conventional charge-coupled-device camera measurements," Appl. Opt., vol 42, No. 10, pp 1789-1797, Apr. 2003.
- [12] J. H. Mathews, K. D. Fink. "Métodos Numéricos con Matlab," 3th ed., Prentice Hall Ed, pp. 304-305, Madrid, 2000.
- [13] Y. H. Hu, J. Hwang, "Handbook of Neural Network Signal Processing," CRC PRESS Ed., 384 p., Boca Raton, 2002.
- [14] H. Demuth, M. Beale, M. Hagan, " Neural Network Toolbox. For use with Matlab," Version 4, The Mathworks, 808 p., Natick MA., 2005.
- [15] J. A. Freeman, D. V. Skapura, "Neural Networks. Algorithms, Applications and Programming Techniques," Addison-Wesley Ed., 401 p., Massachusetts, 1991.
- [16] N. A. Arias, " Sistema de visión artificial para la estimación del peso del ganado bovino," Tesis de Maestría en Física, Universidad Industrial de Santander. Bucaramanga 2004.
- [17] P. D. Burns and R. S. Berns, "Quantization in multispectral color image acquisition," in Proc. IS&T /SID Seventh color imaging conference: Color science, systems and applications, IS&T , 1999, pp 32-35.
- [18] P. Carcagni , A. Della Patria, R. Fontana, M. Greco, M. Mastroianni, M. Materazzi, E. Pampaloni and L. Pezzati, "Multispectral imaging of paintings by optical scanning," Optics and Laser in Engineering, vol 45, pp. 360-367, Jun. 2006.
- [19] H.G. Volz, "Industrial Color Testing: Fundamental and Techniques," 2nd Ed. Wiley-VCH, pp. 243-245, New York, 2001.
- [20] L. G. Valdivieso, J. Guerrero, "Numerical Reconstruction of Spectral Reflectance Curves of Oil Painting on Canvas," in Proc. XII STSIVA, Barranquilla, 2007. Edición electrónica: UNINORTE y Rama estudiantil IEEE Colombia.

Desarrollo del prototipo de un sistema de control y monitoreo para el vuelo de un cohete

A. Mejía andresjosemejia@gmail.com , J. Urrego jonatan_urrego@hotmail.com , Á. Forero alanfova@hotmail.com ,
Dir. F. R. Jiménez L. fjimenez@ustatunja.edu.co Grupo de Investigación GITELCOM, Universidad Santo Tomás – Seccional Tunja

Abstract— This project consists of a control and monitoring of rocket prototype, through a card , It captures longitudinal speed and acceleration in the axis x, y, z, angular speed and acceleration in the shafts Pitch, Yaw and Roll and as the altitude, temperature and position of the rocket respect to its mass center in a plane based on the cardinal points (North, South, East and West). With the acquisition of these variables through a stage of instrumentation, the system is able to process them and deliver them to a control subsystem which aims to rocket position in the desired direction through flaps located in the four wings of the rocket. Besides the prototype RF communication with a PC on earth, where they can monitor variables measured via a control panel designed in MATLAB.

Index Terms— Rocket, Pitch, Roll, Yaw, Aircraft Control, Instrumentation, Communications.

I. INTRODUCCION

ESTA investigación nace de la culminación de una etapa del proyecto "Desarrollo del Prototipo de un Sistema de Control y Monitoreo para el Vuelo de un Cohete" perteneciente a una rama del Grupo de investigación GITELCOM de la Universidad Santo Tomas, seccional Tunja. Como trabajo de investigación aplicada, este proyecto abarca la problemática de diseñar y elaborar un sistema capaz de sensar y monitorear las variables presentes en un cohete en vuelo y así mismo controlar la dirección de éste.

Para lo cual se desarrollaron tres tarjetas que se pueden integrar a cualquier cohete o planta en movimiento.

Se desarrolló el prototipo del sistema de control, monitoreo, protección y recuperación basado en sensores de inclinación,

Artículo escrito el 18 de Junio de 2008. La investigación ha sido soportada por el Grupo de Investigación GITELCOM de la Facultad de Ingeniería Electrónica, Universidad Santo Tomas, Seccional Tunja, Boyacá, Colombia.

F. R. Jiménez López es Docente de la Facultad de Ingeniería Electrónica de la Universidad Santo Tomas de Tunja, Boyacá, Colombia. Sus intereses de investigación incluyen la ingeniería de control, el control digital, la electrónica de potencia, y el Procesamiento Digital de Señales (E-mail: fjimenez@ustatunja.edu.co).

A. J. Mejía. Ingeniero Electrónico, Facultad de Ingeniería Electrónica, Universidad Santo Tomás - Tunja. (E-mail: andresjosemejia@gmail.com).

Jonatan Fabián Urrego. Ingeniero Electrónico, Facultad de Ingeniería Electrónica, Universidad Santo Tomás - Tunja. E-mail: (jonatan_urrego@hotmail.com).

Á. A. Forero, Ingeniero Electrónico, Facultad de Ingeniería Electrónica, Universidad Santo Tomás - Tunja. (E-mail: alanfova@hotmail.com).

giro, aceleración, presión, temperatura y campo magnético de la tierra. Este sistema posiciona al cohete por medio de actuadores para mantener una dirección deseada y garantiza la activación de un sistema de recuperación basado en un paracaídas con el fin de reutilizar el sistema. Este proceso es monitoreado vía RF por medio de un computador en el cual se visualizan las variables presentes en el cohete y permite variar la dirección deseada o activar el sistema de recuperación en cualquier instante de tiempo. Es de suma importancia aclarar que en el proceso de investigación no se contempla la prueba final del vuelo del cohete, sino que se desarrolla el prototipo basado en unas tarjetas de control e instrumentación que cumplen con lo descrito anteriormente.

El sistema de control desarrollado puede ser aplicado para varios proyectos aparte de este, siendo una herramienta versátil, ya que puede monitorear cualquier tipo de sólido estático o en movimiento teniendo en cuenta su aceleración, velocidad, inclinación, rotación, temperatura y altura. Además por el tipo de instrumentación que se implementa, se pueden determinar vibraciones, giros, choques y movimientos fuertes.

II. DESCRIPCIÓN DEL SISTEMA A CONTROLAR

A. Partes de un Cohete

En general son cuatro y se encuentran ubicados como se observa en la Figura 1: carga útil, cámara de instrumentos y/o control, el sistema de propulsión y alerones estabilizadores.



Fig. 1. Partes del Cohete.

1) *Carga Útil*: En este compartimento se coloca el propósito del cohete. Ej.: satélites, instrumentos

metereológicos, paracaídas, entre otros.

2) *Cámara de instrumentos*: Aquí se colocan los sistemas electrónicos de monitoreo, control y comunicaciones necesarios para asegurar que se cumpla el objetivo del vuelo.

3) *Sistema de Propulsión*: Es una parte vital del cohete, ya que proporciona la energía para el movimiento. Este sistema se basa en fundamentos químicos y físicos, los cuales provocan las reacciones necesarias para conseguir el empuje suficiente para alcanzar el vuelo.

4) *Alerones Estabilizadores*: Estas piezas se encargan de mantener estable y en curso el vuelo. Aquí se manifiesta el sistema de control por medio de unos flaps que permiten variar las rotaciones del cohete, por medio de la incidencia del aire.

B. Movimientos de Control en el Cohete

Un cohete en vuelo presenta movimientos en 3 ejes: pitch, roll y yaw, como se ve en la figura 2. Estas rotaciones permiten mover el cohete en la dirección deseada por medio del control.

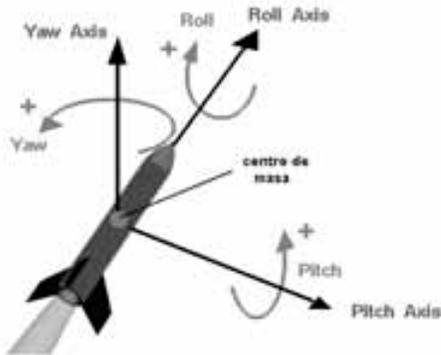


Fig. 2. Rotaciones del cohete. Fuente: <http://exploration.grc.nasa.gov/education/rocket/rotations.html>

C. Movimientos del cohete a partir de los flaps

El cohete prototipo propuesto cuenta con cuatro flaps ubicados en los alerones distribuidos tal y como se observa en la figura 3, de la siguiente forma:

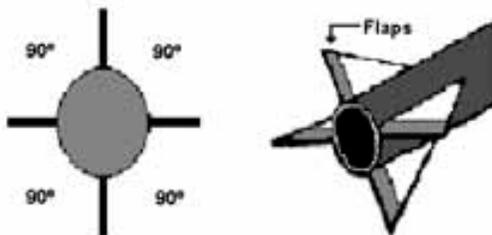


Fig. 3. Ubicación de los flaps en el cohete.

Por medio del movimiento de estos flaps el cohete puede variar su posición en Pitch, Roll y Yaw como se observa en las figuras 4, 5,6.



Fig. 4. Movimiento en Pitch

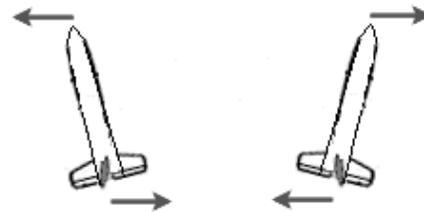


Fig. 5. Movimiento en Yaw

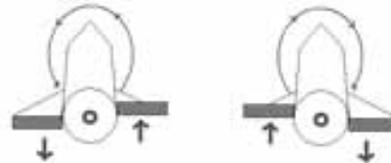


Fig. 6. Movimiento en Roll

III. DESCRIPCIÓN DEL PROTOTIPO

El prototipo diseñado se basa en el diagrama de bloques de la Fig. 7, en el cual se pueden observar todos los procesos del sistema y la secuencia en que se realizan.

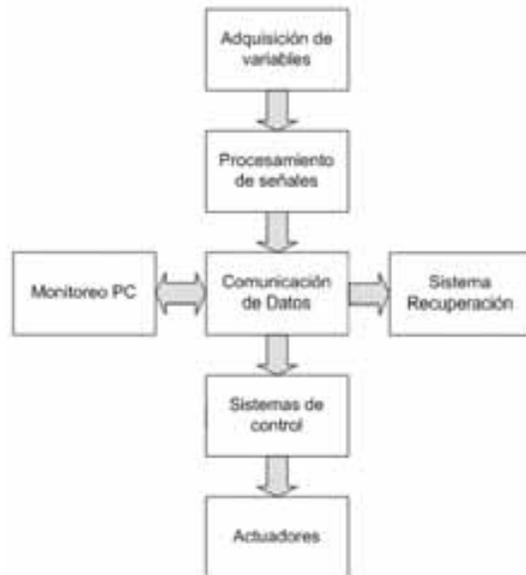


Fig. 7. Diagrama de bloques del sistema

A. Adquisición de Variables

La adquisición de variables esta a cargo de una etapa de instrumentación compuesta por varios sensores especializados, seleccionados cuidadosamente, teniendo en cuenta que las variables a medir para realizar los procesos de monitoreo y control son: la aceleración y velocidad angular y lineal, la altura, temperatura y dirección en la cual se esta desplazando el cohete en un plano de referencia. Para cada una de estas variables se tuvo en cuenta la siguiente instrumentación:

1) *Aceleración y velocidad lineal:* Para este caso se usan los sensores de aceleración ADXL78 y ADXL278 de uno y dos ejes respectivamente, con salida análoga, y un rango de medida de ± 70 gravedades. Estos dispositivos se adquirieron teniendo en cuenta las grandes aceleraciones que puede experimentar un cohete en vuelo. Para conocer la velocidad lineal del cohete se deriva la aceleración sensada, este proceso es realizado en tierra.

2) *Aceleración y velocidad angular:* Para el control del cohete es necesario conocer las velocidades angulares en Roll, Pitch y Yaw, estas variables se adquieren por medio de los giróscopos ADIS16100 con comunicación serial SPI de 16 bits. Este dispositivo permite la medición de la velocidad angular de un eje con un rango de $\pm 300^\circ/s$. Para adquirir el valor de la velocidad es necesario que los 3 sensores se encuentren ubicados en el centro de masa del cohete o mejor dicho el eje de rotaciones de este. Para conocer la aceleración angular del cohete se integra la velocidad sensada, este proceso es realizado en tierra.

3) *Altitud:* Para determinar la altura a la cual se encuentra el cohete se usa la presión atmosférica por medio de la relación no lineal de la Ec. (1), donde se obtiene la altura en metros a partir de la presión en kilo pascales.

$$h = \frac{\left(1 - \left(\frac{P_1}{101.325}\right)^{0.19026}\right) 288.15}{0.00198122} \times 0.3048 \quad (1)$$

Donde h es la altura y P_1 la presión atmosférica, que se adquiere gracias al sensor MPX4100A análogo que tiene un rango de medida de 15 a 115 kPa o 2.2 a 16.7 psi.

4) *Temperatura:* La temperatura de la cámara de instrumentos se puede adquirir por medio del sensor LM35, el cual entrega una salida de voltaje análogo lineal con un rango de 2 a $+150^\circ C$.

5) *Dirección en la cual se esta desplazando el cohete en un plano de referencia:* Para conocer en que dirección se esta desplazando el cohete es necesario adquirir los grados de inclinación y su orientación en un plano de referencia, como se ve en la Fig. 8 a., respecto a los ejes cardinales y la gravedad de la tierra.

Para la inclinación se utiliza el sensor ADIS16201 inclinómetro de dos ejes con comunicación SPI y rango de medida de $\pm 90^\circ$. Este sensor nos permite conocer la inclinación en los ejes X y Y. Por medio de la Ec. (2) se puede

conocer la inclinación total del cohete, representada por Phi en la Fig. 8. b.

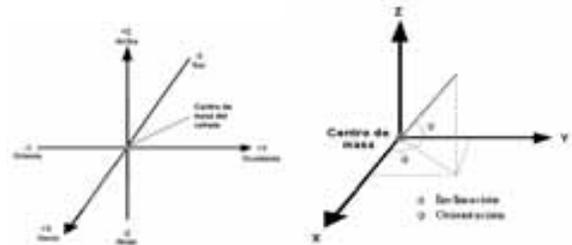


Fig. 8. a. Plano de referencia del cohete. 8. b. Dirección de vuelo representada por Theta y Phi.

$$\Phi = \tan^{-1} \left(\tan(\text{incli}_x) \cos \left(\tan^{-1} \left(\frac{\tan(\text{incli}_x)}{\tan(\text{incli}_y)} \right) \right) \right) \quad (2)$$

Para conocer la orientación del cohete respecto a los ejes cardinales de la tierra, representada por Theta en la Fig. 9, se implementaron los sensores magnéticos HMC1051 y HMC1052 que se configuran en forma de brújula electrónica de tres ejes capaz de sensar el campo magnético de la tierra. Los sensores miden un máximo en el norte y un mínimo en el sur siendo el oriente y el occidente los valores medios, como se ve en la Fig. 9. De tal forma que con una configuración ortogonal de los sensores se puede conocer el ángulo de su dirección respecto al norte magnético por medio del arreglo matemático de la Ec. (3) a partir de los sensores A y B (en el eje X y en el eje Y respectivamente). De esta forma se conoce la dirección de vuelo del cohete:

$$\begin{aligned} \text{Dirección}(A > 0) &= 90 - \left(\arctan \left(\frac{x}{y} \right) \right) \times \frac{180}{\pi} \\ \text{Dirección}(A < 0) &= 270 - \left(\arctan \left(\frac{x}{y} \right) \right) \times \frac{180}{\pi} \\ \text{Dirección}(A = 0, B < 0) &= 180 \end{aligned} \quad (3)$$

$$\text{Dirección}(A = 0, B > 0) = 0$$

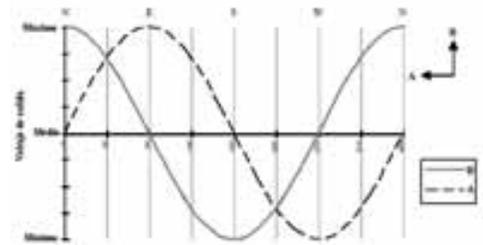


Fig. 9. Comportamiento de los sensores magnéticos.

B. Procesamiento de Señales

En el procesamiento de señales del sistema se tiene en cuenta que existen señales digitales en formato serial (Serial

Peripheral Interface - SPI) de 16 bits (inclinómetro y giróscopos) y señales análogas que deben ser adquiridas por el microcontrolador GP32 de Freescale (Motorola). Para las señales análogas se implementó un convertor análogo digital de 10 bits con comunicación digital SPI de 16 bits, para así poder obtener todos los datos de los sensores por medio de la misma comunicación.

Todos los sensores magnéticos tienen una salida de forma diferencial del orden de los milivoltios. Por tanto necesitan de un ajuste por medio de un amplificador de instrumentación que toma esta diferencia de voltaje para obtener una señal amplificada que puede ser interpretada por el convertor análogo digital y es proporcional al campo magnético sentido.

C. Comunicación de Datos.

Para la comunicación de datos es necesario aclarar que el sistema se dividió en tres subsistemas diferentes como se observa en la Fig. 10.

1) *Subsistema de monitoreo de sensores:* Es el maestro de todo el sistema. Este utiliza la comunicación SPI para los sensores, adquiriendo todas las variables que después transmite a los subsistemas de control y de comunicación también en el mismo formato.

2) *Subsistema de control:* Toma los datos necesarios para realizar el proceso de control a partir de una petición realizada por medio de un bit. Estos datos son procesados generando una señal PWM para los actuadores del sistema.

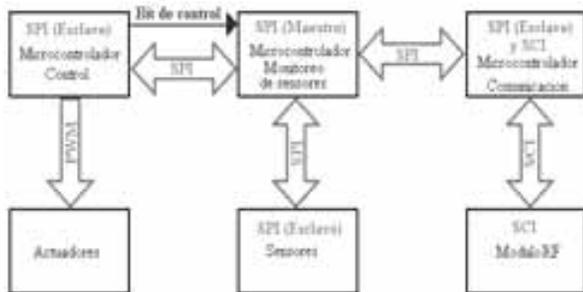


Fig. 10. Configuración de la comunicación de datos.

3) *Subsistema de comunicaciones:* Es el enlace entre el usuario y el sistema general, por medio de la comunicación SCI con un módulo de Radio frecuencia. Este transmite a un PC en tierra todos los datos medidos y a la vez recibe las peticiones del usuario.

Para el manejo de todos los datos de los sensores en cada uno de los subsistemas se utilizó una trama de comunicación, en donde los 4 bits más significativos representan la variable (sensor o control) y los 12 restantes el dato, como se ve en la Fig. 11.

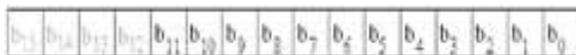


Fig. 11. Trama de datos del sistema.

D. Monitoreo PC

Para la adquisición de datos en el PC se usó una comunicación con el cohete conformada por dos módulos RF X09-009NSC de MaxStream, utilizando la configuración que se observa en la Fig. 12.



Fig. 12. Configuración de comunicación PC - Cohete

El monitoreo en el PC se realizó por medio del software MATLAB. Gracias al módulo RF y el puerto serial RS232 del PC, se adquieren todas las variables para procesarlas y visualizarlas en el panel de control de la Fig. 13.

Este panel permite observar la inclinación del cohete, la orientación respecto a los ejes cardinales de la tierra, la altura, la velocidad, la aceleración lineal, la temperatura en la cámara de instrumentos, el tiempo de vuelo, las magnitudes de los ángulos pitch, rol y yaw y sus respectivas aceleraciones angulares. Además este panel ofrece 4 botones al usuario que permiten iniciar la adquisición de variables, empezar la ignición, abortar el vuelo y enviar un dato de control para la dirección del cohete.

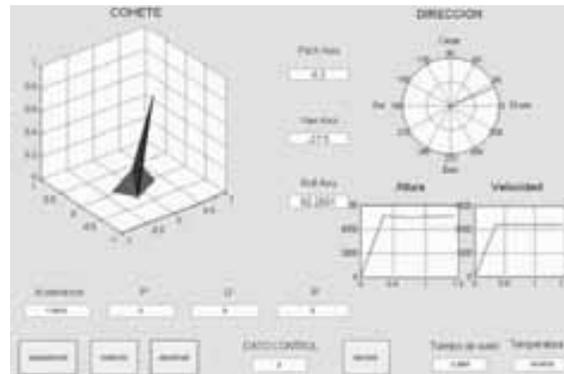


Fig. 13. Panel de monitoreo y control en Matlab.

E. Sistema de Recuperación

El sistema de recuperación (Fig. 14) está diseñado para activar la expulsión de un paracaídas a partir de una capsula de aire comprimido, cuando el sistema detecta que la aceleración lineal del cohete en el eje Z, es cero o empieza a hacerse negativa, o en el caso que el usuario en tierra lo considere necesario.

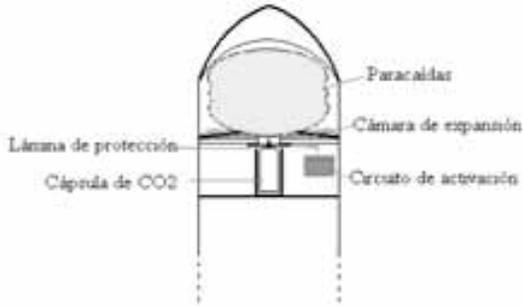


Fig. 14. Sistema de recuperación.

F. Sistema de Control

Ya se conoce que el sistema a controlar es un cohete por medio de unos flaps, en sus alerones, que permiten que se aplique un torque para corregir el ángulo de desviación del mismo cuando se presenta una perturbación en el ángulo deseado. Es importante señalar que para realizar el control se consideraron las siguientes condiciones:

- La masa del sistema no varía durante la implementación del control. Esta condición radica principalmente en que, en los cohetes experimentales, los tiempos de combustión son pequeños, por lo tanto la masa del propulsor se agota rápidamente, sin embargo, el tiempo de vuelo del cohete es lo suficientemente largo para llevar a cabo una tarea de control.

- El control se realizará de forma secuencial y condicional. Esto quiere decir que hay un controlador independiente para cada rotación posible del cohete y además se establece que la rotación sobre su mismo eje (rotación Roll) es la primera a controlar, si esto no se cumple las acciones de control de las otras rotaciones no se llevan a cabo.

1) *Ecuación característica del sistema*: Los tres sistemas a controlar poseen las mismas características, dos de estos son exactamente iguales (Pitch y Yaw) puesto que su inercia es la misma y el otro (Roll) en donde cambia este valor. Teniendo en cuenta lo anterior solo se realiza el modelamiento del sistema de control una vez, y al final se ajustan los valores de cada controlador utilizando la realimentación de estados.

El torque aplicado por 2 flaps para cada rotación está descrito por la Ec. (4).

$$I_{xx} \ddot{\beta} = \rho A L v_1^2 \sin^3 \beta \quad (4)$$

En donde I_{xx} es el momento de inercia en la rotación, $\ddot{\beta}$ es la aceleración angular, ρ es la densidad del aire, A es el área de los flaps, L es la longitud de entre el flap y el punto de rotación, v_1 es la velocidad del cohete y β es el ángulo de deflexión del flap y representa la entrada al sistema. A partir de lo anterior se puede realizar el modelamiento del sistema en espacio de estados. Sea el sistema representado por:

$$\dot{X} = AX + BU$$

$$Y = CX + DU$$

Las variables de estado y su entrada se encuentran en la Ec. (5).

$$\begin{aligned} x_1 &= p \\ x_2 &= \dot{p} \\ \beta &= u \end{aligned} \quad (5)$$

Donde p es la posición angular del cohete, \dot{p} es la velocidad angular y u es el ángulo de entrada. Se puede apreciar que la velocidad del cohete no se toma como una variable de estado, esto se debe a que durante el vuelo, después de la combustión, se toma un valor promedio de la velocidad para realizar el control. Teniendo ya las variables de estado y la Ec. (4), se pueden escribir las ecuaciones de la siguiente manera:

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= \frac{\rho A L v^2 \sin^3 u}{I_{xx}} \end{aligned} \quad (6)$$

Definiendo una constante c en la cual se incluyen las constantes de la segunda ecuación de estado se tiene:

$$c = \frac{\rho A L}{I_{xx}} v^2 \quad (7)$$

Reemplazando c en la segunda parte de la ecuación (6), se obtiene la Ec. (8) así:

$$\dot{x}_2 = c \sin^3 u \quad (8)$$

Estas ecuaciones se pueden escribir de forma matricial, Ec. (9).

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ c \lambda \end{bmatrix} u \quad (9)$$

Lambda (λ) hace las veces del seno cubo de la ecuación.

2) *Linealización*: Ya se tienen las ecuaciones con las cuales se puede modelar el sistema, pero se observa claramente que es un sistema no lineal por el seno al cubo que se presenta. Teniendo en cuenta esto, se realizó la linealización del sistema en un punto de equilibrio, pero el resultado arrojó que el sistema no es controlable.

Para solucionar este problema se decide expresar la función seno al cubo en una función a trazos que permita en cada uno de estos ser lineal, o por lo menos, en caso de que uno de sus trazos no sea lineal, obtener un punto de equilibrio que al ser evaluado durante la linealización arroje como resultado un sistema controlable. Con la ayuda del Toolbox de Matlab para ajuste de curvas y rectas se realiza la representación del seno cubo $(\sin(u))^3$ en 3 trazos, Fig. 15, teniendo en cuenta que los valores de u están limitados entre $-\pi/2, \pi/2$, con lo cual la función queda expresada como se observa en la Ec. (10).

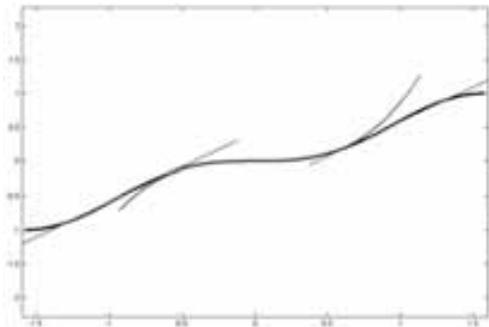


Fig. 15. Función seno cúbico y función $Fs(u)$ (a trazos)

$$\sin^3 u = Fs(u) = \begin{cases} 1.023u + 0.434 & \text{Para } (-1.4; -0.57) \\ 0.8678u^3 + 0.006971u & \text{Para } (-0.57; 0.57) \\ 1.023u - 0.434 & \text{Para } (0.57; 1.4) \end{cases} \quad (10)$$

Al realizar el cambio de función no se alteran las variables de estado (Ec. (11)), pero si las ecuaciones que representan el sistema, entonces, ahora el sistema en el espacio de estados queda representado de la siguiente forma:

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= c \times Fs(u) \end{aligned} \quad (11)$$

Como la función $Fs(u)$ (Ec. (10)) esta dada por trazos se debe realizar el proceso de modelamiento por cada trazo de la misma, teniendo especial cuidado en el segundo intervalo, pues es una función no lineal y se debe usar de nuevo un método de linealización.

3) *Diseño de controladores:* Este sistema de control, esta basado en las consideraciones del diseño del prototipo, como la velocidad promedio del cohete 180 m/s, longitud del cohete 1.3m, radio de 0.15m y peso de 8Kg entre otras.

Todas las funciones de transferencia tienen las mismas características sin importar el intervalo en donde se encuentre, poseen 2 polos en el origen por lo tanto son inestables y presentan la misma respuesta a la función escalón. En primer lugar se realiza el diseño para el comportamiento dinámico del sistema y luego teniendo este presente se ajustan las ganancias correspondientes al comportamiento estático, estos dos procesos se realizan utilizando la realimentación de estados e implementando precompensadores.

Sea la función de transferencia, Ec (12):

$$G(s) = \left[\frac{c_1}{s^2} \right] \quad (12)$$

En donde 'c' puede tomar cualquier valor, ya se sabe que este valor depende de varias constantes, entonces se realiza el cambio del polinomio característico por medio de la realimentación de estados. Se pretende que éste sea de la forma de la Ec. (13).

$$s^2 + 2\rho w_n s + w_n^2 \quad (13)$$

Y que cumpla las siguientes condiciones:

- Porcentaje de sobre pico máximo: $Mp = 0.15$
- Tiempo de establecimiento $t_2 = 0.5$ seg

Por tanto el polinomio característico resulta como en la Ec. (14).

$$s^2 + 18.4s + 147.4657 \quad (14)$$

En la Fig. 16 se observa la respuesta del sistema (para cualquier intervalo) a una entrada escalón, cumpliendo con las condiciones establecidas pero por debajo de la respuesta estática deseada.

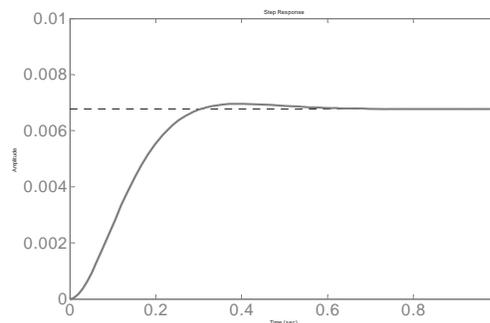


Fig. 16. Respuesta al escalón para cualquier intervalo

Para solucionar la respuesta estática se utilizan precompensadores, uno para cada sistema, como se ve en la Fig. 17.

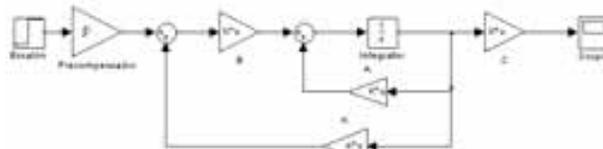


Fig. 17. Diagrama de realimentación por variables de estado con precompensador

Obteniendo como resultado una salida optima para el objetivo del proyecto, y con una respuesta al escalón como se ve en la Fig. 18.

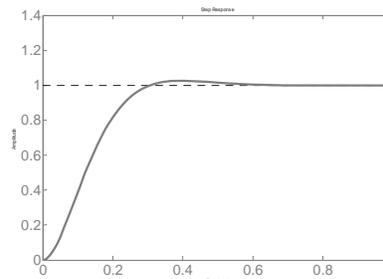


Fig. 18. Respuesta al escalón con precompensador

El sistema de control diseñado es implementado en un microcontrolador GP32.

G. Actuadores

Los actuadores usados en el sistema son servomotores BMS-410 que permiten un movimiento de hasta 180°, controlados a partir de la señal PWM proveniente del subsistema de control. Estos servomotores mueven los flaps para generar un cambio en la dirección del vuelo del cohete. La señal PWM utiliza un periodo de 10ms y el ancho del pulso varía entre 1 y 2 ms, aunque estos valores pueden cambiar según el motor usado.

IV. CONCLUSIONES

Para llegar a resultados más precisos se deben realizar pruebas reales como en un túnel de viento o diferentes lanzamientos del cohete para así identificar aspectos que puede llegar a mejorar el sistema. Sin embargo, como el objetivo fue diseñar un sistema de control electrónico para mantener la posición del cohete, y así no se hayan tenido en cuenta algunos aspectos particulares como aerodinámica, esfuerzos mecánicos de cada una de las piezas, determinación del punto de presión, variación de la masa del sistema, entre otros, el sistema puede ajustarse a cualquier tipo de cohete, modificando las ecuaciones que rigen el comportamiento del sistema.

El desarrollo final de este trabajo se integra en tres tarjetas que son capaces de realizar todo el proceso de instrumentación, comunicación y control. La ubicación de cada tarjeta en el cohete debe cumplir con las condiciones planteadas para la adquisición de variables generando un plano de referencia y valores correctos del movimiento respecto al centro de masa del cohete.

REFERENCIAS

- [1] J.D. Anderson, "Introduction to flight". Ed. Mc. Graw Hill. 3ªEd.
- [2] E. Avendaño, "Matlab aplicado a la solución de problemas de ingeniería". Edición Preliminar. Universidad Santo Tomas Tunja. 2007.
- [3] R. Coughlin, "Amplificadores Operacionales y circuitos integrados lineales". 5ªEd
- [4] O. Duarte, "Introducción a los Sistemas no Lineales". Facultad de Ingeniería Universidad Nacional de Colombia
- [5] C. A. Duque, "Dinámica de gases de un pequeño motor-cohete-modelación por CFD". Trabajo de grado, Ing. Mecánica-Universidad Nacional de Colombia.
- [6] D. B. French, "Thesis. Hybrid control strategies for rapid, large angle satellite slew maneuvers". Air force institute of technology. Ohio, USA.
- [7] D. A. Garzón, "Introducción General a la tecnología de propulsión". Universidad Nacional de Colombia.
- [8] H. Khalil, "Nonlinear Systems". Prentice Hall, NJ, 1996. 2ªEd
- [9] Military Handbook. "Design of aerodynamically stabilized free rockets". Departement of Defense, USA.
- [10] W. J. Palm, "System dynamics". McGraw-Hill, 2005
- [11] C. PÉREZ, "Matlab y sus aplicaciones en las ciencias y la ingeniería". Ed. Prentice Hall.
- [12] H. Sira; R. Márquez; F. Rivas, "Control de sistemas no lineales".
- [13] W. Tomasi, "Advanced electronic communication systems". Prentice Hall International, cop. 1992

[14] (Jonson Space Center Portal). (2007, Octubre 21). Disponible en: <http://history.msfc.nasa.gov/rocketry/>

[15] R. Nakka's, (2007, Octubre 25). Richard Nakka's Experimental Rocketry Web Site. Disponible: <http://www.nakka-rocketry.net/>

[16] V. R. González, (2007, Noviembre 3). Curso de Robótica Centro de Formación de Profesores en Innovación Educativa Valladolid, España. Disponible en : http://cfievalladolid2.net/tecnocyr_01/robotica



Jiménez López Fabián Nacido en Sogamoso, Colombia, en Marzo 13 de 1977. Recibió el título de Ingeniero Electrónico de la Universidad Pedagógica y Tecnológica de Colombia, Seccional Sogamoso, Colombia en el 2000, y el título de Especialista en Automatización Industrial de la Universidad Pedagógica y Tecnológica de Colombia, Seccional Sogamoso en el 2002. En el 2001, ingreso a la Facultad de Ingeniería Electrónica de la Universidad Santo Tomas Tunja.

Aspirante a Maestría en Electrónica - Automatización y Control, Universidad Nacional de Colombia. Actualmente labora como docente en las áreas de Control, Tratamiento Digital de Señales y Automatización. En el 2004 crea el Grupo de Investigación ARMÓNICO – DSP, y es director del Grupo en la línea de Instrumentación y Control.

Medidas antropométricas para la identificación de características en imágenes del rostro

Julián Fernando Garzón Sánchez, María Paula Otero Díaz, Juliana Reinemer Valencia, Pedro Vizcaya¹
j.garzon@javeriana.edu.co motero@javeriana.edu.co juliana.reinemer@javeriana.edu.co
Departamento de Electrónica, Pontificia Universidad Javeriana, Bogotá, Colombia

Resumen—En este trabajo se presenta el problema de investigar, desarrollar y evaluar medidas antropométricas para el registro de puntos útiles y comparables del rostro humano para la identificación de similitudes de dos imágenes faciales de la misma persona, tomadas con una diferencia de tiempo apreciable. Se usa el Discriminante de Fisher y el vector de dirección de mayor separación de clases para proyectar la información de un par de fotografías. El proceso fue utilizado para la selección de características antropométricas útiles del rostro humano. Como resultado se obtiene un método generalizado para comparar un grupo de fotografías pertenecientes a una base de datos para la verificación de identidad, que presenta un porcentaje de error de clasificación de las fotos.

Palabras Clave—Antropometría, Discriminante de Fisher, Interpolación, verificación de identidad, error de clasificación.

I. INTRODUCCIÓN

Las fotografías son una fuente de información valiosa al hablar de verificación de identidad, pues son un registro de los cambios que ha sufrido una persona a lo largo de los años. Los documentos de identidad en su gran mayoría presentan una foto para identificar al portador y en muchos casos son solicitados para tener certeza de quien es la persona, sin embargo, en general las fotos fueron tomadas al cumplir la mayoría de edad, y la verificación de identidad es difícil de realizar.

La investigación que se muestra a continuación toma las características faciales de imágenes del rostro pertenecientes a una base de datos para establecer un porcentaje de error al verificar la identidad de una persona.

II. DESCRIPCIÓN

La obtención de la base de datos de fotografías es el punto de partida de la investigación, y a su vez es la primera fase de

cuatro de las que se compone la investigación. Las siguientes etapas son: selección de característica de normalización, escogencia de los parámetros más importantes y cálculo del error.

Con la información que aporta la investigación se realizaron pruebas con una nueva base de datos para evaluar el modelo elaborado.

En paralelo con esta investigación se realizó un experimento manual de verificación de identidad.

III. DESARROLLO

A. Obtención de la Base de Datos

Para el estudio se construyó una base de datos que contiene una serie de fotografías de hombres y mujeres desde los dieciocho años de edad hasta los sesenta. Esta base de fotos contiene como mínimo dos imágenes frontales del rostro por persona, deben ser sin expresión y no se tuvieron en cuenta los casos de gemelos, personas con deformaciones, ni con cirugías plásticas faciales. Este fue un proceso constante a lo largo del desarrollo del proyecto ya que en paralelo a las demás fases, se construyó la base de datos de prueba con la que se validó el modelo creado con el prototipo.

Sobre las fotografías se tomaron las medidas establecidas por la antropometría facial teniendo en cuenta que por ser medidas manuales se incluye un error de medición, que se redujo promediando las medidas realizadas por dos personas o más. Se miden en total catorce características de las cuales nueve son horizontales y cinco son verticales, como se muestra a continuación:

¹ Manuscrito enviado el 26 de Agosto de 2008. Este artículo es el resultado del trabajo de final de grado presentado para optar al título de Ingeniero Electrónico.

P. Vizcaya es profesor titular de la Facultad de Ingeniería de la P. Universidad Javeriana (pvizcaya@javeriana.edu.co) del área de Comunicaciones y Análisis de Señales.

J. Garzón, Ma. Paula Otero, J. Reinemer, son ahora egresados de la carrera de Ingeniería Electrónica de la Pontificia Universidad Javeriana.

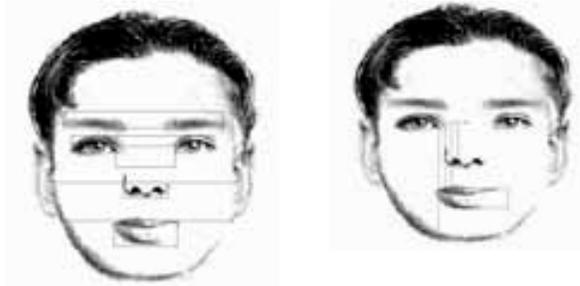


Fig. 1. Medidas horizontales y verticales realizadas sobre las fotografías²

B. Normalización de las medidas

Las medidas realizadas se registraron por separado para hombres y mujeres, junto con un código que diferencia un sujeto de otro. Como el tamaño de las fotografías no es el mismo, las medidas no son comparables entre sí, lo que hizo necesario normalizar estas mediciones. Esta escogencia se hizo con base en el resultado del Discriminante de Fisher, el cual es un método de clasificación que permite encontrar un vector de proyección que maximiza la distancia entre las medias de las clases (entre clases) y minimiza la varianza de cada una de las clases (dentro de la clase). El discriminante se expresa mediante la siguiente ecuación:

$$J(a) = \frac{a^T S^{(b)} a}{a^T S^{(w)} a} \quad (1)$$

Donde

$$S^{(w)} = \sum_{i=1}^l \sum_{j:y_j=i} (x_j - \mu_i)(x_j - \mu_i)^T \quad (2)$$

$$S^{(b)} = \sum_{i=1}^l (\mu_i - \mu)(\mu_i - \mu)^T \quad (3)$$

$S^{(w)}$ es la matriz de correlación cruzada dentro de la clase y $S^{(b)}$ es la matriz de correlación entre las clases. El vector a es el que mejor separa las clases, y se obtiene al solucionar el problema de los vectores y valores propios generalizado. Con esto se escogen los parámetros de normalización vertical y horizontal.

C. Selección de características más importantes

Las doce características se convierten en setenta y ocho al relacionarlas entre sí, con el fin de tener más información por fotografía. Con estas características normalizadas se realizó el cálculo del Discriminante de Fisher.

Para diferenciar un sujeto de otro dentro de la matriz de medidas se realizó una máscara que consiste en una matriz cuadrada cuya longitud es la cantidad total de fotos dentro del estudio. Esta matriz contiene recuadros con unos en las posiciones donde se tienen fotos que corresponden a una misma persona. A partir de estas dos matrices (máscara y medidas) se realiza una comparación foto a foto y se genera una matriz de diferencias en la que la diagonal es cero, pues se compara una foto con ella misma. En las demás posiciones se

tiene la diferencia del valor correspondiente a una foto con otra, es decir en la coordenada (x,y) se tendrá el valor de la foto x menos el valor de la foto y .

D. Cálculo del error de clasificación

El proceso de Fisher arroja además de J , el vector a , que separa dos clases, en este caso estas clases son: se trata de la misma persona o por el contrario, son personas diferentes. Este vector es utilizado para proyectar sobre él un vector que contiene la información de la comparación de una foto con el resto. Como resultado de esto se tiene una matriz cuadrada simétrica cuyas coordenadas corresponden al resultado del producto punto del vector de dirección arrojado por el proceso de Fisher y la diferencia de una foto con otra. La diagonal de esta matriz, por lo tanto también tiene sus valores en ceros. Si se ordena esta matriz de manera ascendente en conjunto con el código asignado a cada foto se identifican en las primeras posiciones las fotos que más se parecen. Con esto se halla el error ubicando las fotos que no caen dentro de las posiciones correspondientes al sujeto en estudio. Este cálculo se realiza mediante la siguiente ecuación en donde se tiene en cuenta la posición en la que se encuentra cada una de las fotografías pertenecientes a un sujeto.

$$\text{error} = \frac{1}{\text{cant}(\text{total} - \text{cant})} \sum_{n=1}^{\text{cant}} (\text{pos}_n - \text{cant}) \quad (4)$$

para $\text{pos}_n > \text{cant}$

Donde *total* es la cantidad total de fotos pertenecientes a la base de datos, *cant* es la cantidad de fotos del sujeto que se está analizando y *pos* es la posición en la que se encuentra la foto que se analiza. La condición $\text{pos} > \text{cant}$ asegura que las fotos que se encuentran dentro del rango del sujeto no aportaran error. Este procedimiento se hace por separado para hombres y mujeres

E. Base de datos de prueba

El mismo procedimiento se realizó con una nueva base de datos generada con las mismas especificaciones de la base de datos utilizada para la generación del modelo. Se normaliza con los parámetros escogidos (ojo-boca centro y pupilas) y se halla el error con la totalidad de las características y con el mismo número de características usadas para alcanzar el 90% de la información.

F. Selección a través de la regresión lineal y extrapolación de características en un rango de edades

Se identificaron las características a través del tiempo y se observó si las distribuciones corresponden a funciones monótonamente crecientes o decrecientes, para esto se calcularon las diferencias entre los valores adyacentes de cada una de las características. Se realiza la regresión lineal de cada una ellas y con estos valores se genera la serie de tiempo de cada uno de los sujetos, es decir, se realiza una extrapolación de las características en un rango de edades y se procede a calcular el discriminante de Fisher que selecciona las características que aportan más información la dirección óptima de separación de las clases de interés.

² Figuras realizadas en el programa Ultimate Flash Facev0.42b. se encuentra en la página: <http://flashface.ctapt.de/index.php>

G. Experimento Manual de verificación de identidad

El experimento fue realizado con la herramienta GUIDE de MATLAB® como se presenta a continuación:



Fig. 2. Ejemplo del experimento manual de verificación de identidad

Se presentan al usuario un par de fotografías y este debe decidir si se trata de la misma persona o no. Se hacen dos experimentos por separado, uno para hombres y otro para mujeres. Esto con el propósito de establecer falsos aciertos y falsos rechazos que se presentan al clasificar un par de fotografías.

IV. CONCLUSIONES

A. Normalización

Para las medidas horizontales se escogió la distancia entre pupilas. Para las verticales se tenían dos posibles normalizaciones: boca vertical y ojo-boca centro. Se escogió la segunda dada la facilidad de medición sobre una imagen del rostro.

B. Selección de características

Para el prototipo de los hombres se tomaron 17 de las 78 características posibles, y en el caso de las mujeres, 28. Se encontraron ocho características en común, sin embargo su aporte en cada uno de los modelos no es el mismo ya que se ubicaron en posiciones diferentes.

C. Error de clasificación

El error de clasificación fue mayor para las mujeres que para los hombres tanto en promedio como en su valor máximo. A continuación se presentan los resultados de la evaluación del prototipo.

HOMBRES BASE DE DATOS PROTOTIPO	
MÁXIMO	PROMEDIO
12.0244%	1.4734%
MUJERES BASE DE DATOS PROTOTIPO	
MÁXIMO	PROMEDIO

27.9412% 3.2447%

Tabla 1. Error para la base de datos prototipo con todas las características

Para simplificar el proceso se tomaron únicamente las características con las que se alcanza el 90% del valor máximo del discriminante de Fisher y se obtienen los mismos resultados, sin embargo, el error en general aumenta para este segundo caso tanto para hombres como para mujeres.

HOMBRES BASE DE DATOS PROTOTIPO (17 CARACTERÍSTICAS)	
MÁXIMO	PROMEDIO
19.0259%	3.1477%
MUJERES BASE DE DATOS PROTOTIPO (28 CARACTERÍSTICAS)	
MÁXIMO	PROMEDIO
64.1176%	5.6820%

Tabla 2. Error para la base de datos prototipo con las características más importantes

Estos resultados se visualizaron en una matriz de confusión como la que se presenta a continuación:

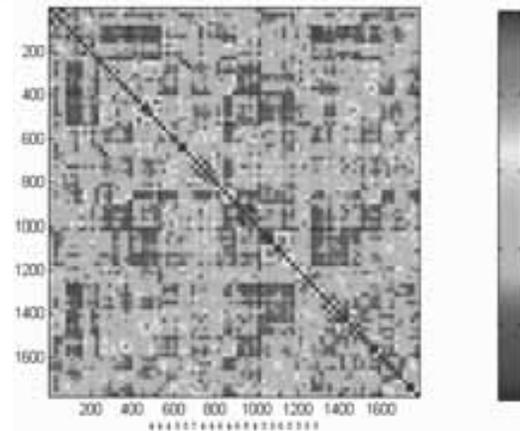


Fig. 2. Ejemplo del Matriz de Confusión

Esta matriz contiene en tonos rojos las fotos que presentan características similares y en azul las que no. La diagonal tiene todos sus valores en café pues es la comparación de una foto con ella misma. Alrededor de la diagonal también se encuentran posiciones rojas pues se trata del mismo sujeto. Se puede ver que se encuentran posiciones en las que fuera de la diagonal hay coordenadas en rojo, estas corresponden a las falsas aceptaciones. Los casos en los que alrededor de la diagonal se tienen puntos azules corresponden a falsos rechazos.

En la matriz de confusión para los hombres, cuando se toman menos características, aparecen más posibles puntos de confusión, lo cual indica un número mayor de falsas aceptaciones (incremento en el error de clasificación). Para las mujeres, disminuye el número de falsas aceptaciones, aunque el error promedio aumenta.

D. Base de datos de prueba

A continuación se presentan las tablas con los resultados del error de la base de datos de prueba.

HOMBRES BASE DE DATOS DE PRUEBA		
MÁXIMO	SUJETO DE MÁXIMO ERROR	PROMEDIO
40.0970%	Foto 38, Sujeto 95	8.3862%
MUJERES BASE DE DATOS DE PRUEBA		
MÁXIMO	SUJETO DE MÁXIMO ERROR	PROMEDIO
38.3333%	Foto 58, Sujeto 135	7.6105%

Tabla 3. Error para la base de datos de prueba con todas las características

HOMBRES BASE DE DATOS DE PRUEBA (17 CARACTERÍSTICAS)		
MÁXIMO	SUJETO DE MÁXIMO ERROR	PROMEDIO
54.1060%	Foto 38, Sujeto 95	11.2766%
MUJERES BASE DE DATOS DE PRUEBA (28 CARACTERÍSTICAS)		
MÁXIMO	SUJETO DE MÁXIMO ERROR	PROMEDIO
36.6667%	Foto 58, Sujeto 135	8.0251%

Tabla 4. Error para la base de datos de prueba con las características más importantes

Con la base de datos de prueba se obtienen los mismos resultados de error, aunque en general el error aumenta

E. Selección a través de la regresión lineal y extrapolación de características en un rango de edades

Al realizar las extrapolaciones para cada uno de los grupos de sujetos, con las características básicas normalizadas sin relacionar entre si y luego se realizó el cálculo del discriminante de Fisher. Se obtuvo que las características que aportan la mayor información son diferentes entre los Hombres y las Mujeres, por esto la hipótesis de hacer el análisis por separado para Hombres y Mujeres fue validada; el grupo de las mujeres tuvo un mayor coeficiente de Fisher que los hombres, sin embargo aunque el aporte de información de las Mujeres es mayor, el 90% de la información se alcanza primero con las características de los hombres. Por otro lado se encontró la necesidad de incluir más observaciones dentro del modelo, en este caso particular son las relaciones entre todas las características

F. Experimento Manual de verificación de identidad

Los porcentajes de falsas aceptaciones y falsos rechazos eran diferentes de la persona que realizara el experimento (hombre o mujer). Los resultados se presentan a continuación.

Grupo	Falsas aceptaciones (%)	Falsos Rechazos (%)
Hombres realizando experimento de verificación a Hombres	16.83	0.6
Hombres realizando experimento de verificación a Mujeres	23.218	1.1494

Mujeres realizando experimento de verificación a Hombres	14.724	1.84
Mujeres realizando experimento de verificación a Mujeres	19.032	0.96

Tabla 5. Resultados del experimento manual de verificación de identidad

Con la información obtenida se concluye que las diferencias no son estadísticamente significativas y que todas representan el mismo experimento. Al analizar los resultados de cada uno de los experimentos se observó que el porcentaje de falsas aceptaciones y falsos rechazos es mayor cuando se quiere identificar mujeres, de igual forma se observó que las personas que realizaron el experimento se fijaban en los rasgos particulares de los sujetos como cicatrices, lunares, forma de los ojos, las cejas, el mentón y los aspectos generales de las fotos como la época y la calidad del color.

REFERENCIAS

- [1] FARKAS, Leslie. Anthropometry of the Head and Face, NY Press, Ed 2 1994
- [2] CHELLAPPA, R. y RAMANATHAN, N. Face Verification across Age Progression. En: Image Processing, IEEE Transactions on Volumen 15; Numero 11; Nov. 2006; p.:3349-361.
- [3] HILLMAN G, GOPE C. y KEHTARNAVAZ N. Zernike Moment Invariants Based Photo-identification Using Fisher Discriminant Model. En: Engineering in Medicine and Biology Society, 2004. IEMBS apos;04. 26th Annual International Conference of the IEEE, Volumen 1; Número 1-5; Sept. 2004;p. 1455 – 1458.
- [4] WALPOLE Ronald E, Probabilidad y estadística para ingenieros; Cuarta edición, Editorial Mc Graw Hill; 1992; p 317-331
- [5] STORKS Henry, Procesos estocásticos; p.259- 269.
- [6] IHRIG Emil y IHRIG Sybil, Manual del escáner para profesionales, traducción José Ángel Vallejo Pinto; Madrid : Osborne-Mcgraw-Hill; 1996; p 89, 136-139.
- [7] DUDA Richard, HART Meter Elliot, STORK Davis, Pattern Classification, Segunda edición, Editorial Wiley; 2001; p.114-123.
- [8] DAVIS, Philip J, Interpolation and approximation, New York : Dover, 1975.

Sistema de clasificación de imágenes basado en técnicas de reconocimiento de patrones aplicado en termografía y robótica

Nini Juliana Vélez, Jorge Humberto Erazo y Humberto Loaiza

Resumen— Este artículo describe el diseño e implementación de un sistema para clasificación de imágenes aplicando técnicas de reconocimiento de patrones. Se utilizan como descriptores cuatro momentos estadísticos y siete momentos invariantes de Hu, se emplea un clasificador estadístico y dos clasificadores neuronales. El sistema se valida con dos aplicaciones: inspección termográfica y robótica.

Palabras claves—Extracción de características, momentos, clasificador bayesiano, redes neuronales, correlación cruzada, termografía infrarroja, sistema robótico.

I. INTRODUCCIÓN

EL problema de diseño de un clasificador comienza con la definición de las clases objeto de estudio y la representación adecuada de la muestra controlada de cada una de ellas (extracción de características). El problema termina cuando se ha obtenido una regla de decisión sobre la cual se puede asignar un nuevo patrón (cuya clase de pertenencia se desconoce) a aquella clase para la que se estime un menor riesgo en la asignación [1].

En este artículo se presenta el proceso de diseño de tres clasificadores: uno bayesiano y dos con redes neuronales: perceptrón multicapa (MLP) y redes de base radial (RBF), para la clasificación de imágenes, utilizando como características cuatro momentos estadísticos y siete momentos invariantes de *Hu*.

El sistema se valida con dos aplicaciones. La primera en termografía que consiste en un prediagnóstico de la operación normal o anómala de un motor eléctrico a partir del análisis de sus imágenes infrarrojas. La segunda consiste en identificar un objeto de un sistema robótico asignándolo a una de las dos posibles clases existentes robot u obstáculo.

Los resultados corroboran la característica discriminante de los descriptores seleccionados y la capacidad de generalización de los clasificadores implementados.

A continuación se presenta una descripción del sistema. Posteriormente se consignan las pruebas y resultados y finalmente las conclusiones.

II. DESCRIPCIÓN GENERAL DEL SISTEMA

Todo el sistema de reconocimiento de patrones es

implementado en Matlab, utilizando los toolbox: *stprtool*, *netlab* y *neural network*.

El siguiente diagrama de bloques ilustra las principales etapas que componen el sistema (figura 1).

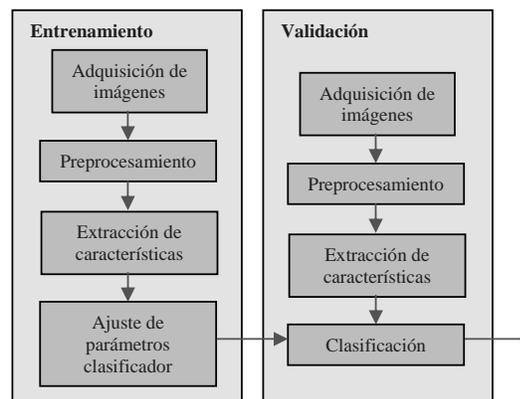


Fig.1. Diagrama de bloques del sistema implementado.

A. Adquisición de imágenes

El sistema tiene como entrada una etapa para la adquisición de información desde una cámara. El resultado corresponde a imágenes en formato RGB.

B. Preprocesamiento

Las imágenes producto de la etapa de adquisición necesitan adecuarse para su posterior uso.

La primera fase del preprocesamiento consiste en realizar un cambio de formato de RGB a escala de grises.

Posteriormente se realiza un proceso de segmentación, debido a que no toda la escena producto de la captura de imágenes es útil. Esta operación se realiza manualmente para que pueda ser independiente de la naturaleza de las imágenes de entrada lo que reduce la complejidad del sistema.

C. Extracción de características

El objetivo de una etapa de extracción de características es la de poder caracterizar un objeto para que éste sea identificado por mediciones cuyos valores sean muy similares entre objetos de la misma clase y muy diferentes para objetos de una clase

diferente [2][3]. Las estadísticas de imágenes digitales en los últimos años han sido ampliamente estudiadas [4]. Teniendo en cuenta lo anterior, se decide utilizar la información que proporcionan los 4 primeros momentos estadísticos y los siete momentos de Hu [5][6], los cuales son invariantes a la rotación, traslación y cambio de escala. El resultado es una matriz de tamaño $n \times 11$, donde n corresponde al número de muestras que han sido procesadas y 11 determina el número de características por cada muestra.

D. Ajuste de parámetros para clasificadores

En este bloque se utilizan 3 tipos de clasificadores: uno estadístico y dos basados en redes neuronales artificiales.

Clasificador estadístico

La teoría de decisión bayesiana es un enfoque estadístico fundamental para resolver problemas que involucren la clasificación de patrones. Se basa en el teorema de Bayes (1)

$$P(C_k/X^l) = \frac{P(X^l/C_k)P(C_k)}{P(X^l)} \quad (1)$$

Donde, $P(C_k/X^l)$ representa la probabilidad (a posteriori) de que tomada una observación X^l ésta pertenezca a la clase C_k , $P(X^l/C_k)$ especifica la probabilidad de observar los datos X^l dado que se tenga la opción C_k y $P(C_k)$ es la probabilidad a priori.

El objetivo es entonces encontrar la probabilidad a posteriori de que un patrón pertenezca a la clase C_k una vez se ha observado cierto vector de características X^l .

La probabilidad de error en la clasificación se minimiza seleccionando la clase C_k que tenga una mayor probabilidad a posteriori [7]. En términos de un conjunto de *funciones discriminantes* $y_1(X^l), \dots, y_c(X^l)$ un vector X^l será asignado a la clase C_k si:

$$y_k(X^l) > y_j(X^l) \text{ para todo } j \neq k \quad (2)$$

Asumiendo que la estructura probabilística sigue la forma de una distribución normal o gaussiana; el teorema de bayes se puede representar así [7]:

$$y_k(X^l) = -\frac{1}{2}(X^l - \mu_k)^T \Sigma_k^{-1} (X^l - \mu_k) - \frac{1}{2} \ln(|\Sigma_k|) + \ln(P(C_k)) \quad (3)$$

En la ecuación (3) μ_k es un vector d -dimensional de medias y Σ_k es la matriz $d \times d$ de covarianza.

El ajuste de los parámetros necesarios para la implementación de un clasificador bayesiano se realiza encontrando tanto el vector de medias como la matriz de covarianza a partir de la información extraída (*características*) de las muestras controladas (imágenes adquiridas).

Clasificadores neuronales

Las redes neuronales artificiales (RNA) se pueden usar también para la clasificación de patrones, ya que son

ampliamente conocidas por su capacidad para realizar transformaciones matemáticas entre conjuntos de variables, su baja sensibilidad al ruido y sus capacidades de aprendizaje y generalización.

Una configuración ampliamente utilizada son los perceptrón multicapa MLP [4].

Las neuronas son entrenadas tomando como base ejemplos de desempeño o un comportamiento deseado. El comportamiento deseado se puede resumir en un conjunto de pares ordenados de entrada y salida $(p_1 t_1, p_2 t_2, \dots, p_Q t_Q)$ donde p es el vector de entrada a la red y t es la salida correcta (objetivo) correspondiente. Se busca entonces reducir el error e que se define como la diferencia entre la salida correcta y la salida real de la red.

$$e = t - a \quad (4)$$

La regla de aprendizaje tipo perceptrón calcula cambios deseados a los pesos y bias de la red, dado un vector de entrada p y el error asociado e . Cada vez que el algoritmo es ejecutado, el perceptrón tendrá mayor probabilidad de producir salidas cercanas a las correctas. Siempre y cuando exista una solución el perceptrón convergerá después de un número finito de iteraciones [5].

Para el sistema implementado se utiliza en las redes MLP un número neuronas en la capa de entrada igual al número de características extraídas por imagen y un número de neuronas en la capa de salida igual al número de clases del sistema. El número de neuronas en la capa oculta es igual al de la capa de entrada.

Las redes de base radial RBF pueden requerir más neuronas que una red MLP común, pero con la ventaja que pueden ser diseñadas en solo una fracción del tiempo que toman estas últimas. Una RBF trabaja mejor cuando hay disponible un grupo de entrenamiento grande.

Para este tipo de redes la expresión para la entrada a sus neuronas es diferente a la que tiene una red MLP. El argumento para la función de activación f de la red es el vector distancia entre su vector de pesos w y el vector de entrada p .

La función de transferencia de una RBF está dada por:

$$f(n) = e^{-n^2} \quad (5)$$

La función de base radial (5) tiene un máximo de 1 cuando su entrada es cero. Como la distancia entre w y p decrece, la salida incrementa. Entonces, una neurona con función de activación de base radial se comporta como un detector que produce 1 cuando la entrada p es idéntica a su vector de pesos w .

E. Clasificación

La finalidad de un sistema de clasificación es asignar un objeto cuya clase de pertenencia es desconocida a una que haga parte de la base de conocimiento del sistema, teniendo en cuenta los parámetros previamente ajustados en el

clasificador. De esta forma se obtiene una métrica que define la clase asignada al objeto en cuestión.

Para el clasificador bayesiano, se evalúan todas las funciones discriminantes del sistema (tantas como clases existan) y se asigna el objeto a aquella clase cuya función haya arrojado el valor más alto (3).

Para las redes neuronales, se asignará el objeto a la clase cuya neurona de salida haya obtenido el valor más alto en la función de activación.

Finalmente para hacer más robusto el sistema se utiliza el método conocido como correlación cruzada, con en el cual los grupos de entrenamiento y validación cambian aleatoriamente de tal forma que se puede observar y comparar el desempeño de los distintos clasificadores obtenidos y seleccionar aquel que se ajuste mejor a las especificaciones deseadas.

III. PRUEBAS Y RESULTADOS

A continuación se presentan las pruebas realizadas sobre el sistema con cada uno de los casos de estudio. Los resultados obtenidos se consignan en tablas utilizando matrices de confusión. En la Tabla I se muestra el significado de los campos asociados a esta matriz.

TABLA I
MATRIZ DE CONFUSIÓN

	C1	C2
C1	N1	N2
C2	N3	N4
	D	NA

C1: clase 1; C2: clase 2; Ñ1: Numero de elementos pertenecientes a C1 y clasificados como C1; Ñ2: Numero de elementos pertenecientes a C1 y clasificados como C2; Ñ3: Numero de elementos pertenecientes a C2 y clasificados como C1; Ñ4: Numero de elementos pertenecientes a C2 y clasificados como C2; D: Desempeño global del sistema en porcentaje de aciertos; NA: Número total de aciertos.

A. Caso de estudio No.1: Detección de fallas de un motor eléctrico a partir de imágenes termográficas

Para la recolección de la base de imágenes a utilizar en esta prueba, se utilizó como motor el prototipo del sistema Feedback ES151 y una cámara termográfica FLIR A20.

Inicialmente se tomaron imágenes, con variaciones periódicas de la unidad de carga magnética incluida en el ES151, la cual tiene diez posibles estados; bajo este procedimiento se archivaron las imágenes con la etiqueta de muestras de operación normal (C1). Para simular condiciones de operación anómala del sistema ES151, se aplicaron perturbaciones no uniformes cuantificables en el incremento de la corriente requerida por el motor; las cuales produjeron en el sistema un aumento en la temperatura normal de operación de la unidad. Estas situaciones se guardaron bajo la etiqueta de muestras de operación no deseada (C2). El conjunto completo de imágenes se tomó en dos paletas de color distintas: hierro (figura 2.a) y en escala de grises (figura 2.b) en formato RGB. La figura 3 corresponde a las mismas paletas, pero con el motor bajo condiciones de funcionamiento anómalas.

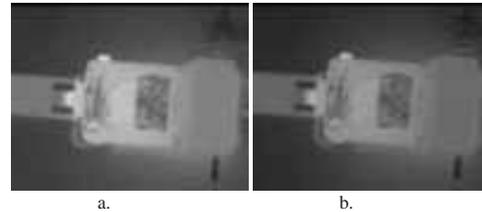


Fig. 2. Imágenes termográficas para un funcionamiento normal del motor. (a) Paleta hierro. (b) Escala de grises.

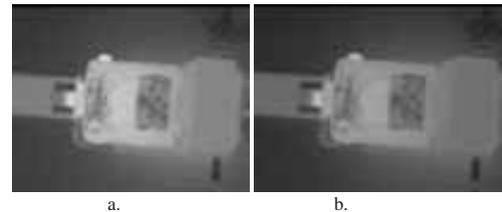


Fig. 3. Imágenes termográficas para un funcionamiento anómalo del motor. (a) Paleta hierro. (b) Escala de grises.

Teniendo en cuenta que no toda la información de la imagen es relevante, se decidió escoger la zona en la cual los cambios de temperatura del motor eran más notables. En la figura 4 se presenta el resultado de la segmentación manual que tuvo que efectuarse para todas las figuras que componen el banco de muestras controladas.

Finalmente se obtiene una base de datos con 120 imágenes (60 condición normal de operación y 60 de operación anómala), de las cuales se utiliza el 50% para entrenamiento y el 50% para validación.



Fig. 4. Imagen segmentada.

Para cada una de las pruebas realizadas se selecciona aleatoriamente un conjunto diferente de patrones de entrenamiento. La validación se realiza con el complemento de estos datos. La Tabla II presenta los resultados de cinco de las pruebas realizadas [P1...P5] que sintetizan el comportamiento observado de los clasificadores.

En la mayoría de las pruebas se puede observar que los clasificadores neuronales presentan un mejor desempeño que los clasificadores estadísticos, lo que permite comprobar la capacidad de generalización de este tipo de máquinas de aprendizaje.

El número de falsos negativos (condición normal clasificada como condición anómala) es mayor que el número de falsos positivos (condición anómala clasificada como condición normal). En la práctica este resultado es más favorable, ya que es preferible después de una inspección termográfica revisar un equipo sano que pasar por alto un equipo con alta probabilidad de falla.

TABLA II
RESULTADOS CLASIFICADORES CASO DE ESTUDIO NO.1

Prueba	BAYES		MLP		RBF	
P1	23	7	29	1	30	0
	0	30	2	28	3	27
	88.33	53	95	57	95	57
P2	17	13	26	4	29	1
	0	30	2	28	7	23
	78.33	47	90	54	86.66	52
P3	19	11	25	5	25	5
	0	30	5	25	0	30
	81.66	49	83.33	50	91.66	55
P4	27	3	28	2	25	5
	0	30	0	30	0	30
	95	57	96.66	58	91.66	55
P5	23	7	30	0	30	0
	0	30	3	27	1	29
	88.33	53	95	57	98.33	59

TABLA III
RESULTADOS CLASIFICADORES CASO DE ESTUDIO NO.2

Prueba	BAYES		MLP		RBF	
P1	18	6	10	14	18	6
	4	26	0	30	8	22
	81.48	44	74.07	40	74.07	40
P2	18	6	17	7	21	3
	3	27	1	29	6	24
	83.33	45	85.18	46	83.33	45
P3	19	5	16	8	16	8
	6	24	0	30	7	23
	79.63	43	85.18	46	72.2	39
P4	14	10	12	12	16	8
	5	25	0	30	8	22
	72.22	39	77.77	42	70.37	38
P5	17	7	12	12	18	6
	7	23	1	29	8	22
	74.07	40	75.92	41	74.07	40

En general los clasificadores presentan muy buen desempeño.

B. Caso de estudio No. 2: Clasificación de imágenes en un sistema robótico

Para este caso se utiliza una cámara de visión global Sony P73, la cual entrega imágenes digitales en formato RGB con una resolución de 640x480 píxeles. Estas imágenes se segmentaron manualmente obteniéndose nuevos tamaños iguales a 330x370, 370x330, 297x333, 333x297, 264x296, 296x264, 198x222, 222x198, 165x185 y 185x165. Se utilizaron diferentes tamaños porque las características extraídas son invariantes, entre otras transformaciones, al cambio de escala. Se definen dos clases: una correspondiente al robot Pioneer 3DX (C1) [8] y otra que comprende un conjunto de obstáculos conformados por cajas de diferentes tamaños, recipientes cilíndricos y sillas (C2).

La base de datos está conformada por 60 imágenes de la clase robot y 75 imágenes de la clase obstáculo (figuras 5 y 6). Se utilizan más imágenes de la clase objeto porque ésta es la que más variaciones presenta.

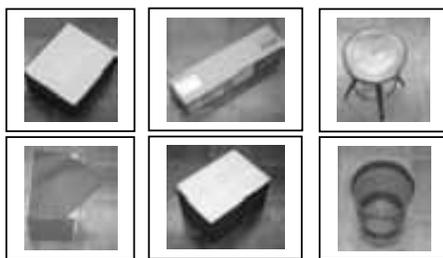


Fig. 5. Imágenes de la clase obstáculo.



Fig. 6. Imágenes de la clase robot

Del conjunto de 135 vectores de características se utilizó el 60% para entrenamiento y el 40% para validación.

De la Tabla III se puede observar que al igual que en el primer caso de estudio se obtienen mejores resultados para el clasificador MLP. Se observa para el clasificador Bayesiano y MLP una tendencia a clasificar incorrectamente la clase C1, lo que significa que se presenta una mayor dificultad en identificar la clase robot que la clase objeto. Esto se puede atribuir a que en la base de datos se empleó un mayor número de patrones correspondientes a la clase C2 (obstáculos) debido a que ésta presenta mayores variaciones.

IV. CONCLUSIONES

Se logró la implementación de tres tipos de clasificadores, utilizando las etapas básicas de un sistema para reconocimiento de patrones aplicado a dos problemas prácticos en diferentes campos: inspección termográfica y robótica.

Aunque los resultados obtenidos en las dos aplicaciones estudiadas son aceptables, los clasificadores en el caso de estudio No.2 presentaron desempeños más bajos que para el caso de estudio No.1. Estos resultados se pueden atribuir a que la composición de la escena de las imágenes adquiridas para cada caso difieren notablemente. En el sistema de inspección termográfica se tiene el mismo tipo de objeto (motor) bajo diferentes condiciones de operación (carga), lo cual se ve representado solamente en un cambio en los niveles de intensidad de zonas específicas de la imagen, mientras que para el sistema robótico las imágenes contemplan una mayor variedad de objetos que afectan en una mayor proporción la distribución de niveles de gris. En el último caso se puede pensar entonces en la utilización de descriptores que se obtengan a partir del color y/o la forma de los objetos de interés.

El diseño de un clasificador requiere una selección cuidadosa del conjunto de patrones, así como del tipo de clasificador. Para este trabajo se obtienen resultados satisfactorios utilizando imágenes provenientes de distintas bandas del espectro electromagnético, lo que verifica la robustez del tipo de descriptores elegidos, así como de los clasificadores implementados.

REFERENCIAS BIBLIOGRÁFICAS

- [1] L. F. Escudero. *Reconocimiento de Patrones*. Madrid, España. Paraninfo, 1977, Cap.1 y 2.
- [2] R. O. Duda, P. E. Hart, and D. G. Stok, *Pattern Classification*. Wiley, London, UK, 2001, Cap.1, 2 y 6.
- [3] F. Van Der Heijden et al. *Classification, Parameter Estimation and State Estimation*. John Wiley & Sons Ltd, England, 2004, Cap 2.
- [4] N. J. W. Morris, "Statistics of infrared images", *IEEE*, 1-4244-1180-7/07/, 2007.
- [5] G. Pajares y J. de la Cruz. *Visión por computador*. México: Alfaomega Grupo Editor, 2002, Cap.5, 9 y 14.
- [6] R. C. González. *Tratamiento digital de imágenes*. Wiomington, Delaware, E.U.A. Addison-Wesley Iberoamericana S.A., 1996, Cap. 3 y 7.
- [7] C. R. Bishop. *Neural Networks for Pattern Recognition*. Oxford. 1995, Cap. 2, 3 y 9.
- [8] Pioneer 3TM & Pioneer 2TM H8-Series. Operations Manual. ActivMedia Robotics. [En línea] <http://www.mobilerobots.com>.

Sistema de localización para un robot móvil utilizando visión global

Nini Juliana Vélez y Humberto Loaiza

Resumen—Este artículo describe la implementación de un sistema de visión para la localización de un robot móvil *Pioneer 3DX* dentro de un ambiente estructurado. Para validar el sistema se realiza una tarea de navegación que combina la información proveniente de la cámara para visión global y la del sistema sensorial local del robot.

Palabras Clave—Sistema robótico, procesamiento de imágenes, modelos de color, captura, calibración, filtrado, segmentación, umbralización, navegación, arquitectura basada en comportamientos.

I. INTRODUCCIÓN

PARTE de la investigación en robótica móvil se centra en poder mover un robot de manera autónoma interactuando con su entorno y en la realización de una tarea específica. La visión por cámara ofrece la posibilidad de extraer información del ambiente en el que se encuentre el robot permitiendo realizar tareas de localización, reconocimiento y navegación.

El presente trabajo se enfoca en la implementación de un sistema robótico conformado por el robot *Pioneer 3DX* y un sistema de visión artificial que permite dar solución al problema básico de navegación de un robot móvil: *Dada la posición inicial del robot y la posición final a la que se quiere llegar, se debe generar un camino que evite los obstáculos constituido por una secuencia de posiciones y orientaciones del robot, desde la posición y orientación inicial hasta una orientación y posición final.* Esta tarea implica minimizar los tiempos de captura y generar algoritmos de procesamiento de imágenes efectivos y robustos ante condiciones no controladas de iluminación.

El sistema de visión global está constituido por una cámara *Sony EVI D30* fija situada sobre la escena, un adaptador de video VA11A de X10 y un conjunto de algoritmos implementados en *Visual Studio .NET 2003* utilizando las bibliotecas de libre distribución *OpenCV*. Su principal misión es reconocer y localizar los objetos (robot y puntos de llegada), utilizando el color como característica fundamental. Para ello este sistema realiza los siguientes pasos: captura, calibración, pre-procesamiento (filtrado), segmentación de la imagen (basada en técnicas de umbralización) y obtención de las posiciones de los centros de masa.

La información obtenida del sistema de visión artificial se utiliza para generar una ruta libre de colisiones, combinándola con la información proveniente de los ultrasonidos del robot, desde la posición y orientación inicial del robot hasta la posición de la meta a través de las funciones disponibles en el *API ARIA* para el control y supervisión del estado del robot *Pioneer 3DX*, produciendo como salida una secuencia de comandos que son transmitidos al robot a través del puerto serie. La arquitectura de control utilizada está basada en comportamientos.

El resultado final es un sistema robótico capaz de tomar decisiones mientras se realizan un conjunto de tareas complejas como la percepción y análisis del entorno para la determinación de una ruta a seguir.

Este trabajo se realiza en el marco de la propuesta de investigación de maestría de la autora: *Sistema de robots móviles cooperativos de realidad aumentada*, con la cual se pretende desarrollar un sistema robótico autónomo cooperativo heterogéneo de realidad aumentada conformado por un robot *Pioneer 3DX*, un robot virtual y un sistema de adquisición de imágenes con una cámara para visión global.

A continuación se describe el sistema robótico implementado desde el punto de vista hardware y software. Posteriormente se consignan las pruebas y resultados y finalmente se presentan las observaciones y conclusiones.

II. DESCRIPCIÓN DE LA PLATAFORMA DE EXPERIMENTACIÓN

El sistema robótico implementado, sobre el cual se realizaron las pruebas, está conformado por el robot móvil *Pioneer 3DX*, que se identifica con dos marcas de color en su parte superior (una roja y otra amarilla), unos puntos de llegada representados por figuras planas de color azul y magenta, unas cajas blancas que conforman los obstáculos y una cámara para visión global ubicada a 245 cm del piso (figura 1). El área cubierta por la cámara es de 219x161 cm². La comunicación entre el robot y el computador se realiza a través del puerto serie.

El robot contiene una puerta de acceso a la batería, 1 batería, 2 ruedas, motores con encoders, un anillo de 8 ultrasonidos, 1 microcontrolador Hitachi H8S/2357 y las respectivas tarjetas de control [1].

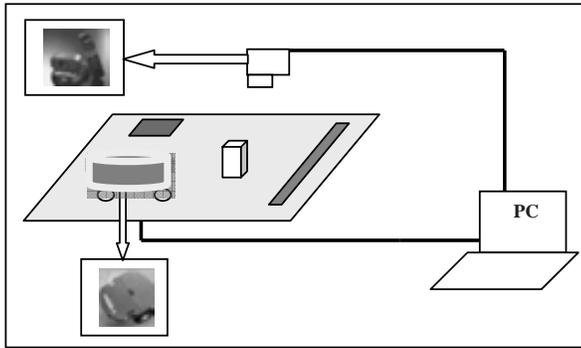


Fig. 1. Sistema robótico implementado.

III. DESCRIPCIÓN FUNCIONAL DEL SISTEMA: SOFTWARE

En esta sección se describen y se especifican los diferentes módulos software que componen el sistema, así como su interacción. Estos módulos incluyen el de visión (calibración de cámara, captura de vídeo, mejoramiento de imágenes, segmentación, formación de blobs, reconocimiento, identificación y localización de objetos de interés en la escena) y el de control del robot (comunicación con el robot, actualización de información sensorial, comportamientos reactivos y deliberativos – evasión de obstáculos y generación de trayectorias -).

Todo el sistema fue programado utilizando hilos POSIX [2] en Visual Studio .NET 2003. Se emplearon funciones de OpenCV y Matlab para la calibración geométrica de la cámara, captura y procesamiento de imágenes, así como el API ARIA para el control y supervisión del estado del robot Pioneer 3DX.

A. Sistema de visión

En la figura 2 se puede observar del diagrama de bloques del sistema de visión que se implementa en un hilo.

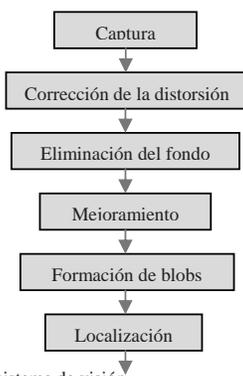


Fig. 2. Módulos del sistema de visión.

- *Captura*: Para obtener imágenes desde el dispositivo de captura se utilizó la librería HighGUI de OpenCV que permite definir la resolución y la velocidad de captura. A través del flujo de vídeo constante establecido en 30 fps se obtienen frames de 320x240 píxeles en formato BGR de 24 bits.

- *Corrección de la distorsión*: Con el fin de corregir la distorsión causada sobre la imagen por defectos ópticos de la cámara se obtiene el modelo interno de ésta a través del *Camera Calibration Toolbox for Matlab*, el cual utiliza un modelo muy similar al utilizado por Heikkilä y Silvén en la Universidad de Oulu [3]. Para ello se construyó un patrón de calibración de tipo tablero de ajedrez de 15x10 cuadrados de lado 10 cm.

Con los parámetros intrínsecos calculados se utilizan las funciones de la librería OpenCV para realizar la transformación entre el sistema de coordenadas de la imagen distorsionada y una imagen no distorsionada.

- *Eliminación del fondo*: Este módulo de segmentación es el responsable de separar los píxeles de la imagen que pertenecen a los objetos de interés de los que no [4][5], es decir, elimina el fondo y deja las marcas del robot y los puntos de llegada.

Debido a las condiciones no controladas de iluminación de la escena se transforma la imagen al espacio YCbCr [6], el cual permite aislar las componentes de intensidad y color. Se calculan los histogramas de los planos Cb y Cr para extraer de forma automática los picos en cada uno de ellos. Posteriormente se realiza una umbralización bimodal, de tal forma que el fondo se deja blanco y los objetos de interés mantienen su color original.

- *Mejoramiento*: Este módulo se encarga de aplicar un filtro mediano a la imagen para eliminar ruido del tipo *salt&pepper*. El tamaño de este filtro es definido desde la interfaz principal de usuario. Preferiblemente se utilizan tamaños de ventana de 3 o 5 píxeles de lado.

Posteriormente se realiza una operación morfológica *closing* que consiste en una erosión seguida de una dilatación con un elemento estructural circular, dado que las marcas del robot son circulares [7].

- *Formación de blobs*: Después de realizar el mejoramiento a la imagen se realiza una transformación al espacio HSV, que al igual que el espacio YCbCr permite aislar las componentes de intensidad y color, con la ventaja de ofrecer en el plano H rangos del mismo tamaño para los colores rojo, amarillo, verde, cyan, azul y magenta. Luego se obtienen de manera automática a partir de los histogramas del plano H y S los límites en este espacio de color para el rojo, amarillo, azul y magenta que corresponden a las marcas del robot y a los puntos de llegada respectivamente, de esta forma se realiza una umbralización multimodal [8] [9].

Posteriormente se forman regiones a partir de los píxeles que han sido segmentados llamadas blobs. Esto se realiza con las funciones de la librería *cvblobslib*.

- *Localización*: Este módulo es el responsable de posicionar a los objetos identificados en un plano de referencia. Para ello encuentra los centros de masa de los blobs obtenidos en la etapa anterior. Los resultados obtenidos se marcan sobre la imagen resultante y su valor numérico se muestra en la interfaz de usuario. Esta información es actualizada con cada cuadro procesado.

Finalmente se realiza la transformación de posiciones de píxeles a mm, teniendo en cuenta que la imagen de 320x240 píxeles corresponde a un área de 2190x1610 mm². También se

obtiene la orientación del robot en grados entre -180° y 180° con base en el sistema de coordenadas del mismo. Esta información se envía al sistema de control del robot para que realice la tarea de navegación.

B. Sistema de control del robot

Este sistema implementa una arquitectura de control híbrida basada en comportamientos [10] utilizando el API ARIA.

En el bajo nivel ARIA tiene un diseño de cliente-servidor. Según este esquema el robot está gobernado por un microcontrolador que hace las veces de servidor (ARCOS). Ese servidor establece un diálogo a través del puerto serie con la aplicación (escrita con ARIA en C++), la cual actúa como cliente. En ese diálogo se envían al cliente las medidas de ultrasonido, odometría, etc, y se reciben las órdenes de actuación o comando para los motores.

Para definir comportamientos complejos como componentes independientes y reusables, a partir de bloques constructivos simples para un control de movimiento del robot dinámico y continuo, ARIA dispone de un sistema de *Acciones* de alto nivel. Las acciones son objetos individuales que independientemente generan requerimientos de movimiento los cuales son evaluados y después combinados en cada ciclo de trabajo para producir un conjunto final de comandos de movimiento. Las acciones son evaluadas por el resolver de acciones en orden descendente de prioridad (la de mayor prioridad primero, la de menor prioridad después).

De esta forma se utilizan comportamientos reactivos con mayor prioridad que los deliberativos.

En este caso se definen los comportamientos de mayor a menor prioridad en la figura 3.

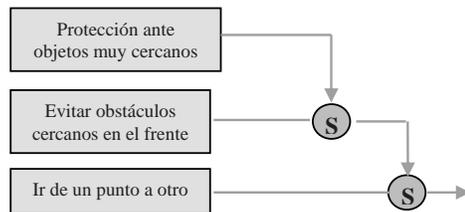


Fig. 3. Comportamientos del sistema de control del robot.

El primer comportamiento que presenta la prioridad más alta le permite al robot protegerse de colisiones haciendo que éste pare cuando detecte obstáculos con los ultrasonidos a 100 mm de distancia. Posteriormente el robot gira 45° .

El siguiente comportamiento le permite al robot evadir obstáculos controlando rotación y traslación. La distancia de detección establecida para llevar a cabo esta acción está en 150 mm, el ángulo de giro mientras evade el obstáculo está en 15° . La velocidad se ha establecido en 200 mm/seg. La información de distancias se obtiene del grupo de ultrasonidos.

El último comportamiento le permite al robot desplazarse hacia una posición establecida sin evitar obstáculos. Se ha establecido una precisión para llegar a la meta de ± 50 mm, una velocidad de desplazamiento igual a 200 mm/seg, una velocidad de giro en 400 mm/seg y ángulos de giro en 7° . La

posición y orientación inicial, así como la posición del punto de llegada son datos que se obtienen del sistema de visión.

La combinación de estos comportamientos y su implementación en un hilo independiente le permiten al robot realizar una tarea de navegación evadiendo obstáculos.

La interfaz de usuario a través de la cual se pueden ingresar parámetros como el tamaño del filtro mediana y el punto de llegada y que permite visualizar resultados se observa en la figura 4.



Fig. 4. Interfaz de usuario del sistema.

IV. PRUEBAS Y RESULTADOS

El primer grupo de pruebas se centraron en el módulo de visión.

Se ubicaron los objetos de interés: robot y marcas de llegada en diferentes posiciones y orientaciones sobre la escena.

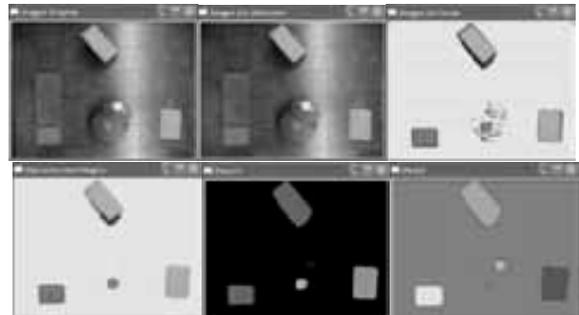


Fig. 5. Imágenes obtenidas en cada etapa del sistema de visión. Pos. del robot: (169, 182) píxeles. Orientación: -115.91° .

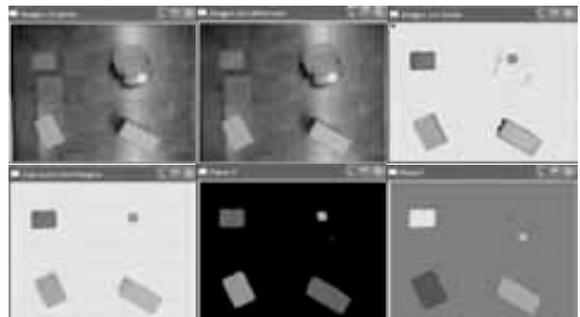


Fig. 6. Imágenes obtenidas en cada etapa del sistema de visión. Pos. del robot: (212, 60) píxeles. Orientación: 117.22° .

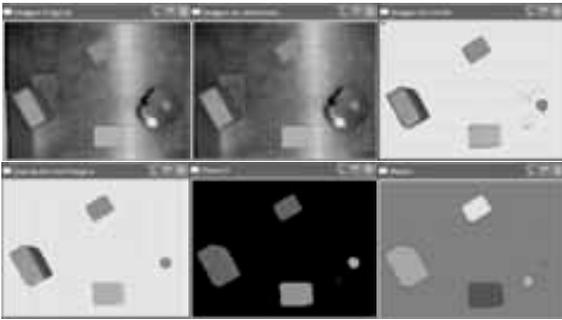


Fig. 7. Imágenes obtenidas en cada etapa del sistema de visión. Pos. del robot: (282, 146) píxeles. Orientación: 50.19°.

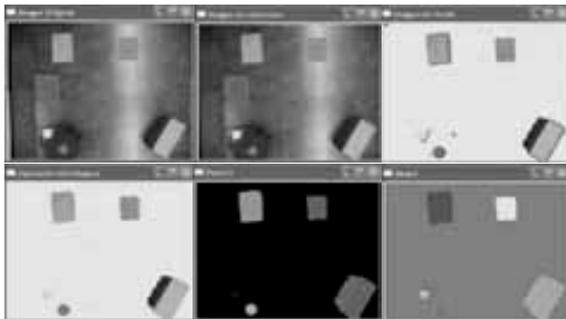


Fig. 8. Imágenes obtenidas en cada etapa del sistema de visión. Pos. del robot: (96, 223) píxeles. Orientación: -43.96°.

Las figuras 5 a 8 ilustran los resultados obtenidos en cada una de las etapas del sistema de visión, para diferentes posiciones y orientaciones del robot. Las posiciones obtenidas se compararon con la posición manual del centroide, obteniendo una precisión de ± 2 píxeles tanto en x como en y . Las orientaciones se encuentran en los rangos esperados.

La etapa de preprocesamiento (eliminación del fondo y mejoramiento de la imagen) permite obtener muy buenos resultados a pesar de las condiciones poco favorables de iluminación lo que confirma la ventaja que presenta el espacio de color YCbCr para aislar la información de color de la de intensidad.

Posteriormente se realizaron las pruebas del módulo de visión integrado con el de control del robot para realizar la navegación de un punto a otro evadiendo obstáculos. La marca de llegada se selecciona desde la interfaz de usuario: azul o magenta. La figura 9 ilustra cuatro de las pruebas realizadas sobre el sistema completo. Los frames de la izquierda ilustran los puntos de partida (primer frame adquirido) y sobre ellos se han realizado marcas circulares correspondientes a los centros de masa de las marcas del robot y del punto de llegada. En los frames de la derecha se ha marcado la trayectoria seguida por el robot (último frame adquirido). La línea cyan corresponde al seguimiento de la marca roja y la línea verde corresponde a la marca amarilla. En todas las pruebas realizadas el robot llega con éxito al punto de llegada y el sistema de visión realiza el respectivo seguimiento del mismo.

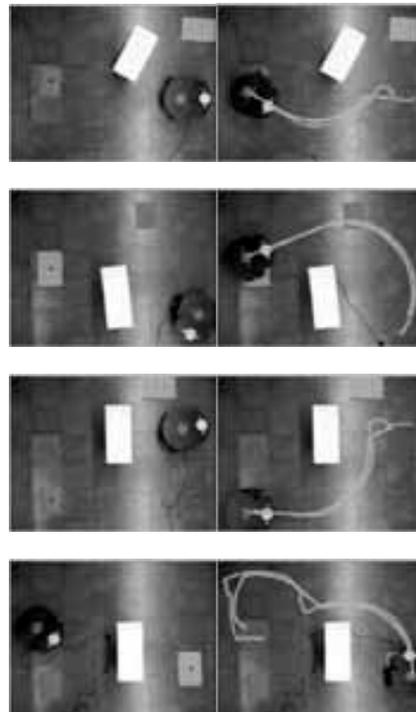


Fig. 9. Trayectorias seguidas por el robot durante cuatro pruebas en la tarea de desplazamiento de un lugar a otro con evasión de obstáculos.

V. CONCLUSIONES

Se logró implementar un sistema de localización para un robot móvil que permite obtener su posición y orientación en una tarea de navegación con evasión de obstáculos. El sistema está implementado en Visual Studio .NET 2003 utilizando una arquitectura multihilos. El módulo de visión realiza la captura de imágenes en formato BGR de 24 bits a 30 fps. El módulo de control utiliza una arquitectura basada en comportamientos que combina la información de la cámara para visión global y la información sensorial local del robot y hace que el robot se mueva a una velocidad máxima de 200mm/seg. Este sistema se convierte en una primera etapa funcional para el proyecto de maestría titulado *Sistema de robots móviles cooperativos de realidad aumentada*, que actualmente está desarrollando la autora.

Durante el procesamiento de las imágenes obtenidas se utilizaron dos modelos de color, el espacio YCbCr y el HSV. El primero arrojó los mejores resultados para la etapa de eliminación del fondo de la imagen dadas las condiciones no controladas de iluminación de la escena. La segmentación de los blobs por color no generó buenos resultados en este espacio de color debido a que en los planos Cb y Cr no se obtiene el mismo rango de valores para los colores utilizados en el sistema (rojo, amarillo, azul y magenta), mientras que en el plano H y S sí.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Pioneer 3TM & Pioneer 2TM H8-Series. Operations Manual. ActivMedia Robotics. [En línea] <http://www.mobilerobots.com>.
- [2] F. Ceballos. *Programación orientada a objetos con C++*. México: Alfaomega Grupo Editor, 2004, Cap.9.
- [3] J. Heikkila, O. Silvén. *A four step Camera Calibration Procedure with implicit image correction*. [En línea] http://www.vision.caltech.edu/bouguetj/calib_doc/papers/heikkila97.pdf
- [4] R. González. *Tratamiento Digital de Imágenes*. Wilmington, Delaware, E.U.A.: Addison-Wesley Iberoamericana S.A., 1996, Cap. 1,2,4 y 7.
- [5] G. Pajares y J. de la Cruz. *Visión por computador*. México: Alfaomega Grupo Editor, 2002, Cap.1,3,4,5,6,7 y 10.
- [6] L. Martínez. *Sistema de visión para el equipo de robots autónomos del ITAM*. Trabajo de grado (Ingeniero en Computación). México, 2004. Instituto Tecnológico Autónomo de México.
- [7] B. Li; E. Smith; H. Hu y L. Spacek. *A Real-time Visual Tracking System in the Robot Soccer Domain*. Proceedings of EUREL Robotics-2000, Salford, England, 12th - 14th April 2000.
- [8] J. Vargas y A. Hurtado. *Vehículo autónomo controlado por visión*. Santiago de Cali, 2003. Trabajo de grado (Ingeniero Electrónico). Universidad del Valle. Facultad de Ingeniería.
- [9] C. Villa y J. de la Cruz. *Cálculo de trayectorias y control de posición mediante procesamiento de imágenes*. Santiago de Cali, 2004. Trabajo de grado (Ingeniero Electrónico). Universidad del Valle. Facultad de Ingeniería.
- [10] R. Arkin. *Behavior-Based Robotics*. Cambridge, Massachusetts: The MIT Press, 1998.

Sistema de seguimiento de un objeto en movimiento usando arquitectura de visión foveal

Natalia Martínez, Javier Pinto Rojas, Carlos Parra Ph.D., Javier Villegas M.Sc. candidate PhD.
Departamento de Ingeniería Electrónica Pontificia Universidad Javeriana, Bogotá, Colombia

Abstract—Este artículo describe un sistema desarrollado por medio de visión artificial, para realizar el seguimiento de un objeto en movimiento usando arquitectura de visión foveal. El sistema está dividido en cinco etapas: adquisición, fovealización, seguimiento, estimación de posición y salida fovealizada. De esta última etapa se obtiene un video fovealizado del objeto en seguimiento, con la mayor resolución centrada en el objeto de interés, a partir del cual se caracteriza el sistema y se analizan ventajas del uso de esta arquitectura como una solución eficiente al problema de seguimiento en visión artificial.

Index Terms— Fotorreceptores ópticos, fovea, mapa de fovealización, resolución variable, MeanShift, segmentación por color, segmentación por movimiento, sistema de seguimiento.

I. INTRODUCCIÓN

En los seres vivos más evolucionados, la retina muestra una distribución no uniforme de celdas visuales ubicadas en la fovea. Con la distribución espacio-variante de las células receptoras en los ojos de los seres vivos se consigue una resolución que también es espacio-variante. De esta manera, se tiene una mayor resolución en el centro (fovea) de la imagen, que es donde se tiene el área de interés, esto manteniendo un campo de visión suficientemente amplio para apreciar los cambios en la periferia, es por esto que la principal característica es la inherente reducción de información debida a la distribución espacio variante[1]. A partir de investigaciones y simulaciones se logra asemejar la distribución espacio variante desarrollando software que emula el comportamiento de los seres vivos. Lo anterior con el fin de obtener mejoras en aplicaciones como transmisión de video donde se necesita que el ancho de banda usado sea el menor posible y que el consumo de energía del sistema también se reduzca [2].

II. DESCRIPCIÓN

A. Etapa de adquisición

La etapa de adquisición se realizó por medio de una cámara web. Mediante ella se ejecutó el desarrollo y las pruebas de la presente investigación, dado que las aplicaciones de este tipo de sistemas de fovealización son usadas en su mayoría en dispositivos portátiles, se requiere poco peso y versatilidad. La captura del video se lleva a cabo directamente por medio del programa Visual C++ .NET, usando la librería CvCam de la biblioteca OpenCV. El video capturado tiene un formato estándar *.AVI a una velocidad de 30 cuadros por segundo y un tamaño de salida de 640 X 480 píxeles.

B. Etapa de fovealización

La fovealización es un proceso en el cual se realiza un muestreo no uniforme sobre una imagen adquirida, la cual permite ver con gran resolución una zona específica (zona de interés), y con baja resolución en la periferia.

La fovealización puede realizarse de varias maneras. Una de ellas consiste en usar dispositivos de captura distribuidos fovealmente en la etapa de adquisición. Otra manera se refiere a la fovealización de la imagen luego de ser adquirida. En este anterior caso, la resolución uniforme del video adquirido se alterará utilizando mapas de fovealización, los cuales toman el video capturado y de acuerdo a un patrón preestablecido reemplazan los píxeles más alejados del centro de la imagen por el promedio de los píxeles cercanos. En esta etapa del sistema se diseñaron tres mapas de fovealización, con el fin de escoger uno de ellos cuyo desempeño se adaptara a la aplicación específica de seguimiento de objetos. En estos mapas la fovealización se inicializa en el punto central de la imagen, correspondiente a la posición $[x=320, y=240]$.

C. Etapa de seguimiento

La etapa de seguimiento del sistema fue diseñada con el fin de seguir un único objeto en movimiento por medio de visión monocular pasiva. Esto significa que la cámara en ningún momento tendrá movimiento alguno. La técnica utilizada para el seguimiento del objeto es el algoritmo de *MeanShift* modificando a imágenes de resolución variable. Este proceso se lleva cabo reubicando cuadro a cuadro (*frame a frame*) el

punto de mayor resolución (fóvea) sobre el objeto en movimiento.

Al inicializar el sistema de seguimiento, la fóvea se ubicará en el centro de la imagen. Esto ocurre para tener una apreciación uniforme que facilite la escogencia del objeto a seguir.

D. Etapa estimación de posición

A partir del resultado de la segmentación que se realiza en la etapa anterior, se obtiene la pareja de coordenadas (x,y) del centro del objeto en seguimiento. Dicho punto central indica la posición del objeto en movimiento. Aquellas coordenadas se envían como parámetro de entrada a la etapa de fovealización. De esta forma, se establece el punto donde se ubicará la fóvea en el siguiente *frame*.

E. Salida fovealizada

La salida es un video de extensión *.AVI fovealizado de dimensiones 640 x 480 píxeles. En este video se da una mayor resolución en la región donde se encuentre el objeto en movimiento

III. MAPAS DE FOVEALIZACIÓN

A partir de una imagen cartesiana adquirida con resolución uniforme, se logra obtener una imagen fovealizada, reemplazando los píxeles más alejados del centro de la imagen por su promedio con los píxeles cercanos de acuerdo a un patrón preestablecido o mapa de fovealización. El promedio de los píxeles tenidos en cuenta aumenta proporcionalmente con la distancia al centro de la imagen y cada región promediada es reemplazada por el valor medio del conjunto de píxeles. Obteniendo así, la imagen foveada. La forma, el tamaño y la posición de las regiones que son promediadas están determinados por el mapa de fovealización utilizado [3].

A. Mapa de baldosa cuadrada

Es un mapa de fovealización que muestra embaldosados de la imagen con figuras geoméricamente similares, en el cual la resolución es función proporcional a la proximidad al centro de la imagen, es decir que tenemos una alta resolución sobre la zona de interés. Esta transformación duplica su tamaño cuando se aleja del centro (fóvea) de la imagen según la zona de resolución. (Figura 1) [3].

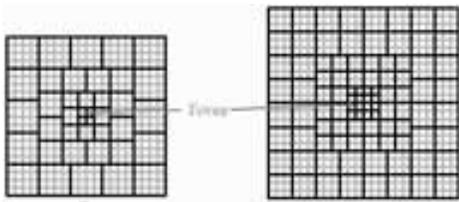


Figura 1. a) y b), Mapas de baldosa cuadrada [3].

Esta transformación geométrica mueve un píxel de coordenadas (x,y) a una nueva posición (x', y') , dada por ecuaciones de transformación de la forma:

$$\begin{aligned} x' &= Tx(x, y) \\ y' &= Ty(x, y) \end{aligned} \quad (1)$$

Donde T_x y T_y son polinomios en x y y , en su forma más simple son lineales en x y y como se muestra en la siguiente ecuación

$$x' = a_0x + a_1y + a_2 \quad (2)$$

$$y' = b_0x + b_1y + b_2$$

Estos polinomios son modificables de la forma que se desee según la transformación requerida.

B. Mapa de baldosa cruz

Pueden generarse mapas de fovealización con una configuración muy similar a la baldosa cuadrada, en la Figura 2 se ve una baldosa generadora en cruz duplicando su tamaño cuando se aleja del centro (fóvea)

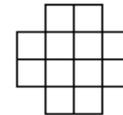


Figura 2. Baldosa generadora en cruz [3]

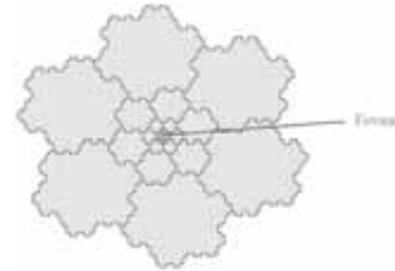


Figura 3. Mapa generado con baldosa en cruz [3].

C. Mapa Log-Polar

Las imágenes Log-Polares pueden ser obtenidas a través de sensores especializados o a partir de la transformación de imágenes cartesianas, ya sea por hardware especializado o mediante software como es el caso de este proyecto. Primero se explicará la notación básica de las imágenes cartesianas y Log-Polares, así como también es necesario explicar la ley matemática que se usa para la transformación del plano cartesiano al plano Log-Polar son:

$$\begin{aligned} \begin{bmatrix} \xi \\ \theta \end{bmatrix} &= \begin{bmatrix} \log(\sqrt{x^2 + y^2}) \\ \arctan\left(\frac{y}{x}\right) \end{bmatrix}, \text{ donde} \\ z &= x + yj \\ \xi &= \log|z| \end{aligned} \quad (3)$$

En el modelo logarítmico una línea que esté ubicada sobre un anillo concéntrico, será representada como una línea recta en el eje θ que representa el ángulo en coordenadas polares de una imagen cortical. Una línea recta que parta desde el centro de la imagen a su vez se representará en el eje de la magnitud del radio, como se observa en la Figura 4.

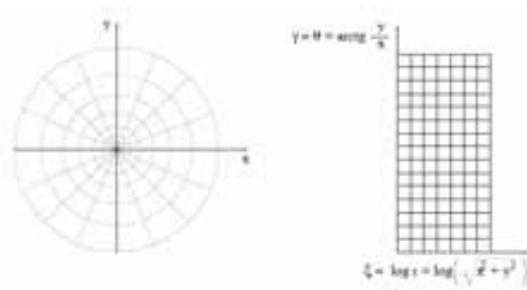


Fig 4. (a) Representación Log-Polar (b) Representación cortical [1]

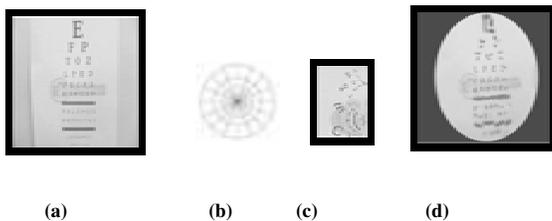


Fig 5. (a) Imagen cartesiana original, (b) Representación Log-Polar (c) Imagen Cortical, (d) Imagen retinal (Después de realizada la transformada inversa).

Este método logarítmico de representación de la información de una imagen tiene algunas propiedades como son:

- Reducción selectiva de la información
Mayor resolución en la zona central, obteniendo un campo de visión en la periferia con poca resolución
- Rotaciones
Las rotaciones en el plano retínico se convierten en simple traslaciones en el plano cortical, si el eje óptico coincide con el centro de la fovea. Por lo que se pueden desarrollar sistemas invariantes a rotaciones.
- Escalamiento
Si el eje óptico coincide con el centro de la fovea, en caso de que se escale la imagen en el plano retinal, se traduciría en el plano cortical en una traslación.
- Alta resolución en el objeto de interés y reducción del número de píxeles a analizar.
- Focalización en un objeto de interés, reduciendo los datos del entorno [1][4].

IV. METODOS DE SEGUIMIENTO

A. Segmentación por movimiento usando diferencia de Imágenes (Básico en escala de Gris)

La detección de movimiento se lleva a cabo, modificando el método de substracción de fondo, que ha resultado ser una técnica rápida y simple. Para ello es necesario almacenar una

estimación del fondo de la escena para poder extraer el objeto que entra en la misma. El método clásico se basa en promediar el nivel de gris de un número determinado de imágenes, al que denominaremos ventana temporal, de acuerdo con la siguiente expresión:

$$B(x, y, t) = \frac{1}{N} \sum_{t'=t-N}^t I(x, y, t') \quad (4)$$



Fig. 6: Secuencia de imágenes ejemplo de implementación del algoritmo de segmentación por movimiento

Donde $B(x,y,t)$ es el fondo estimado en la posición (x,y) en el instante t , $I(x, y, t')$ representa el valor de la imagen en la posición (x,y) en el instante t' , y N el número de imágenes de la ventana temporal. En una matriz binaria $M(x,y,t)$ se almacenan los valores que presenta valores distintos a cero solo en aquellas posiciones donde se ha detectado movimiento, es decir:

$$M(x, y, t) = 1 \quad \text{si} \quad |B(x, y, t-1) - I(x, y, t)| \geq U$$

$$M(x, y, t) = 0 \quad \text{en otro caso} \quad (5)$$

Donde $B(x,y,t-1)$ se refiere al fondo previamente estimado, $I(x,y,t)$ es la imagen actual y U un umbral estimado heurísticamente. El valor de U influye en la detección de movimiento en una imagen dada, porque los objetos o la porción de ellos cuyo nivel de gris sea similar al del fondo no serán detectados. En la figura 6, se observa la figura que detalla el funcionamiento del algoritmo desarrollado [5].

B. Segmentación por movimiento usando diferencia de imágenes con distribución Gaussiana.

Con este método se modela cada píxel como una mezcla de distribuciones Gaussianas aplicando un esquema selectivo para ir actualizando la imagen referencia del fondo. Las distribuciones Gaussianas son evaluadas para determinar cuáles píxeles hacen parte del fondo. Este método es más robusto ante los cambios de iluminación del fondo y otras situaciones como movimientos pequeños y repetitivos.

Es por esto que cada píxel en la escena es modelado como una mezcla de k distribuciones Gaussianas para que basándose en la persistencia y la varianza de cada una de las Gaussianas se determine si el píxel en determinado tiempo t pertenece o no al fondo. Se consideran los valores de cada píxel cambiando a través del tiempo. La función de densidad de probabilidad Gaussiana a partir de la cual se clasifican los píxeles como del fondo o del objeto en movimiento se define en la ecuación 6.

$$\eta(X_t, \mu, \Sigma) = \frac{1}{|\Sigma|^{\frac{1}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1} (X_t - \mu)}$$

(6)

El número de distribuciones por píxel es determinado por la capacidad de memoria del computador. En general se usa este valor de 3 a 5. La matriz de covarianza está definida como

$$\Sigma_{i,t} = \sigma_{i,t}^2 I$$

(7)

Para ir actualizando, el valor de cada píxel X_t se compara con cada una de las k distribuciones Gaussianas, hasta que se encuentre una coincidencia. Se encuentra que hay coincidencia si el valor del píxel tiene una desviación estándar menor que 2,5 de la distribución. Este sería el valor de umbralización. La primera distribución que coincida se actualiza usando las siguientes ecuaciones.

$$\begin{aligned} \omega(k,t) &= (1 - \alpha)\omega(k,t) - 1 + \alpha(M_{k,t}, t) \\ \mu_c &= (1 - \rho)\mu_c - 1 + \rho X_t \\ \sigma_c^2 &= (1 - \rho)\sigma_c^2 - 1 + \rho(X_t - \mu_c)^T (X_t - \mu_c) \end{aligned}$$

(8)

Donde α se denomina factor de aprendizaje y la constante de tiempo $1/\alpha$ determina la velocidad con la cual cambian los parámetros de distribución. Si ninguna de las k distribuciones coincide con el valor actual del píxel, la distribución menos probable es reemplazada con el valor de la media del píxel actual, una varianza inicial de un valor alto y un peso de bajo valor. Donde T es el umbral de la mínima fracción del modelo del fondo, es decir la mínima probabilidad de que el fondo esté presente en la escena. Las k distribuciones son ordenadas de acuerdo al valor ω/σ . Este valor incrementa cuando la distribución toma más peso y disminuye al aumentar la varianza. Cuando se compara el nuevo valor de un píxel se empezará a comparar con las distribuciones que tengan este valor más alto. Así las que representen mas el fondo serán comparadas primero, y las de menor valor en caso de que no coincida con el valor del píxel serán reemplazadas como se explico anteriormente. Las primeras B distribuciones son escogidas como el modelo del fondo [6][7].

$$B = \operatorname{argmin}_b \left(\sum_{k=1}^b \omega_k > T \right)$$

(9)

C. Segmentación por color usando el algoritmo CamShift

El algoritmo de *CamShift* se basa en la técnica de *MeanShift*. La distribución de la intensidad de cada componente de color puede ser visto como una función de densidad de probabilidad. El vector *MeanShift* es la diferencia entre la media de la función de probabilidad en una área local y el

centro de esta región. En términos matemáticos, el vector de *MeanShift* asociado con una región S_x centrado en x puede ser representado como:

$$\bar{V}(x) = \frac{\int_{y \in S_x} p(y) (y - x) dy}{\int_{y \in S_x} p(y) dy}$$

(10)

Donde $p(y)$, es la función de densidad de probabilidad; la técnica de *MeanShift* significa que el vector es proporcional al gradiente de la función de probabilidad de $\nabla p(x)$ y reciproca a la densidad de probabilidad $p(x)$.

$$\bar{V}(x) = c \frac{\nabla p(x)}{p(x)}$$

(11)

Donde c , es una constante y al estar el vector *MeanShift* a lo largo de la dirección de la máxima densidad de probabilidad, entonces se puede aprovechar esto para encontrar la localización actual de la máxima densidad.

Un sólo paso del *MeanShift* es realizado definiendo una vecindad β alrededor del actual punto en el espacio. Esta vecindad se calcula por el método de estimación de densidad de ventana de *Parzen*. El procedimiento es calcular la densidad de los puntos que caen debajo de esta vecindad β , y luego desplazar hacia la media mediante un movimiento de la ventana para que esté centrada alrededor del punto. Por lo tanto, el paso fundamental es calcular la media de todos los puntos de la vecindad β alrededor del punto dado; específicamente, si P es un punto del espacio característico, entonces la media $Mh(P)$ de P es definida usando un *kernel* radialmente simétrico K , con un radio h en el espacio característico con un perfil monótonamente decreciente [8].

La utilización de *CamShift* en el seguimiento de objetos de color en una escena de video empieza con obtener una distribución de probabilidad del color deseado. Primero se crea un modelo en el sistema HSV (*Hue, Saturation, Value*), el cual corresponde a la proyección del sistema RGB (*Red, Green, Blue*) a través de la diagonal principal que viaja del color blanco al negro. El espacio HSV separa el color, la saturación de color y el brillo de la imagen, con lo cual es posible crear un histograma de una dimensión en el canal H (*Hue*) [8].

El color derivado de los píxeles del objeto ubicado en la imagen, genera un histograma de una dimensión. Cuando el muestreo sobre toda la imagen se ha realizado el histograma es completo. El color generado en el histograma se segmentará en la imagen, a la cual se le aplicará la técnica de *MeanShift* para el seguimiento del móvil.

V. ANALISIS DE RESULTADOS

Las pruebas realizadas fueron hechas en un PC Centrino Core 2 Duo de 1.5GHz, memoria RAM de 2 GB, usando una

webcam integrada de 640 x 480, con una velocidad de 30 cuadros por segundo. El código ha sido escrito en Visual C++ .NET 2003, biblioteca OpenCv.

Como se observa en la figura 7, el algoritmo que toma menos tiempo de procesamiento es el mapa Log Polar. El mapa de fovealización de baldosa cuadrada es igualmente rápido y fácilmente adaptable a otras aplicaciones. Caso contrario ocurre con el mapa de fovealización basado en un patrón de baldosa tipo cruz que toma un gran tiempo de procesamiento, lo cual se debe a la necesidad de cargar archivos externos correspondientes a las máscaras y patrones de máscaras que se aplican a la imagen. La medida de estos tiempos se realizó a una tasa de adquisición de 30 cuadros por segundo, durante un igual número de ciclos de captura para todos los casos.

Otra razón por la cual los tiempos de procesamiento de los algoritmos de los mapas Log Polar y Baldosa cuadrada son significativamente menores es debido a que el mapeo se basa en una transformación matemática formal para pasar una imagen de resolución uniforme a una de resolución variable.

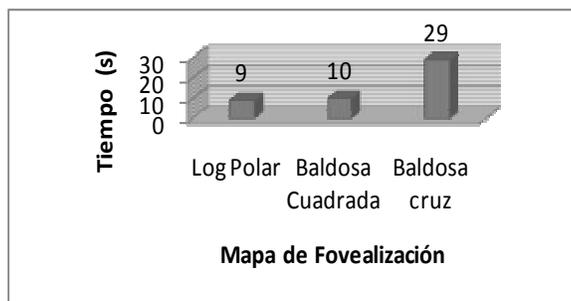


Figura 7. Comparación tiempos de ejecución algoritmos de fovealización

En el mapa de baldosa cuadrada la reducción del número de píxeles entre la imagen original y foveada se diseñó con tres etapas de resolución, de esta manera se tiene una reducción de 6 en el número de píxeles en la imagen foveada con respecto a la imagen original.

El mapa baldosa tipo cruz se diseñó con un total de 90 receptores, correspondientes al número de máscaras. Podría implementarse un mapa con una cantidad considerable de máscaras, lo cual conllevaría tener más resolución. Sin embargo dichos patrones harían aun más lento el algoritmo debido al diseño externo de los mismos.

El mapa Log Polar no presenta una reducción significativa en el número de receptores, debido a que el número de filas de la imagen cartesiana se mapean en el espacio logarítmico en el número de círculos concéntricos y el número de columnas son las secciones angulares, con lo cual existe mayor concentración de los receptores en la fovea y pocos en la periferia.

El mapa de fovealización elegido es el de la figura 4. Esta elección se debe en su mayoría a la aplicación objeto de la presente investigación, la cual requiere un bajo tiempo de procesamiento. Adicional a esto el mapa logarítmico crece lentamente, permitiendo obtener más detalle del móvil a seguir. El criterio más relevante para la elección del mapa de

fovealización fue la aplicación de seguimiento. Esta requiere un rápido reposicionamiento del centro de la fovea y esto se logra eficientemente al desarrollar el algoritmo.

Los tres métodos de seguimiento comparados permitieron seguir un proceso para la elección del más adecuado para enfrentar los problemas que se pueden presentar durante la aplicación del sistema en un ambiente real. A continuación se describen las ventajas más relevantes del método por segmentación de color utilizando el algoritmo de *CamShift* en comparación a los métodos por segmentación de movimiento.

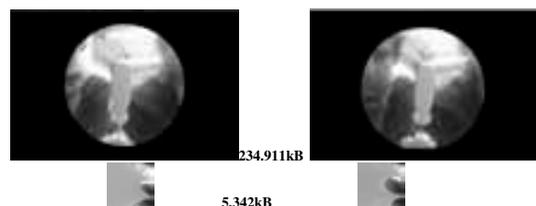
- Velocidad en el seguimiento del móvil.
- Convergencia de la fovea ante ocultaciones, cambios de luz, rotación y traslaciones del móvil.
- Si el móvil sale del campo visual de la cámara, la fovea queda posicionada en el último registro de centro de masa. En el caso del algoritmo por segmentación de movimiento utilizando una distribución gaussiana la fovea se posiciona en el centro nuevamente.
- La segmentación por color explota la ventaja biológica más relevante de la fovealización, ya que en la fovea se tiene alta concentración de conos, los cuales se especializan percepción del color.

En la figura 8 se aprecia una secuencia de imágenes que describen paso a paso el seguimiento de un objeto mediante el método de segmentación por color utilizando el algoritmo de *CamShift*. En la primera imagen, la fovea está centrada en la posición $x = 320$, $y = 240$, punto medio de la ventana de captura. En la segunda imagen se observa que por medio del mouse se selecciona el objeto que se desea seguir. Una vez seleccionado dicho objeto es posible visualizar el histograma en el cual se aprecia el color al que se realizará la segmentación, como se ve en la última imagen de la secuencia. En las imágenes restantes se aprecia la secuencia de la traslación del objeto seleccionado, tanto la imagen retinal como la imagen segmentada.



Figura 8. Secuencia de cuadros para el método de segmentación por color utilizando *CamShift*.

Una vez definido el mapa de fovealización Log Polar, y el método de segmentación por color para hacer el seguimiento, se realizaron pruebas para evaluar la funcionalidad y robustez del sistema.



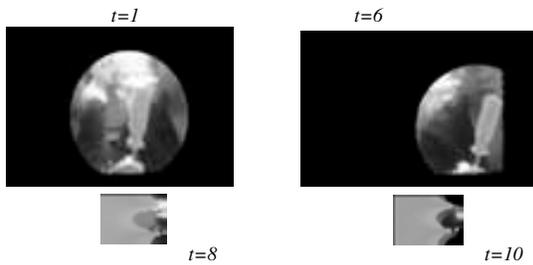


Figura 9. Secuencia de cuadros para el método de segmentación por color

En los cuadros que se muestran para diferentes tiempos se observa el desplazamiento del objeto a seguir con respecto a la imagen de referencia. Este algoritmo funciona de manera muy eficiente, ya que presenta ventajas con respecto a los algoritmos de segmentación por movimiento. Las ventajas más relevantes son la rapidez de respuesta que alcanza el sistema, su correcto funcionamiento tanto en ambientes controlados como en ambientes reales y la posibilidad de que el objeto realice traslaciones amplias, de más de 30 píxeles de magnitud en tiempos cortos, sin que el centro de la fovea deje de posicionarse sobre el objeto. Durante las pruebas se guardaron los videos en formato .AVI, y se encontró que la reducción del tamaño del video cortical es significativamente alta. La disminución del tamaño del archivo es aproximadamente 43 veces menos.

El sistema es capaz de responder ante rotación del objeto que se desea seguir, pero con traslaciones más cortas. Teniendo en cuenta esta condición es posible realizar rotaciones completas del objeto.

Al incrementar la velocidad del objeto llega un momento en la cual la fovea pierde el seguimiento debido a la velocidad alcanzada por el móvil. Como conclusión de esta prueba se encontró que el objeto debe desplazarse a una velocidad no mayor de 6.66 cm/s a una distancia de la cámara de 120 cm.

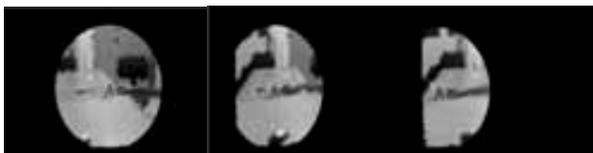


Figura 10. Secuencia de cuadros para el método de segmentación por color.

En caso de que el objeto se salga del campo visual de la cámara, según la implementación de este sistema, la fovea quedara centrada sobre la última posición donde encontró el objeto, si el objeto regresa al campo visual ya sea por el mismo lado por donde desapareció de la escena o por otro lado diferente, el sistema es capaz de continuar el proceso de seguimiento. Para que el objeto sea detectado nuevamente tiene que volver a aparecer en el cerca al área donde la fovea quedo posicionada por última vez.

En la figura 10 se muestra secuencias de un papel con la letra A en color azul de distintos tamaños. Se posicionaron los objetos a una misma distancia de la cámara y se movieron a lo

largo de la misma trayectoria. La fovea se posiciona con mayor facilidad en objetos de mayor tamaño. Para que el objeto sea detectado es necesario que ocupe al menos un 5 % del total de la fovea.

VI. CONCLUSIONES Y TRABAJOS FUTUROS

El sistema para seguimiento de objetos desarrollado provee una solución sencilla y eficiente, lo cual permite su aplicabilidad en diferentes campos como robótica, videotelefonía, entre otras. La principal ventaja de este sistema, es la reducción de datos a procesar dadas las transformaciones realizadas el video original.

Se desarrollo un sistema eficiente ante cambios de parámetros como: iluminación, forma del objeto, tamaño y rotaciones. Sin embargo se deja abierta la posibilidad de realizar mejoras e implementaciones en Hardware.

La fovealización por medio del mapa Log Polar produce una reducción en el video cortical de hasta 43 veces el tamaño del video original, con lo cual es posible reducir la cantidad de información en un proceso de transmisión.

Al investigar el funcionamiento biológico de la fovea de los seres vivos, se orientó el trabajo hacia el uso de la segmentación por color, debido a la gran concentración de receptores de color en la fovea. Concluyendo así que emular sistemas biológicos constituye una buena idea para que desde la ingeniería se realicen investigaciones del funcionamiento de algunos sistemas en la naturaleza.

Es posible emular distintos mapas de fovealización de acuerdo al tipo de ambiente en el cual se utilizará el sistema. Lo anterior puede ser desarrollado teniendo en cuenta un sistema biológico real. Así como también características y ventajas del ambiente en el cual se desenvuelve. Para la generación de este tipo de imágenes existen trabajos con los algoritmos recopilados por Leo Grady y Eric Schwartz en la Universidad de Boston, los cuales permiten generar mapas de fovealización a partir de configuraciones retinales de varios animales, como por ejemplo de palomas, de lobos y de primates.

De acuerdo a las pruebas realizadas se encontró que a pesar de las limitantes de procesamiento (equipo de ejecución), el sistema funciona correctamente. Sin embargo se podría fusionar la etapa de adquisición con la etapa de fovealización en un dispositivo de captura foveal. Logrando así un mejor desempeño del sistema.

Algunos desarrollos que podrían estudiarse son los siguientes:

- Desarrollar un sistema Transmisor-Receptor de videos fovealizados. El transmisor fovealiza y transmite los videos corticales. El receptor realiza la fovealización inversa. Con esto se reduce el tiempo de transmisión y el ancho de banda utilizado para transmitir.

- Implementar un dispositivo de visión móvil controlado, el cual permita posicionar la fovea para tener un mayor rango de visión.
- Diseñar un sensor de fovealización basado en VLSI con el fin de economizar etapas en el proceso y permitir el seguimiento de objetos con mayor velocidad.
- En el campo de la videotelefonía existe una creciente demanda por el desarrollo de que permitan que los interlocutores no deban permanecer en cierto lugar o cierta posición para realizar la comunicación, pudiendo moverse libremente.

BIBLIOGRAFÍA

- [1] BOLUDA José A., Pardo Fernando Pardo, Francisco Blasco, Joan Pelechano. Una arquitectura segmentada para el cálculo del tiempo al impacto con visión log-polar. Instituto de Robótica. Universidad de Valencia.
- [2] QI Wang, KE Zhang, YOUYI Jiang, XIANZE Xiong, The Discrete Algorithm of Log-polar Transformation.
- [3] VILLEGAS Javier. Evaluación del uso de arquitectura foveal en visión estéreo. Tesis de maestría en ingeniería electrónica y de computadores. Bogotá, Colombia, 2002 Universidad de los Andes.
- [4] TRAVER J, PRAA F. "Similarity motion estimation and active tracking through spatial domain projections on log polar images", en Computer vision and image understanding 2005 209-241
- [5] J. A. RODRÍGUEZ, C. URDIALES, P. CAMACHO, F. SANDOVAL. Detección Jerárquica de Móviles sobre Geometrías de Fovea Adaptativa. Universidad de Málaga España.
- [6] STAUFFER Chris, GRIMSON Wilson , Adaptive background mixture models for real-time tracking The Artificial Intelligence Laboratory Massachusetts Institute of Technology Cambridge, M A 02139.
- [7] TRAKULPONG Kaew, BOWDEN Richard. Vision and Virtual Reality group, Department of Systems Engineering, Brunel University An Improved Adaptive Background Mixture Model for Realtime Tracking with Shadow Detection.
- [8] RAMIREZ Alonso. CIMAT. Reporte e Investigación del método Mean Shift, descripción del mismo y aplicaciones a Filtrado preservando bordes, Segmentación y Tracking de objetos no rígidos en tiempo Real .[on line].1. [Citada Abril 3 de 2008] Disponible en Internet: <http://www.cimat.mx/~alram/VC/MSAA.htm>
- [9] SUÁREZ Hernán A., CUCHO Gonzalo. Diseño de una arquitectura para la segmentación de imágenes basados en MeanShift en un FPGA. Grupo de procesamiento digital de señales e imágenes – GPDSI sección electricidad y electrónica – Pontificia Universidad Católica del Perú

Sistema de localización y seguimiento de un conjunto de robots móviles utilizando visión global

Breyner Posso y Humberto Loaiza, Miembros IEEE

Resumen—En este artículo se presenta el diseño e implementación de un sistema de localización y seguimiento para un conjunto de robots móviles utilizando una cámara para visión global en una aplicación de fútbol robótico. Se requiere que este sistema identifique de forma confiable seis categorías de colores: amarillo, azul, naranja, cian, verde claro y magenta y pueda hacer el seguimiento de los mismos empleando poco tiempo de procesamiento.

Palabras clave—fútbol de robots, procesamiento de video, segmentación en colores, análisis de blobs.

I. INTRODUCCIÓN

El fútbol robótico es una excelente plataforma para la experimentación en el área de los sistemas multiagente [1], [2], e involucra: comunicación inalámbrica, control, inteligencia artificial y procesamiento de imágenes. En este tipo de aplicaciones se requiere controlar cada agente para que se mueva hacia una posición específica evadiendo obstáculos y al mismo tiempo lograr la colaboración de todos los miembros del equipo para alcanzar un objetivo común a través de estrategias de grupo.

Existen diferentes organizaciones que promueven el desarrollo de las competencias robóticas entre ellas se encuentran *Robocup Initiative (Robocup)* [3] y *FIRA (Federation of International Robot-soccer Association)* [4].

El sistema propuesto está inspirado en la *Small Soccer League* de la *Robocup* [5] (también denominada categoría F-180) y su diagrama general se observa en la figura 1. En términos generales, en este tipo de sistemas se tiene un conjunto de robots móviles y una pelota que se mueven sobre un entorno físicamente acotado. Mediante una cámara para visión global se realimenta el estado del sistema y la información es llevada a un computador que se encarga de dos tareas: el procesamiento de imágenes y el control de cada agente del equipo. Los comandos son enviados mediante un enlace inalámbrico a cada uno de los robots para que ellos ejecuten los movimientos de acuerdo a los sistemas de control local que posean.

En ese tipo de sistemas uno de los módulos más importantes es el de procesamiento de imágenes ya que mediante él puede obtenerse el estado cinemático de cada uno de los objetos involucrados en el juego [6]: posición y orientación de los

robots y posición de la pelota. La posición y orientación de los robots se determina mediante unas marcas de color de 5 cm de diámetro ubicadas en su parte superior: amarillo o azul como colores de equipo y cian, verde claro o magenta, para identificar de manera única cada robot dentro del equipo [5] y la posición de la pelota se obtiene identificando el objeto de color naranja en la escena.

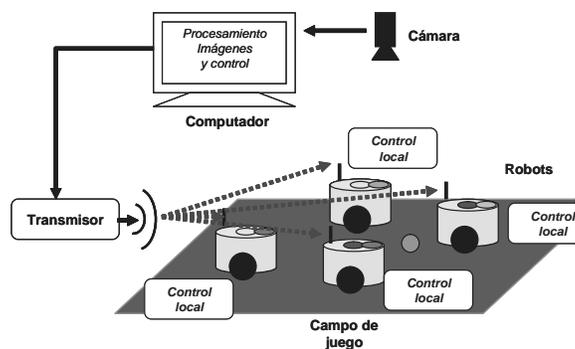


Fig. 1. Diagrama general del sistema para fútbol robótico.

Se busca que el sistema de visión sea rápido, preciso y robusto para que pueda lidiar con los cambios de iluminación, el ruido inherente en la captura y digitalización de las imágenes y el pequeño tamaño de las marcas de color [7], [8], [9], [10].

En el presente artículo se presenta un sistema de visión que implementa algoritmos sencillos que permiten una extracción rápida y efectiva de la información de las imágenes. El desarrollo de este módulo está enmarcado en el proyecto de investigación de maestría titulado “Selección e implementación de una arquitectura de control para un conjunto de robots móviles cooperativos basada en sistemas multiagente” que actualmente es llevado a cabo por el autor.

Este artículo está distribuido de la siguiente forma: en la sección 2 y 3 se describen la arquitectura hardware y software del sistema, respectivamente; en la sección 4 se consignan las pruebas realizadas y los resultados y en la sección 5 se presentan las conclusiones derivadas del trabajo realizado.

II. ARQUITECTURA HARDWARE

A nivel hardware el sistema está compuesto por los siguientes elementos: una cámara, un frame grabber, un campo de juego y un conjunto de robots móviles.

A. Cámara

Se utiliza una cámara SONY EVI-D70 que entrega una señal de video en formato NTSC con una resolución de 640x480 píxeles y 8 bits de profundidad. La cámara se encuentra ubicada a 2.48 m del campo de juego.

B. Frame grabber

Se utiliza un frame grabber Matrox Morphis [12] que permite la conexión de la señal de video en formato NTSC y va alojada en el puerto PCI del computador.

C. Campo de juego

El campo de juego mide 1.60 m x 1.74 m, se encuentra encerrado por unas paredes blancas de 15 cm de altura y cubierto por un paño de color verde que minimiza los reflejos producidos por la luz.

D. Robots

Se construyen 2 robots móviles utilizando el kit de Lego MindStorms NXT [13], con las siguientes características: locomoción diferencial, una rueda libre para apoyo y un tamaño de 18 cm x 18 cm y 15 cm de altura, un sistema para impulsar el balón (pateador) y con un círculo negro que contiene las marcas de colores en la parte superior, tal como se observa en la figura 2. Otra característica importante de los robots es su capacidad de comunicación bluetooth con el computador.

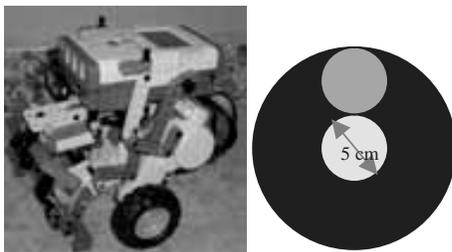


Fig. 2. Agente robótico y su marca de colores.

III. ARQUITECTURA SOFTWARE

El sistema para el procesamiento de imágenes se ha desarrollado en el entorno de programación Visual C++ 6.0 utilizando el paradigma de programación orientada a objetos y usando la *Matrox Imaging Library* (MIL) [14]. Esta librería de alto nivel tiene funciones que permiten, entre otras: capturar, grabar y mostrar imágenes en pantalla, umbralizar, realizar análisis de blobs, operaciones aritméticas, calibración, etc.

Se desarrolló la interfaz de usuario mostrada en la figura 3, a través de la que se permite: previsualizar el video, capturar el área a ser procesada, ver los planos R, G, B, H, S, y L del área seleccionada, ajustar manualmente los umbrales de segmentación para los 6 colores: amarillo, azul, naranja, cyan, verde claro y magenta e introducir los parámetros para la calibración; ver los resultados de la umbralización, la erosión,

la dilatación y el análisis de blobs e iniciar el procesamiento de video



Fig. 3. Interfaz gráfica aplicación de procesamiento de imágenes.

En la figura 4 se observa un diagrama general de bloques del módulo de visión.

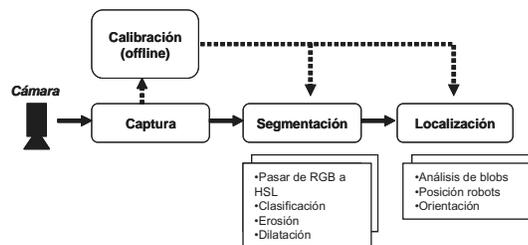


Fig. 4. Diagrama de bloques general del módulo de visión.

A continuación se explican las características más importantes de cada etapa.

A. Calibración

El proceso de calibración se centra en dos aspectos fundamentales: ajustar los umbrales de segmentación y eliminar el efecto de distorsión en la imagen capturada debido a la lente de la cámara. El primer proceso se realiza de forma manual a través de la interfaz, sin embargo ya se tienen unos valores predefinidos que se obtuvieron a través del análisis del histograma de los planos H y S de la imagen en Matlab. Estos valores son usados por el módulo de segmentación.

Para el segundo proceso se utiliza un toolbox desarrollado en Matlab denominado *Calib* [15], el cual arroja las constantes de calibración de acuerdo al método de Tsai [16] que son usados por el módulo de localización para determinar la posición real de los objetos. Se utilizó un patrón tipo tablero de ajedrez con rectángulos de 10 cm de lado y 20 imágenes, algunas de ella se observan en la figura 5.

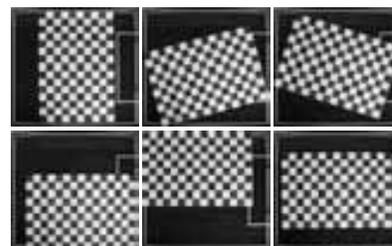


Fig. 5. Algunas imágenes usadas como entrada al toolbox Calib.

B. Captura

Este módulo se encarga de capturar una imagen de 640x480 y 8 bits de profundidad en el espacio de color RGB y mostrarla en la interfaz gráfica.

C. Segmentación

En este módulo se llevan a cabo las siguientes procesos:

1. Se pasa la imagen del espacio RGB al espacio HSL.
2. Cada píxel se clasifica de acuerdo a su valor de tono (plano H) y saturación (plano S) en uno de dos buffers binarios. El primer buffer para los colores Amarillo, Azul o Naranja y el segundo para los colores Cyan, Verde claro o Magenta. Esta clasificación en dos buffers se hace tratando de separar los blobs que puedan estar más cercanos y evitar que el módulo de análisis de blobs considere que dos colores muy cercanos pertenecen a la misma región [11].
3. Se aplica la erosión a cada buffer binario para eliminar píxeles ruidosos.
4. Se aplica la dilatación para dar mayor tamaño a las regiones antes de pasarlas al módulo de análisis de blobs.

En la figura 6 se observan los elementos estructurales utilizados en la erosión y dilatación.

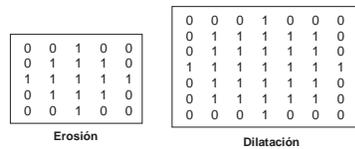


Fig. 6. Elementos estructurales.

D. Localización

En este módulo se llevan a cabo los siguientes procesos:

1. Se utiliza el módulo de análisis de blobs para extraer el número de blobs, el centroide de cada uno y su área.
2. Se eliminan los blobs que no cumplen un tamaño mínimo.
3. Se determina el color de cada blob evaluando el color del píxel de su centroide.
4. Se clasifican los blobs en 6 vectores, uno para cada color.
5. Se determina cuales blobs pertenecen a un mismo robot, teniendo en cuenta que el azul y amarillo son colores de identificación de equipo y cyan, magenta, y verde claro se utilizan para identificación de cada robot. Esto se hace mediante distancias euclidianas.
6. Se calcula la orientación del robot usando el centroide de las dos marcas.

IV. PRUEBAS Y RESULTADOS

Se llevaron a cabo tres pruebas en el sistema: verificación de segmentación de todos los colores, seguimiento de robots en el campo de juego, seguimiento de robots y balón en un enfrentamiento delantero vs arquero.

A. Prueba 1: validación de la segmentación

Para esta prueba se distribuyen parches de colores en todo el área de juego con el objetivo de identificar el centroide de cada uno de ellos mediante una cruz. En la figura 7 se

observan los resultados paso a paso del proceso de identificación.

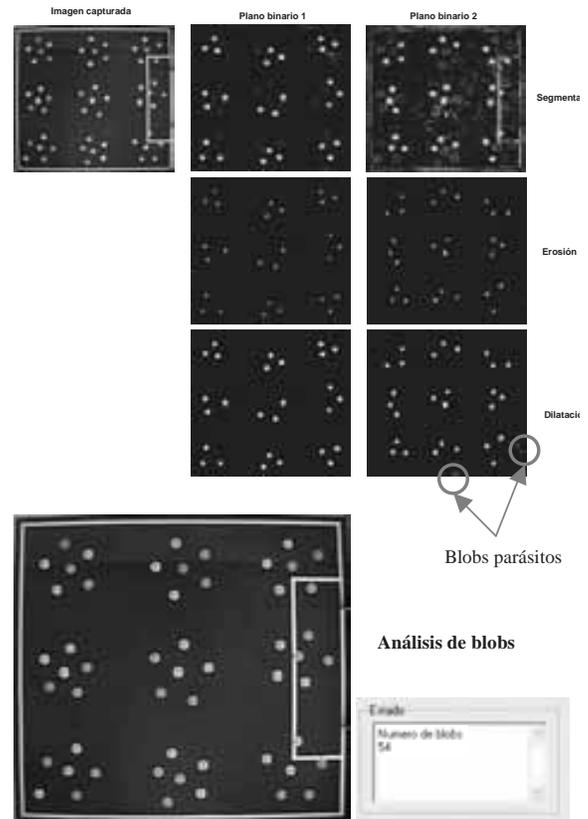


Fig. 7. Pasos en el proceso de identificación de parches de colores.

En el primer paso la imagen capturada es segmentada en dos planos binarios: el primero contiene los colores amarillo, azul y naranja y el segundo los colores cyan, verde claro y magenta. Se observa que después de la segmentación el plano binario 2 presenta muchos más píxeles ruidosos, esto se debe a que el fondo verde a veces es clasificado como cyan y las líneas que demarcan el terreno a veces se confunden con el verde claro. Este efecto se minimiza variando los umbrales de segmentación, sin embargo, para el caso del cyan, si se estrecha mucho este rango ya no puede clasificarse adecuadamente este color.

En el siguiente paso del proceso (erosión) se elimina la gran mayoría del ruido, sin embargo al aplicar la dilatación aparecen unas regiones parásitas. Estas regiones se eliminan en el análisis de blobs ya que no cumplen el tamaño mínimo para ser consideradas un parche de colores. Este tamaño mínimo ha sido determinado de forma experimental. El máximo error en la localización del centroide es de 1.5 cm y se obtiene en las regiones más oscuras del campo, tal como puede observarse en la imagen final del proceso. Este error puede minimizarse si se ubica otra lámpara que permita una iluminación un poco más uniforme sobre el campo. El número de blobs identificados también es mostrado en la interfaz gráfica.

Después de tener el sistema calibrado puede realizarse el procesamiento de video.

B. Prueba 2: localización y seguimiento de dos robots

Para esta prueba se ubican dos robots en el campo de juego cada uno con sus marcas de colores y se ponen a deambular con el objetivo de realizar el seguimiento durante todo el recorrido. Al final, se marca la trayectoria con líneas de colores. En la figura 8 se observa el resultado de esta prueba.

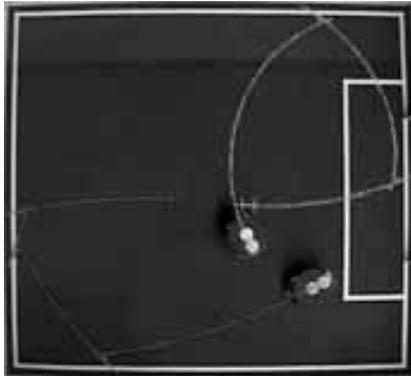


Fig. 8. Resultado de la prueba 2.

Se logran procesar en promedio 20 frames/seg, utilizando la técnica de doble buffer la cual consiste en procesar una imagen capturada mientras se captura la siguiente, esto garantiza que la pérdida de frames sea mínima y el procesamiento sea más eficiente, debido a que se eliminan los tiempos de espera para procesar los nuevos frames. Esta característica se logra mediante algunas funciones de la MIL. En la imagen 8 se observa que el sistema hace el seguimiento correcto de cada uno de los robots. El robot del equipo amarillo se ha puesto a deambular en una trayectoria curva, mientras el robot del equipo azul describe trayectorias más rectilíneas.

C. Prueba 3: movimiento de ataque portero vs delantero

Para esta prueba se ubica un robot en la portería (arquero) haciendo un movimiento rectilíneo tratando de bloquearla, mientras que otro robot (delantero) dribla el balón cierta distancia y dispara. El objetivo de la prueba es identificar los robots y el balón en una escena típica de fútbol robótico. En la figura 9 se observa el resultado de esta prueba, donde puede notarse que el sistema hace el seguimiento de todos los objetos dentro de la escena. Cabe resaltar que en este punto los robots no están siendo controlados por el computador, sino que obedecen a la programación local que se les ha dado. Uno de los problemas principales en las escenas de fútbol robótico es que a veces el balón está siendo driblado y podría no identificarse bien. Se han planteado diferentes soluciones [9], [11] desde las que utilizan una cámara auxiliar, hasta las que dotan al sistema de control de suficiente inteligencia para que puedan determinar donde se encuentra el balón de acuerdo al estado del juego.

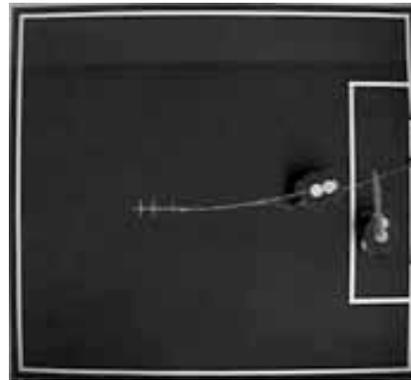


Fig. 9. Resultado de la prueba 3.

V. CONCLUSIONES

Se implementó un sistema que permite procesar el video de una escena de fútbol robótico a una velocidad de 20 frames/seg identificando seis categorías de colores: amarillo, azul, naranja, cian, verde claro y magenta, y que puede asociar a cada robot dos marcas de estos colores utilizando distancias euclidianas.

Se utilizó un algoritmo de segmentación muy sencillo que simplemente compara el valor de cada uno de los píxeles en los planos H y S, con unos umbrales que son generados manualmente. A pesar del ruido inherente de este tipo de segmentación, las operaciones morfológicas de erosión y dilatación ayudan a contrarrestar este problema. Al final, el análisis de blobs toma en cuenta los tamaños de cada una de las regiones para determinar cuáles corresponden a marcas de colores y cuales no, lo que da mayor robustez al sistema.

El proceso de localización arroja un error máximo de 1.5 cm, el cual puede ser disminuido mejorando las condiciones de iluminación sobre el campo.

REFERENCIAS

- [1] M. Veloso, P. Stone, y K. Han, "The cmunited-97 robotic soccer team: perception and multiagente control", in AGENTS'98: Proceedings of the first international conference on autonomous agents. New York, USA: ACM Press, 1998, pp. 78 – 85.
- [2] J. Kim, "Cooperative multi-agent robotic systems: from the robotic soccer perspective", 1997. [En línea]. Disponible en: citeseer.ist.psu.edu/kim97cooperative.html
- [3] "Robocup Initiative. <http://www.robocup.org/>".
- [4] "Fira – federation international of robot-soccer associated. <http://www.fira.net/>".
- [5] "Small Size Robotic League. <http://small-size.informatik.uni-bremen.de/>".
- [6] P. Costa, P. Marques, A. Moreira, A. Sousa y P. Costa, "Tracking and identifying in real time the robots of a F-180 team", in Proceedings IJCAI'99 - International Joint Conference on Artificial Intelligence, The Third International Workshop on RoboCup. Stockholm, 1999.
- [7] J. Bruce, T. Balch, M. Veloso, "Fast and inexpensive color image segmentation for interactive robots", in Proceedings of the IROS 2000. Japón, 2000.
- [8] B. Li, E. Smith, H. Hu, L. Spacek, "A real time visual tracking system in the robot soccer domain", in Proceedings of EUREL Robotics-2000. Salford, England, 2000.
- [9] R. Garcia, J. de la Rosa, A. Oller, J. Battle, A. Figueras, J. Ramon, "A fast vision system for tracking in robot soccer", in Proceedings IEEE of Mirost-98. Paris, 1998.

- [10] G. Gallinelli, G. Grittani y J. Ramirez, "FutBotIII: towards a robust centralized vision system for robocup small league", 1999. [En línea]. Disponible en: <http://www.ep.liu.se/ea/cis/1999/006/>
- [11] J. Loomis, J. Palmer y P. Pandit, "Performance development of a real-time vision system", 2003. [En línea]. Disponible en: www.cis.cornell.edu/boom/2005/ProjectArchive/robocup/documentation/robocup/2003/2003Vision.pdf
- [12] "Matrox Morphis. <http://www.matrox.com/imaging/products/morphis/home.cfm>"
- [13] "Lego MindStorms NXT. <http://mindstorms.lego.com/eng/Overview/default.aspx>"
- [14] "Matrox Imaging Library. <http://www.matrox.com/imaging/products/mil/home.cfm>"
- [15] "Camera Calibration Toolbox for Matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/"
- [16] R. Tsai, "Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using off-the-shelf TV Camera Lenses" en: IEEE Journal of Robotics and Automation. Vol. RA-3, No 4. (1987). pp. 323-344.

AUTORES



Breynner Posso Bautista. Ingeniero electrónico de la Universidad del Valle (2004). Actualmente estudiante de maestría en ingeniería con énfasis en automática en la Universidad del Valle y profesor auxiliar en la misma universidad. Áreas de interés: robótica móvil, sistemas multiagente, inteligencia artificial, procesamiento de imágenes.
breyposo@univalle.edu.co



Humberto Loaiza Correa. Ingeniero electricista (1991) y M. Sc. en Automática (1994) de la Universidad del Valle. Ph.D. en Robótica de la Université d'Evry, Francia (1999). Profesor titular de la Universidad del Valle. Áreas de investigación en robótica móvil, visión artificial, procesamiento de imágenes y señales.
hloaiza@univalle.edu.co

Utilización de matrices lógicas programables por campo FPGA, para la implementación de un sistema de evaluación de redes neuronales orientadas al reconocimiento de patrones de imagen

Ing. Ilber Adonayt Ruge R., iruge@unicundi.edu.co, ilberruge@yahoo.es Jose David Alvarado, elejosedavid@gmail.com
Grupo de Investigación en Tecnologías de la Información y las Comunicaciones GITEINCO Universidad de Cundinamarca - Fusagasuga

Resumen- Se presenta la utilización de Matrices Lógicas Programables Por Campo (FPGA) para la implementación de un sistema de evaluación de redes neuronales artificiales (RNA), y la aplicación de reconocimiento de patrones de imágenes de caracteres alfabéticos, en el que se utilizan diferentes modelos de redes neuronales. Esto se desarrollo para evaluar el desempeño de las redes neuronales implementadas en la FPGA y para explorar y analizar las características de funcionamiento de la FPGA Spartan-3 XCS200. El proyectó se realizo en cuatro fases las cuales son: establecer los modelos de redes neuronales, desarrollo de una plataforma de entrenamiento de RNA utilizando Labview, diseño, implementación y verificación de los modelos de RNA, y por último se estableció los parámetros y se realizo la evaluación del desempeño de las RNA implementadas en la FPGA Spartan-3 XCS200.

Palabras clave – ART, Hopfield, Labview, VHDL, FPGA, reconocimiento de patrones.

I. INTRODUCCION

Una FPGA es un dispositivo lógico programable que se caracteriza por su funcionalidad (flexibilidad y versatilidad), y en la actualidad se encuentra presente en el desarrollo de aplicaciones en diferentes áreas de la ingeniería, por tal motivo se busca abordar este tipo de dispositivos para el desarrollo de una aplicación que nos permita explorar y evaluar esta herramienta tecnológica.

La aplicación seleccionada para el desarrollo del proyecto es el reconocimiento de patrones aplicado a caracteres alfabéticos utilizando redes neuronales artificiales (RNA), y se busca utilizar diferentes modelos de RNA que dentro de sus aplicaciones involucren esta aplicación, para posteriormente evaluar y analizar el rendimiento de los diferentes modelos implementados en la FPGA XCS200.

II. SISTEMA DE EVALUACION DE RNA.

El sistema de evaluación busca comparar el funcionamiento de diferentes modelos RNA implementadas en una FPGA Spartan-3 XCS200^[5], con el ánimo de establecer qué modelo de red tiene un mejor rendimiento con relación a la aplicación de reconocimiento de patrones.



Fig. 1. Sistema de evaluación de RNA

III. SELECCIÓN DE LOS MODELOS DE RNA

En el proceso de la selección de la RNA utilizadas para el desarrollo de este proyecto se analizaron diferentes aspectos de la topología de las redes, con respecto al número de capas que sea lo menor posible para realizar una buena clasificación, y que la información que procese la red sea digital (redes discretas) esto se debe al tipo de información que manejan las FPGAs.

Después de analizar la información recopilada y de realizar un estudio de las diferentes topologías de RNA teniendo en cuenta lo anterior, los modelos que mejor cumplen con estas características son el modelo de resonancia adaptativa (ART) y el modelo Hopfield.

IV. PLATAFORMA DE ENTRENAMIENTO RNA

La plataforma es un software diseñado para realizar el entrenamiento de RNA para la aplicación de reconocimiento de patrones enfocado a caracteres alfabéticos, este es desarrollado en Labview que es una herramienta de instrumentación virtual que ofrece ventajas para el desarrollo

de este tipo de aplicaciones, la plataforma cuenta con una interfaz de usuario que ofrece facilidades para realizar el entrenamiento y posteriormente efectuar la verificación del funcionamiento de RNA de una manera práctica, donde se encuentran los controles y las herramientas necesarias para este fin, así como la opción de exportar la información obtenida del entrenamiento de RNA.



Fig. 2. Plataforma de entrenamiento de RNA

La plataforma de entrenamiento cuenta con un modulo que permite realizar el entrenamiento y la verificación del funcionamiento del modelo de resonancia adaptativa, de igual manera cuenta con un modulo para el modelo Hopfield que cumple la misma función, el modulo de comunicación con la FPGA permite presentar los patrones para ser transmitidos del PC a la FPGA, y por último la opción de salir para terminar la ejecución de la aplicación.

A. Algoritmo de entrenamiento del modelo de resonancia adaptativa (ART)

El proceso de entrenamiento de las RNA^[2] consiste en el proceso de aprendizaje por el cual se modifican los pesos de la red en respuesta a una información de entrada, este tipo de mecanismo de aprendizaje se diferencia en cada modelo.

El modelo de resonancia Adaptativa ART es de tipo ON LINE^[1], este mecanismo consiste en que el número de neuronas y el de los pesos pueden variar durante el entrenamiento y el funcionamiento de la red.

$$V_{ij}(t=0) = 1 \quad W_{ji}(t=0) = \frac{1}{1+N} \quad (1)$$

Para iniciar el proceso de adaptación de los pesos la dimensión de red dependerá del número de neuronas de entrada y de patrones que se entrenaran inicialmente, cuando se presenta el patrón que se desea entrenar la información se propaga a través de la conexiones hacia adelante y se almacena en los pesos V_{ij} y los pesos W_{ji} se actualizan a partir de la siguiente ecuación:

$$W_{ji}(t+1) = \frac{V_{ij} * (t) e_i^{(k)}}{\gamma + \sum_{i=1}^N V_{ij} * (t) e_i^{(k)}} \quad (2)$$

Cuando la red entra en funcionamiento la relación de semejanza es el parámetro encargado de establecer si es necesario de que un nuevo patrón sea aprendido por la red. El valor de relación de semejanza que se utilizara es de 90% ya que los caracteres alfabéticos tienen bastante similitud, por tal razón hay que ser muy selectivos para poder realizar una buena clasificación.

$$R = \frac{\|E_k * X\|}{\|E_k\|} = \frac{\|E_k * V_j\|}{\|E_k\|} \quad (3)$$

La adaptación de los pesos de la red cuando se encuentra en funcionamiento se realiza utilizando la ecuación 2 para las conexiones feedforward W_{ji} , y para las conexiones feedback V_{ij} , básicamente lo que se realiza es que el patrón de entrada se almacena en la matriz. La arquitectura de la red se modificara agregando nuevas neuronas en la capa de salida.

B. Algoritmo de entrenamiento del modelo Hopfield

La red neuronal tipo Hopfield tiene un mecanismo de aprendizaje OFF LINE^[1], es decir, que inicialmente se realiza el proceso de aprendizaje, y cuando este procedimiento finaliza comienza el proceso de funcionamiento de la red, a diferencia del aprendizaje ON LINE, donde el proceso de aprendizaje y funcionamiento son paralelos o simultáneos.

El funcionamiento del modelo Hopfield se divide en dos partes la primera es entrenamiento en donde se presentan diferentes patrones que son almacenados o memorizados en la red neuronal, y la segunda parte es el funcionamiento que consiste en que si se le presenta a la entrada algunas de las informaciones almacenadas, la red se propaga hasta estabilizarse, ofreciendo entonces en la salida la información correspondiente a la información almacenada. Si, por el contrario, la información de entrada no coincide con ninguna de las almacenadas, por estar distorsionada o incompleta, la red se propaga generando en la salida lo más parecido posible a la información almacenada.

Para realizar a la adaptación de pesos a la matriz W_{ji} se realizara a partir de los patrones de entrada representados por los vectores E, la adaptación se realizara ha utilizando la siguiente ecuación:

$$W = \sum_{k=1}^M [E_k^T E_k - I] \quad (4)$$

Donde la matriz E_k^T es la transpuesta de la matriz E_k , e I es la matriz identidad de dimensiones NxN que anula los pesos de las conexiones autorrecurrentes.

V. IMPLEMENTACION DE LAS RNA EN FPGA

Para el desarrollo de este proyecto se cuenta con la FPGA Spartan-3 XCS200 de xilinx, por lo cual, el software para la realización de aplicaciones basadas en esta FPGA^[4] deben ser

las herramientas desarrolladas por el fabricante, el cual provee diferentes herramientas de síntesis. El software utilizado es el Xilinx ISE 9.1, que incorpora herramientas para la síntesis, implementación, diseño y programación del dispositivo.

El software de desarrollo permite realizar el diseño utilizando lenguaje de descripción de hardware^[3], en los que se puede encontrar el lenguaje *ABEL* (Advanced Boolean Expression Language), *VHDL* (VHSIC Hardware Description Language), *Verilog*. El lenguaje de descripción seleccionado para el desarrollo del proyecto es el VHDL, ya que este lenguaje es uno de los más populares y la información para desarrollar aplicaciones basadas en este es amplia.

A. Diseño del modelo de resonancia adaptativa (ART)

EL diseño del modelo de ART utilizando VHDL se realizó incorporando el algoritmo de funcionamiento de modelo, en donde es necesario adicionar algunos componentes para el diseño del modelo como: reloj del sistema, demultiplexor y el sistema para la visualización.

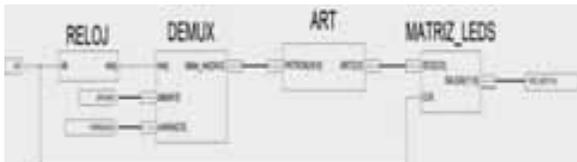


Fig. 3. Diagrama completo del modelo ART

El reloj del sistema es el encargado de la sincronización, el demultiplexor es utilizado para ajustar la información que proviene del PC (puerto paralelo) a 8 bits a los 35 bits que se requieren para representar la imagen, el bloque ART tiene incorporado el algoritmo de funcionamiento del modelo de la red neuronal, y por último tenemos el proceso de la visualización de la imagen que ajusta la señal de salida del modelo ART para representar el patrón en una matriz de 7x5.

B. Diseño del modelo Hopfield utilizando VHDL

Para realizar la implementación del modelo Hopfield se desarrolla el algoritmo de funcionamiento en VHDL, en donde los bloques que hacen referencia al reloj del sistema y el demultiplexor se utilizan los descritos anteriormente en el modelo ART (ver figura 3).

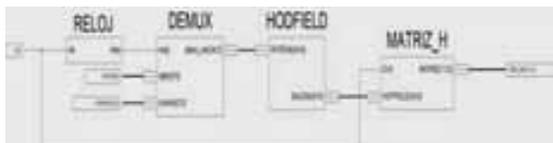


Fig. 4. Diagrama completo del modelo Hopfield

En el bloque Hopfield se desarrolló el algoritmo de funcionamiento del modelo de red neuronal utilizando VHDL, en el cual el módulo de visualización realiza el ajuste de la señal de salida del módulo Hopfield para representarla en una matriz de 7x5.

C. Adquisición y procesamiento de la imagen.

El proceso de adquisición de la imagen se realiza utilizando una herramienta la cual permite plasmar una imagen de tamaño 350x250 píxeles, posteriormente se ejecuta un procesamiento de la imagen realizando una caracterización de una imagen obteniendo así un patrón representativo de la original.

VI. IMPLEMENTACION DE LAS RNA EN FPGA

Una vez concluido la etapa de diseño del código en VHDL que describe cada uno de los módulos necesarios para la implementación de los modelos de redes seleccionados, el software Xilinx ISE utilizado para el desarrollo de aplicaciones basadas en FPGA permite realizar la metodología de diseño^[3] necesaria para la implementación, los pasos necesarios para realizar la programación (reconfiguración de hardware) del dispositivo.

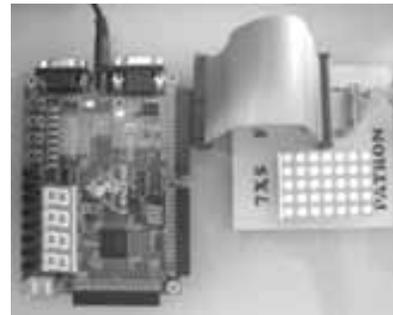


Fig. 5. Sistema de evaluación de RNA

A. Implementación del modelo ART

El modelo ART es entrenado para clasificar diez patrones diferentes compuestos por píxeles que representan los primeros caracteres del alfabeto, obteniendo una arquitectura del modelo red neuronal de 35 neuronas en la capa de entrada y de 10 neuronas en la capa de salida alcanzando 350 conexiones e igual número de pesos entre las neuronas.

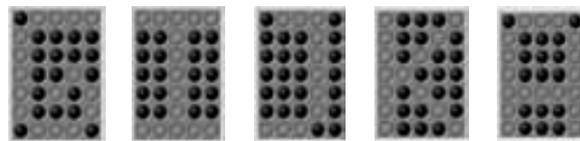


Fig. 6. Características de los Patrones entrenados en el modelo ART y Hopfield

B. Implementación del modelo Hopfield

El modelo Hopfield es entrenado para clasificar cinco patrones diferentes, el número de patrones utilizado se debe a que el autor establece una limitación de la capacidad del número de patrones que se pueden entrenar para obtener una buena clasificación, debe ser menor que el 13.8% del número de neuronas de la capa de entrada, otra limitación que presenta

el modelo es que los patrones entrenados deben tener una diferencia igual o mayor a 30%, teniendo en cuenta lo anterior los patrones entrenados son: G, I, J, K, I, en la figura 10 se encuentran las características de los patrones.

La arquitectura de red obtenida al realizar el entrenamiento es de 35 neuronas en la capa de entrada y 35 neuronas en la capa de salida para un total de 1190 conexiones entre la neuronas de la red, vale la pena resaltar que le numero de pesos obtenido en el entrenamiento no modificara si se modifica el numero de patrones entrenados en la red.

VII. EVALUACIÓN DE RNA IMPLEMENTADAS

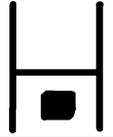
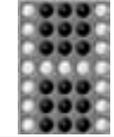
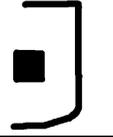
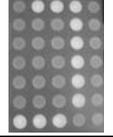
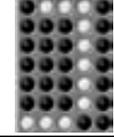
La evaluación de RNA se realizo en dos etapas: lo primero es efectuar la verificación del funcionamiento de los modelos de redes implementadas en la FPGA para posteriormente evaluar el desempeño de las redes implementadas con respecto a la aplicación de reconocimiento de patrones.

A. Verificación Modelo ART

Utilizando el modulo de comunicación para presentarle diferentes patrones a la RNA implementada en la FPGA, los resultados que se obtienen se pueden ver en la Tabla 2:

A continuación se muestra los resultados obtenidos al presentarle a la red neuronal patrones con ruido o perturbaciones:

TABLA 1
FUNCIONAMIENTO PARA EL MODELO DE ART UTILIZANDO
PATRONES CON RUIDO

PATRON DE ENTRADA	PATRÓN DE SALIDA FPGA	PATRÓN DE SALIDA SIMULACIÓN
		
		

El funcionamiento de red neuronal implementada en la FPGA tiene un comportamiento óptimo comparado con los resultados obtenidos en la simulación. La plataforma solo tiene problemas en la clasificación de los patrones correspondientes a los caracteres E y F, y los patrones C y G debido a la similitud que presentan estos, pero los resultados obtenidos en la plataforma corresponden a los obtenidos en la simulación.

Verificación del Modelo Hopfield

El proceso de verificación del funcionamiento no se logró realizar debido a que la capacidad de la FPGA utilizada para el desarrollo del proyecto no cuenta con los recursos necesarios para la implementación de la red neuronal.

Debido a que no se pudo realizar la implementación del sistema completo, se utilizo un número menor de neuronas para verificar el funcionamiento del modelo Hopfield, los resultados obtenidos al comparar el funcionamiento de 5 neuronas implementadas en la FPGA con los resultados de la simulación es que tenían un correcto funcionamiento, debido a esto se puede decir que el código VHDL diseñado con el fin de describir el funcionamiento del modelo Hopfield es correcto.

VIII. EVALUACIÓN DEL DESEMPEÑO DE LAS RNA IMPLEMENTADAS EN LA FPGA SPARTAN-3 XCS200

Después de realizar un breve análisis ala sistema de evaluación, el parámetro más relevante que se estableció es la cantidad de recursos con lo que dispone la FPGA Spartan-3 utilizada para el desarrollo del proyecto. Esto se estableció teniendo en cuenta las características y la arquitectura de las RNA para la aplicación de reconocimiento de patrones, debido a la cantidad de recursos que necesita cada modelo para ser implementadas en la FPGA.

TABLA 2
COMPONENTES Y CAPACIDAD DE LA FPGA SPARTAN-3

COMPONENTES	DISPONIBLES
Slice Flip Flops	3840
LUTs	3840
Slices	1920
IOBs	173
GCLK	8€

Los recursos utilizados por el modelo ART y el modelo Hopfield se muestran en la siguiente tabla en donde encontramos los recursos usados y el porcentaje de utilización.

TABLA 3
RECURSOS USADOS POR EL MODELO ART Y HOPFIELD

COMPONENTES	ART		HOPFIELD	
	USADOS	%	USADOS	%
Slice Flip Flops	490	25%	5495	286%
LUTs	106	2%	106	2%
Slices	869	22%	10047	261%
IOBs	56	32%	56	32%
GCLK	2	25%	2	25%

Los recursos utilizados por el modelo ART son relativamente bajos, con respecto a la arquitectura de la red neuronal, y se puede establecer que cada neurona utiliza algo más del 2% de los recursos de la FPGA XC200, en base a esto es posible entrenar de 30 a 35 patrones adicionales ya que se dispone del 70% de los recursos de la FPGA XC200.

Para el modelo Hopfield no se puede realizar la implementación del sistema propuesto debido a que los recursos de la FPGA no son suficientes como se observa en la tabla 4, esto se debe en parte a las características de la red como la representación de la información de entrada que en este caso utilizan valores binarios de 1 y -1, y debido a esto el valor de alguno de los pesos de la red tengan valores negativos lo que implica un gasto computacional elevado, otro aspecto es la arquitectura de la red que en este caso es de 35 neuronas en la capa de entrada y 35 neuronas en la capa de salida que generan 1190 conexiones entre las neuronas de la red.

Debido a que no se pudo realizar la implementación del modelo Hopfield en la FPGA XC200 se buscaron alternativas de diseño para optimizar el código VHDL y a partir de esto reducir la cantidad de recursos requeridos. La alternativa que se utilizó para realizar la optimización del código VHDL es la herramienta de xilinx System Generator, este cual permite diseñar una aplicación para FPGA utilizando el simulink de Matlab.

TABLA 4
RECURSOS UTILIZADOS POR LA OPCIÓN A Y LA OPCIÓN B DEL
MODELO HOPFIELD

Componente	Hopfield A		Hopfield B	
Slice Flip Flops	5495	286%	4644	241%
LUTs	106	2%	109	2%
Slices	1004 7	261%	8376	218%
IOBs	56	32%	68	39%
GCLK	2	25%	2	25%

Si comparamos los resultados obtenidos entre la opción A y la opción B, el System Generator realmente permite optimizar los recursos utilizados por la FPGA XCS200 Spartan-3, debido a que se presenta una disminución del 40% de la opción B con respecto a la opción A, pero la implementación del modelo Hopfield no se puede realizar por que la reducción de los recursos utilizados por la aplicación superan la capacidad de FPGA con la que se cuenta para el desarrollo del proyecto.

Debido a que la FPGA utilizada para el desarrollo del proyecto no posee los recursos necesarios lógicos para realizar la implementación del modelo Hopfield, a través de software Xilinx ISE se realizaron simulaciones con el ánimo de establecer que FPGA tiene los recursos necesarios para la implementación, después de realizar diferentes simulaciones se obtuvo que la FPGA Spartan-3 XCS1000 que dispone de

1000k compuertas, cuenta con los recursos necesarios para realizar la implementación del modelo Hopfield.

IX. CONCLUSIONES

La utilización de una FPGA para la implementación de RNA en un hardware, ofrece ventajas para el diseño y desarrollo de este tipo de aplicación como la disminución del tiempo de desarrollo, la reducción de los costos de fabricación, el buen rendimiento del sistema, una alta escala de integración y de capacidad, y la más determinante es que permite la reconfiguración de hardware sin realizar procesos que impliquen grandes cambios de diseño comparados con otras tecnologías como DSP, VLSI, PCB y ASIC.

El modelo de red neuronal que mejor desempeño demostró es el modelo ART, esto se debe principalmente a que la FPGA Spartan-3 XCS200 contaba con los recursos suficientes para la implementación del modelo de red neuronal y al realizar la verificación entre la FPGA con respecto a la simulación el resultado que se obtuvo fue que el sistema implementado presento un correcto funcionamiento. Por otra parte el número de patrones que puede clasificar el modelo ART es ampliamente mayor al modelo Hopfield en este caso, aunque este margen se podrá reducir al aumentar las dimensiones del patrón de entrada.

BIBLIOGRAFÍA

- [1] Hiler, José. Martines, Víctor. Redes Neuronales Artificiales fundamentos modelo y aplicaciones. RAMA, 1995.
- [2] Freeman, James. Skapura, David, Redes Neuronales Algoritmos, aplicaciones y técnicas de programación, Adison Wesley, 1993.
- [3] Pardo, Fernando. Baluda, José. VHDL lenguaje para síntesis y modelado de circuitos. Alfaomega, 2004
- [4] PARNELL, Karen. METHA, Nick, Programmable logic Design Quick Start Hand Book. Fourth Edition ISE 5.1i, Xilinx, 2003.
- [5] SPARTAN-3 STARTER KIT BOARD USER GUIDE, UG130 (v1.1), XILINX, 2005.

Identificación de color, tamaño y forma mediante técnicas de visión artificial y lógica difusa en banda transportadora

Camilo Bello, Alejandro Triviño, Diego Zárate, Felipe Zarta Ing. Sandra Cancino Directora Proyecto, Escuela Colombiana de Ingeniería

Resumen- Este es un trabajo que aplica y demuestra conceptos, técnicas y algoritmos. En este artículo se presenta un estudio acerca de técnicas para el reconocimiento de características de objetos en color, tamaño y forma, a través de visión artificial y lógica difusa, para ser utilizadas en una banda transportadora. Así mismo, se presenta un control difuso de posición del motor de la banda transportadora. Finalmente se presenta un análisis de los resultados obtenidos por ambas técnicas de forma que se establezcan ventajas y desventajas de la implementación de cada técnica frente a la otra.

Palabras Clave- Identificación, Color, Tamaño, Forma.

I. INTRODUCCIÓN

En un proceso industrial es importante la visualización, identificación y reconocimiento de características propias del producto, de forma que sea posible clasificar una línea de producción, identificar posibles imperfecciones u otras métricas importantes para la gestión en la empresa. Con el fin de optimizar el proceso, reduciendo costos y tiempos adicionales de un procesamiento manual, se utilizan técnicas de visión artificial y lógica difusa para la automatización de procesos.

Tanto la identificación por Visión Artificial, como por Lógica Difusa utilizan las mismas entradas, sin embargo los resultados pueden ser diferentes.

El desarrollo del artículo es como sigue: inicialmente elementos de *hardware*, seguido de control difuso de posición. A continuación se presenta la adecuación y extracción de información de la imagen adquirida, que consiste de tres etapas: preprocesamiento, binarización y segmentación. Finalmente se presentan los resultados obtenidos para las tres características tomadas del objeto (color, tamaño y forma), así como su análisis estadístico.

II. METODOLOGÍA

ESTRUCTURA FÍSICA Y *HARDWARE*

- Banda Transportadora
- Cámara de video para PC (288x352)
- PC

- Servomotor (12 Vdc, 800mA)
- Microcontrolador 16f877A
- Amplificador de Corriente L293B

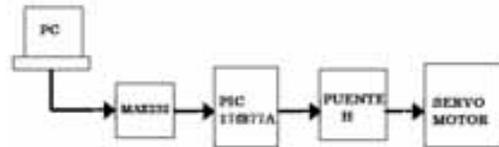


Fig. 1 Hardware para el Control Difuso de Posición del Servomotor.

A. DISEÑO DEL CONTROL DIFUSO DE POSICIONAMIENTO DE LA BANDA TRANSPORTADORA

Se implementa un control difuso de posición para el motor. Las entradas al controlador difuso de posición son: *Error* y Variación del Error ($\Delta Error$), definidas en la ecuación 1:

$$Error = referencia - fila$$

$$\Delta Error = Error_{actual} - Error_{anterior}$$

Ec. 1. Definición de Entradas al Controlador Difuso

En la figura 2 se muestra el esquema del lazo de control difuso de posición para el servomotor.

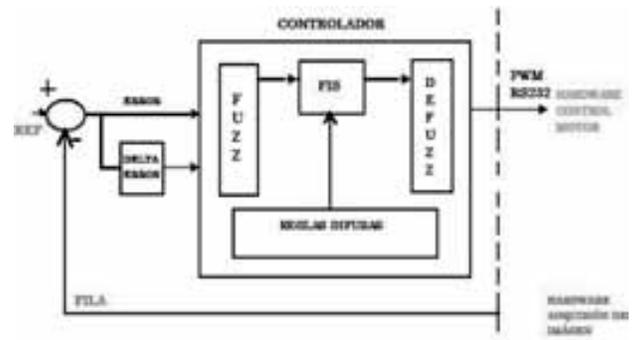


Fig. 2 Lazo de Control Difuso de Posición

Para la definición de los rangos de los conjuntos difusos de la variable ERROR se definen zonas específicas de la posición de la muesca, que a su vez define el valor de la variable Fila (punto medio de la muesca), tal como se muestra en la - 1 - figura 3.

1 Institución a la que pertenecen los autores www.escuelaing.edu.co.



Fig. 3 Zonas definidas para la variable Error (Visualización desde la interfaz gráfica)

Con respecto a la Variación del Error, se espera no tener una diferencia significativa entre la posición actual de FILA y su posición anterior., la salida es un valor numérico entre 0-255 el cual será representado por una modulación por ancho de pulso PWM.

B. PROCESAMIENTO INICIAL DE LA IMAGEN

La aparición de variaciones en intensidad debidas al ruido, por deficiencias en la iluminación, o la obtención de imágenes de bajo contraste, hace necesario un preprocesamiento de la imagen. El contraste incrementa el cambio de luminosidad entre las zonas más oscuras o más claras. Inicialmente fue necesario utilizar una fuente de luz (lámpara de luz amarilla 40 W) para proporcionar una iluminación uniforme al cuadro de la imagen y así poder modificar el brillo de la imagen adquirida por la cámara Web.

La segmentación de una imagen consiste en la división o separación de una imagen en regiones de atributos similares. La técnica utilizada es: Aplicación de umbrales de niveles de gris "*Segmentación por umbralización*". La *umbralización de propiedades* es el proceso de Segmentación más simple, sin embargo resulta muy eficaz en muchas aplicaciones. La idea se basa en suponer el rango de valores de un objeto, respecto a una determinada propiedad. Aquellos píxeles que estén en ese rango serán parte del objeto y aquellos fuera de ese rango serán parte del resto de la imagen. La dinámica de aplicación hace que el proceso de umbralización sea muy rápido computacionalmente la parte más difícil es encontrar el valor correcto del umbral.

Al aplicar un umbral, la imagen quedará binarizada; etiquetando con „1“ los píxeles correspondientes al objeto y con „0“ aquellos que son del fondo. El proceso de segmentación se muestra en la figura 5. A la imagen con muesca se le aplica una máscara para eliminar su muesca (figura 5c)

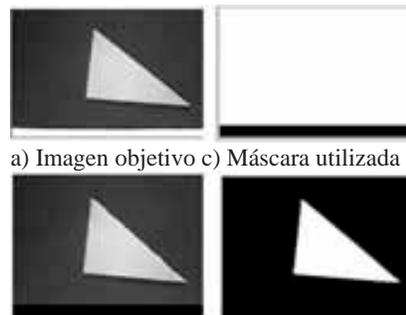


Fig. 5 Proceso seguido para la segmentación de la imagen.

Debido a la resolución de la cámara y la diferencia de luz, es necesario aplicar filtros a la imagen para poder obtener un valor más uniforme de píxeles, para esto se ha aplicado un filtro de mediana.

Es necesario promediar la imagen para poder obtener un único valor para cada componente RGB, y así realizar la respectiva identificación del color con solo 3 valores (RGB) que identifican el color del objeto, ya sea por técnicas de visión artificial o por lógica difusa.

Debido a que es posible que no exista objeto ubicado en la banda transportadora, es importante agregar esta validación de objeto para evitar que el motor se detenga innecesariamente.

C. CÁLCULO DE PARÁMETROS PARA IDENTIFICACIÓN DE TAMAÑO

Especificaciones:

- Pequeño (área figura < 19 cm²)
 - Mediano (19cm² ≤ área figura ≤ 38cm²)
- Grande (38 cm² < área figura).

A partir de la imagen binarizada (Figura5d) es posible obtener el área (píxeles x píxeles) que ocupa la imagen; como ya se ha obtenido el valor máximo que puede tomar la imagen definido por la resolución de la cámara (252 X 352) y como se conoce el área máxima en cm² definido por el cuadro de la imagen adquirida, se puede obtener el cambio de píxeles cuadrados a cm² definido por el área que ocupe la imagen a procesar.

Identificación de Tamaño por Visión Artificial

A partir del dato de Área en cm² y los rangos definidos (pequeño, mediano, grande), se determina a que rango pertenece cada valor y se determina si es grande, pequeño o mediano.

Identificación de Tamaño por Lógica Difusa

Para el reconocimiento de tamaño por lógica difusa se utiliza la misma entrada definida por visión artificial (Área en cm²).

Conjuntos de Entrada:

- Pequeño (P). Función: Z [12 27]
- Mediano (M) Función: Gaussiana [5 32]
- Grande (G) Función: S [36 48]

Conjuntos de Salida

- Pequeño: Trapecio [0 0 8 19].
- Mediano: Trapecio [16 21 34]
- Grande: Trapecio [37 43 60 61]

D. CALCULO DE PARÁMETROS PARA IDENTIFICACION DE COLOR

Para identificación de color se utiliza la imagen promediada, como entrada por ambas técnicas, estableciendo un único valor de cada componente RGB. Cada imagen adquirida tiene tres (3) matrices de 252 x 352 pixeles cada una, cada matriz representa un color primario, en este caso Rojo, Verde y Azul, formando la matriz RGB. Es entonces como cada píxel tiene un porcentaje de color Rojo, Verde y Azul, respectivamente. Los colores a identificar por ambas técnicas son: Café, Rojo, Rosa, Naranja, Amarillo, Verde, Cian, Azul, Blanco, Púrpura

Identificación de Color por Visión Artificial

Para la identificación por Visión Artificial se establecieron rangos para cada componente RGB, dependiendo de los valores observados en las pruebas realizadas en el entorno en el cual se desarrolla la identificación.

Identificación de Color por Lógica Difusa

Para la identificación de color por lógica difusa se definen tres (3) entradas Rojez, Verdez y Azules, respectivamente, con 5 conjuntos difusos cada una (Bajo (B) Medio Bajo (MB), Medio (M), Medio Alto (MA), Alto (Alto)); definidos con funciones de pertenencia Gaussianas, debido que el histograma de una imagen muestra un comportamiento Gaussiano. El establecimiento de los conjuntos de color se realizó en base a un estándar de componentes de color RGB.

E. CALCULO DE PARÁMETROS PARA IDENTIFICACIÓN DE FORMA

Las formas identificadas corresponden a figuras geométricas: cuadrado, rectángulo, trapecio, triangulo, circulo, elipse, rombo y estrella de cinco puntas, cruz o ninguna de las anteriores. Partiendo de la imagen binarizada (figura 5d), se

establecen las siguientes métricas de la figura: excentricidad, área, perímetro y número de esquinas.

A continuación se establecen variables que permitirán identificar la forma del objeto. Las variables establecidas (Ecuación 2) definen características de figuras geométricas. Es entonces como se puede clasificar cada figura geométrica, definiendo sus características típicas frente a la relación1, relación2 y su excentricidad

$$\text{Relación 1} = \frac{4\pi(\text{Area})}{(\text{Perimetro})^2} \quad \text{Ec. 2}$$

$$\text{Relación 2} = \frac{\text{Area}}{\text{Perimetro}}$$

Identificación de forma por Visión Artificial

Para la identificación por visión artificial se establecen experimentalmente los rangos típicos de relación1, relación2 y excentricidad que tiene cada una de las figuras geométricas a identificar y se clasifica según se ha preestablecido.

Identificación de forma por Lógica Difusa

Para el reconocimiento de forma por métodos de lógica difusa se toman las tres funciones de entrada que se utilizan en visión artificial y se dirigen hacia el FIS. La salida del FIS es entonces el nombre de la figura geométrica identificada, en el caso que sea posible.

III. PRUEBAS Y RESULTADOS

El material utilizado para todas las pruebas fue *fommy*, de 0.002 m de espesor. Inicialmente se realizaron las pruebas con figuras “perfectas”, a continuación se realizaron pruebas con figuras “imperfectas”, es decir, con colores de diferentes matices, formas con áreas adicionales o con ausencia de área. Todas las figuras son ubicadas en una posición aleatoria en la banda transportadora.

Pruebas con figuras “perfectas”

Resultados de Tamaño

Tipo	Desaciertos V.A	Desaciertos L.D
Pequeño	0	0
Mediano	0	16
Grande	0	17

Tabla1. Resultados sobre cincuenta (50) muestras.

Resultados de Color

Color	Desaciertos V.A	Desaciertos I.D
Café	3	2
Rojo	0	0
Rosado	0	0
Amarillo	1	1
Naranja	0	0
Cian	3	1
Azul	1	0
Púrpura	2	3
Verde	2	1
Blanco	0	0

Tabla2. Resultados sobre diez (10) muestras.*Resultados de Forma*

Figura	Desaciertos V.A	Desaciertos I.D
Círculo	1	1
Elipse	2	0
Rombo	3	2
Cuadrado	1	1
Rectángulo	2	0
Trapezio	2	2
Triángulo	2	1
Cruz	1	1
Estrella	2	2

Tabla3. Resultados sobre diez (10) muestras.*Pruebas con figuras “imperfectas”**Resultados de Tamaño*

Tipo	Desaciertos V.A	Desaciertos I.D
Pequeño	0	0
Mediano	0	22
Grande	0	18

Tabla4. Resultados sobre cincuenta (50) muestras.*Resultados de Color*

Color	Desaciertos V.A	Desaciertos I.D
Café	7	2
Rojo	2	1
Rosado	6	4
Amarillo	2	0
Naranja	3	2
Cian	3	2
Azul	4	1
Púrpura	8	3
Verde	5	1
Blanco	0	0

Tabla5. Resultados sobre diez (10) muestras.*Resultados de Forma*

Figura	Desaciertos V.A	Desaciertos I.D
Círculo	3	1
Elipse	3	2
Rombo	0	2
Cuadrado	0	3
Rectángulo	8	2
Trapezio	7	2
Triángulo	8	3
Cruz	8	3
Estrella	9	4

Tabla6. Resultados sobre diez (10) muestras.**IV. CONCLUSIONES**

- Es necesario realizar un preprocesamiento a las imágenes adquiridas debido a la presencia de ruido y otras perturbaciones.

- Por medio del control difuso de posicionamiento del servomotor, se tiene un adecuado perfil de velocidad con arranque y frenado suaves.

- La incidencia de luz es el principal factor que afecta los procesos de identificación, afectando en mayor forma la identificación de color.

- Para la identificación de color la técnica de visión artificial requiere definir y tener en cuenta todos los rangos para los valores R, G, B, en base a las condiciones del entorno de la figura. La técnica de lógica difusa puede definir los rangos en base a las características estándar de cada color

- La identificación de tamaño por visión artificial presenta menor tasa de desaciertos debido a que los rangos área se encuentran bien definidos en las especificaciones

- En base a los resultados se observa una menor tasa de desaciertos por técnica de lógica difusa cuando existen figuras imperfectas, debido a su manejo de incertidumbre y robustez frente a variaciones de las entradas de cada proceso.

V. BIBLIOGRAFIA

- [1] Reconocimiento de formas y Visión Artificial. Darío Maravall Gómez. Editorial ADDISON – WESLEY. Capítulo1. Introducción al reconocimiento automático de formas.
- [2]http://campusvirtual.uma.es/tdi/www_netscape/TEMAS/Tdi_29/index2.php#Filtro%20de%20mediana%20II
- [3] http://64.233.169.104/search?q=cache:4_eta4-lxpkJ:www4.ujaen.es/~satorres/practicas/practica3_vc.pdf+funcion+graythresh&hl=es&ct=clnk&cd=3&gl=co&lr=lang_es
- [4]http://gva1.dec.usc.es/~antonio/docencia/2003tratamientocomputacionaldeimagenes/P4_sbs/P4_sbs.html
- [5]<http://www.sia.eui.upm.es/master/VA05.pdf>

Viabilidad de la detección precoz de cáncer de cuello uterino por medio del análisis de la respuesta fluorescente natural en muestras ex-vivo

Diego R. Mahecha. Grupo CIR-IONIX, Grupo UNITEL-USTA. Jáder E. Guerrero, Grupo GOST-UIS

Resumen — Al intentar presentar un método específico para establecer el diagnóstico de lesiones precursoras y cáncer de cérvix, se consideraron diversas metodologías de estudio del análisis de la respuesta espectral fluorescente natural en muestras ex-vivo y sin el uso de fluorocromos; la primera se centra en el estudio microscópico de fluorescencia se realizaron estudios bajo la microscopía de fluorescencia y la segunda en el análisis de la respuesta de espectroscopía. La microscopía de fluorescencia y la detección precoz de cáncer de cuello uterino en muestras ex-vivo sin uso de fluorocromos, fue un método en donde se pretendía encontrar alguna respuesta a la excitación con luz ultravioleta(UV) en las imágenes obtenidas. Luego de innumerables pruebas con el microscopio se encontró que mediante esta técnica no era posible mostrar alguna diferencia entre una imagen obtenida de una muestra negativa y una positiva. También se realizó mediante reflectancia difusa y transmitancia, en ambos casos al obtener los espectros para su análisis, se encontró que no se presentaban características distintivas, aún con análisis de redes neuronales backpropagation. La conclusión del trabajo de investigación es que mediante la fluorescencia natural de las muestras de cuello uterino, no es posible proponer técnicas para la diferenciación entre una placa negativa y una con lesión precursora o con cáncer de cérvix.

Palabras claves—Fluorescencia, respuesta espectral, cáncer de cérvix, microscopio de fluorescencia, espectrofotometría.

I. INTRODUCCIÓN

A nivel mundial el cáncer de cérvix o cáncer de cuello uterino es la segunda causa de muerte más frecuente entre las mujeres después del carcinoma de glándula mamaria. Es un problema de salud pública que afecta a la mujer, especialmente en los países en vía de desarrollo; la alta tasa de mortalidad en su mayoría se debe a que las pacientes no acuden al examen oportunamente, en Colombia la tasa de mortalidad ha venido en aumento en donde se paso de una tasa de mortandad del 13.7 en el 2000 a 18.2 en el 2002¹. La no asistencia oportuna a la citología es ocasionada por el temor existente ante el examen o a la distancia que debe recorrer la paciente para la toma y luego la entrega del resultado. Por ello se estudian a nivel mundial otras formas más eficientes y rápidas de diagnóstico que la técnica tradicional del Papanicolaou.

¹ Fuente: agencia internacional para investigaciones en cáncer (IARC)

Uno de los campos de estudio es el de la fluorescencia y su aplicabilidad a la detección de cáncer de cérvix, los estudios muestran que es posible la discriminación de una muestra negativa de una con lesión precursora o con cáncer de cérvix, cuando se realiza el examen *in vivo* como se indica en [Big97], [Cha05], [Dre02], [Fer03], [Kon04], [Mac03], [Mej97] y [Ric04].

Tomando la idea de las referencias sobre estudios de fluorescencia y adaptándolas a nuestra realidad se generó la pregunta al problema de investigación, ¿es posible la diferenciación entre una muestra negativa y una muestra con lesión precursora o cáncer de cérvix, mediante el análisis de la respuesta espectral de muestras *in-vivo* excitadas con luz ultravioleta (UV) y sin el uso de fluorocromos?

Es necesario contrastar el resultado de la citología (expresado en la nomenclatura del sistema Bethesda 2001) con el análisis de la respuesta espectral de fluorescencia, para su correcta interpretación este debe realizarse sobre la misma muestra.

El desarrollo de investigaciones sobre el análisis de respuesta fluorescente es nuevo en nuestro medio, por ello, a medida que se profundizó más en el mismo, se llegó a resultados que hicieron cambiar la forma de adquirir la respuesta espectral de la muestra.

II. MARCO TEÓRICO

El cáncer cervicouterino empieza creciendo lentamente, los tejidos normales pasan por un proceso conocido como displasia, durante el cual empiezan a aparecer células anormales, después, las células cáncerosas comienzan a crecer y se diseminan con mayor profundidad en el cuello uterino y en las áreas circundantes. La citología con tinción de Papanicolaou muestra las células pre-malignas.[Ech02b]

Debido a que no hay síntomas asociados con el cáncer cervicouterino, suelen ser necesarias una serie de pruebas para diagnosticar la enfermedad: Citología con la tinción de Papanicolaou y si se encuentran células anormales se realiza una Biopsia.

De acuerdo a la lesión encontrada se clasifica según dos sistemas el SISTEMA PAP y Bethesda 2001, la Tabla presenta la nomenclatura para citología cervical. Se determinó usar la nomenclatura Bethesda debido a que era la utilizada por la Liga contra el Cáncer y PROFAMILIA.

TABLA I
NOMENCLATURA EN CITOLOGÍA CERVICAL²

SISTEMA PAP	SISTEMA BETHESDA	
Epitelio:escamoso/ y glandular	Epitelio escamoso	Epitelio glandular
PAP I – II Normal	Negativo para lesión intraepitelial o malignidad	Negativo para lesión intraepitelial o malignidad
	ASCUS* ASC-H*	Atipias glandulares
PAP III Sospechoso dudoso	EBG – LSIL* Cambios por VPH NIC I (displasia leve) LIEAG – HSIL* NIC II (displasia moderada) NIC III (displasia severa) Sospechosa de invasión	Atipias que favorecen neoplasias • Células endocervicales • Células glandulares
PAP IV Altamente sugestivo de malignidad CA IN SITU		Adenocarcinoma Endocervical in situ (AIS)
PAP V Positivo para malignidad	Carcinoma de células escamosas	Adenocarcinoma

Para el trabajo se irradió con luz ultravioleta (UV) las muestras citológicas, este proceso conocido como fluorescencia sucede cuando un sistema es excitado por medio de luz, con lo cual se pasa de un estado de menor energía a uno de mayor energía, cuando el sistema tiende a restablecerse el cuerpo comienza a emitir la energía que ganó de la excitación, la energía expulsada sale en forma de radiación electromagnética –REM– la cual es posible registrar. Una de las características más atractivas de los métodos de fluorescencia es su sensibilidad inherente, motivo por el cual, se escogió para el análisis de patologías de tejidos, [Mie82] [Mej97] [Per00].

III. ESTUDIO DE LOS DATOS

La espectroscopia de fluorescencia mide las características ópticas del tejido que se presentan por cambios bioquímicos o estructurales que acompañan las lesiones precursoras o el cáncer de cérvix. La nicotidamida adenina dinucleótido reducida (NADH) y la flavina adenina dinucleótido (FAD) son moléculas fluorescentes presentes en la mitocondria que tienen un importante rol en el metabolismo celular [Cha05].

El incremento de las coenzimas NADH y FAD, produce el aumento de la fluorescencia en la célula [Pal03] [Mil03] y en el tejido [Pav03], que se relaciona con un incremento en la actividad metabólica, y dependiendo del aumento en el porcentaje de fluorescencia se predice el grado de lesión en el que se encuentra la muestra [Cha05].

El diagnóstico se deriva de la extracción y análisis directo de la información más significativa de las mediciones de fluorescencia. Modelos matemáticos pueden ser desarrollados con base a los datos obtenidos de las mediciones, encontrando características que permitan la discriminación de células normales o negativas de células anormales o con presencia de lesiones precursoras o cáncer [Cha05].

Para la realización de las pruebas se trabajó con muestras citológicas de la Liga Santandereana contra el cáncer, muestras que previamente se encontraban clasificadas y analizadas por el citólogo de la Liga. La Tabla II presenta el número de placas y la distribución según la clasificación que presentaron.

TABLA II
PLACAS DE CITOLOGÍAS ORGANIZADAS SEGÚN EL DIAGNÓSTICO

Negativa	Negativa	15
Célula escamosa atípica	Ascus	15
	ASC-H	1
L-SIL	NIC I	15
	NIC II	15
	NIC III	15
H-SIL	Carcinoma in situ	12

No sólo se consideró el tipo de lesión, además se tuvo en cuenta los datos más relevantes del paciente y la composición de los reactivos utilizados para la preparación de las muestras.

Por medio del registro de la intensidad luminosa de la muestra al ser estimulada con luz ultravioleta sin la adición de fluorocromos, se realizó el proceso se con el microscopio Axiolab para luz transmitida y epifluorescencia (marca CARL ZEISS), el cual consta de una fuente de mercurio y una cámara CCD marca COHU, el *software* de captura fue LEICA QFISH V1.01 que permite la captura de las imágenes a diferentes tiempos de exposición de irradiación UV.

Al irradiar la placa con luz UV se produce una respuesta del tejido a la excitación, produciendo una imagen resultado de la fluorescencia, capturada por la cámara monocromática con un tiempo de exposición variante entre 0 y 20s.

La imagen mostrada en la Fig. 1 presentan la forma de las células dentro de una de las placas de estudio, al evaluarlas se determinó que la visual fluorescente natural no arrojó información contundente en el área que han sido establecidas como precancerosas por los métodos tradicionales. Y por tanto es difícil que sean empleadas como factor discriminante entre muestras negativas y positivas de lesión cervical, diferente a si se utilizan fluorocromos como en el estudio realizado por [Reb03].

Al evaluar el método para determinar las causas de la no obtención de los resultados esperados se encontró que: la potencia de la fuente de luz de estimulación es demasiado baja para la estimulación de fluorescencia a nivel celular. El trabajo

² Fuente: Liga Santandereana de lucha contra el cáncer

con muestras *ex vivo*, uno de los pilares de la investigación es el estudio de placas citológicas.¹



FIGURA 1. PLACA CON DIAGNÓSTICO NIC II³

Debido a lo expuesto anteriormente, se buscó la forma de incrementar la potencia de la fuente de excitación, pero en el laboratorio de genética humana es imposible realizar un cambio al sistema debido a que con ese equipo se trabaja en pruebas de laboratorio de cariotipos; por ello se examinó la posibilidad de trasladar la investigación.

Una nueva forma de realizar el análisis de medición de espectroscopia UV-Visible se realizó con el equipo de espectrofotometría Shimadzu serie UV- 2200 y el software UVPROBE en la opción espectro.

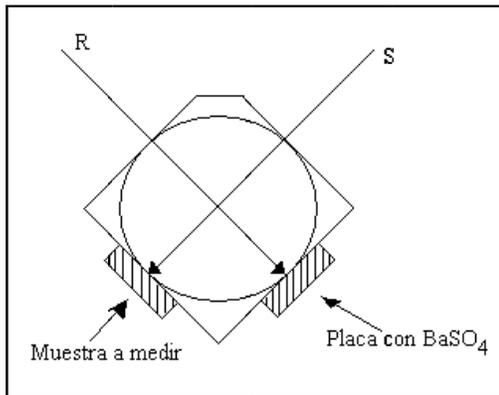


FIGURA 2. MEDICIÓN DE REFLECTANCIA.

La toma de datos se realizó en dos etapas, la primera se utilizó la reflectancia difusa para la toma de datos y en la segunda la transmitancia. Las Figuras 2 y 3 muestran la forma

¹ En estudios de doctorado se tiene pensado el abordar la pregunta ¿cómo afecta la excitación con luz UV al tejido del cérvix en muestras *in vivo*?

³ Placa de la Liga número de referencia 04-09944. Tiempo de exposición 20 [s], objetivo x20.

de colocación de la muestra dentro de la cámara para realizar la medición correspondiente.

La figura 2 muestra cómo dentro de la cámara dos haces de luz inciden sobre el dispositivo uno de ellos es el de referencia y el otro incide sobre la muestra, al caer el haz de luz sobre la muestra en ella se produce una reflexión que es medida y comparada con la línea base.

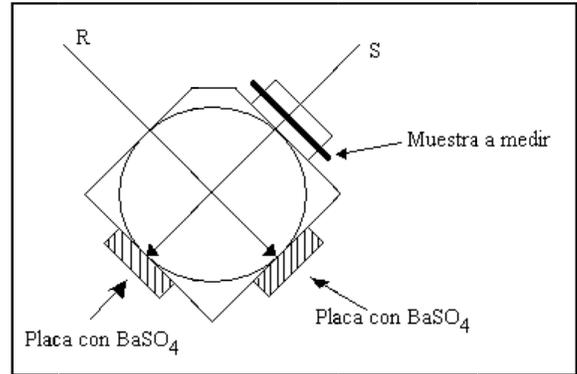


FIGURA 3. MEDICIÓN DE TRANSMITANCIA.

La figura 3 muestra el proceso para la medición de transmitancia en una placa, existen dos haces de luz uno pasa a través de la muestra y el otro es el de referencia. Los resultados obtenidos al colocar las muestras dentro de la cámara por medio del análisis de reflectancia y transmitancia se muestran a continuación.

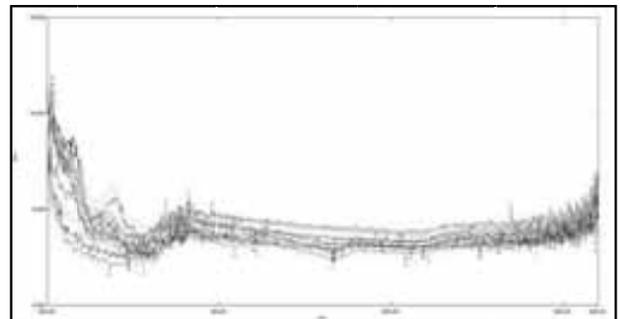


FIGURA 4. SEÑALES ESPECTRALES DE REFLECTANCIA PARA PLACAS POSITIVAS

La figura 4 muestra las señales espectrales de muestras citológicas con lesiones de alto grado y carcinoma.

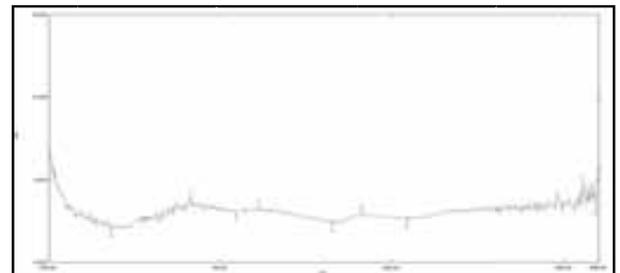


FIGURA 5. PROMEDIO DE LAS SEÑALES ESPECTRALES POSITIVAS²

² Se tomaron las 14 señales espectrales para sacar el promedio

TABLA III
DATOS MÁXIMOS Y MÍNIMOS PROMEDIO DE LAS PLACAS POSITIVAS³

No.	Pico/valle	Longitud de onda (nm)	% R
1	⬆️	822,00	4,701
2	⬆️	564,00	2,936
3	⬆️	445,00	3,281
4	⬆️	364,00	3,793
5	⬆️	836,00	3,288
6	⬇️	617,00	2,589
7	⬇️	530,00	2,373
8	⬇️	418,00	3,065
9	⬇️	272,00	2,046

La tabla III muestra los valores más representativos de la señal espectral de la figura 5 con el respectivo porcentaje de reflectancia producido. Los datos que presentan una flecha hacia arriba representa los valores pico de la señal, los mostrados con una flecha hacia abajo representan los valles de la misma.

Un procedimiento similar se realizó para las muestras negativas arrojando los resultados expresados en la tabla siguiente.

TABLA IV
DATOS PICO PARA LA SEÑAL ESPECTRAL DE M07A, PLACA NEGATIVA

No.	Pico/valle	Longitud de onda (nm)	% R
1	⬆️	801,00	3,348
2	⬆️	554,00	2,708
3	⬆️	423,00	3,159
4	⬆️	364,50	3,822
5	⬆️	206,50	23,357
6	⬇️	806,00	2,759
7	⬇️	628,00	2,454
8	⬇️	539,00	2,583
9	⬇️	494,00	2,690
10	⬇️	421,00	3,049
11	⬇️	289,00	2,129
12	⬇️	194,50	20,204

Al comparar los resultados de las placas con los promediados y por grupos no se encontró ninguna característica particular entre señales pico se muestras positiva y negativas.

Como último recurso se realizó el análisis de los datos mediante dos programas realizados en MATLAB; se desarrollaron dos programas el primero encargado de seleccionar los datos, la distribución de las muestras de entrenamiento y prueba de la red, y el segundo una red neuronal⁴ *backpropagation*.

⁴ Las redes de neuronas artificiales (RNA) son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Consiste en simular las propiedades observadas en los sistemas neuronales biológicos a través de modelos

El primer programa tomaba los datos obtenidos por las pruebas de reflectancia y transmitancia, tomaba dos grupos de estudio una para entrenamiento de la red y otro para la validación de los datos, se normalizaban los mismos y se arreglan en un solo vector para su ingreso a la red neuronal.

En programa de la red neuronal *backpropagation* se introducen el vector de datos encargado de seleccionar la mejor configuración para la red, el proceso dura 50 horas continuas, tiempo en el cuál se producen algunas redes para las pruebas; al tomar el segundo conjunto de datos para analizar los resultados obtenidos no se obtuvieron los mejores resultados.

La comprobación no arrojó ninguna red que pudiese discriminar una muestra negativa de una positiva. Al buscar las posibles causas para los resultados obtenidos se encontró que el factor que podía tener mayor relevancia es la toma de datos con muestras *ex-vivo*.

IV. CONCLUSIONES

Este trabajo es un primer acercamiento en al estado del arte y el aspecto metodológico de la determinación cáncer de cuello uterino mediante técnicas de microscopia y espectroscopia de fluorescencia.

Se encontró que al realizar la investigación para la determinación del estudio de la respuesta fluorescente natural en muestras *ex-vivo*, no es posible realizar una separación entre placas negativas y positivas.

Se desarrollo un procedimiento numérico para clasificar los tejidos sanos, con lesión o carcinoma a través de MATLAB, dicho procedimiento reforzó la hipótesis nula⁵ de la investigación. Siempre correlacionando los resultados obtenidos con el sistema internacional Bethesda 2001.

La investigación abre un camino para la realización de estudios en el campo de la fluorescencia especialmente al desarrollo de técnicas para la respuesta fluorescente de muestras *in-vivo*⁶. Estos nuevos investigadores poseerán una herramienta para direccionar su investigación.

Los principales actores responsables de la fluorescencia en las células del cérvix son el NADH y FAD, estas coenzimas participes del metabolismo requieren estudios adicionales para que futuras investigaciones den más luces en la forma de trabajo de ellas.

matemáticos recreados mediante mecanismos artificiales (como un circuito integrado, un ordenador o un conjunto de válvulas). El objetivo es conseguir que las máquinas den respuestas similares a las que es capaz el cerebro que se caracterizan por su generalización y su robustez

⁵ Mediante la fluorescencia natural de las muestras de cuello uterino *ex-vivo*, no es posible diferenciar el tejido anormal del tejido sano por medio del análisis de la respuesta espectral.

⁶ Este nuevo trabajo de investigación será realizado en estudios de doctorado.

REFERENCIAS Y BIBLIOGRAFÍA

- [Alt94a] Altman, D. Bland, J. *Statistics Notes: Diagnostic tests 1: sensitivity and specificity*. 1994.
- [Alt94b] Altman D.G., Bland J.M. *Statistics Notes: Diagnostic tests 2: predictive values*. 1994.
- [Ben03] Benavides, J.M. Et al. *Multispectral digital colposcopy for in vivo detection of cervical cancer*. *Optics Express*. Vol 11, No 10. Mayo 2003.
- [Bel01] Belinson, J. Qiao, Y.L. Et al. *Shanxi Province Cervical Cancer Screening Study: A Cross-Sectional Comparative Trial of Multiple Techniques to Detect Cervical Neoplasia*. *Gynecologic Oncology*. Septiembre de 2001.
- [Big97] Bigio, I. *Ultraviolet and visible spectroscopy for tissue diagnostics: fluorescent spectroscopy and elastic-scattering spectroscopy*. *Phys. Med. Biol.* 42. 1997.
- [Car05] Carlson, K. Et al. *Confocal microscopy: imaging cervical precancerous lesions*. *Gynecologic oncology*. Septiembre 6 de 2005.
- [Cap04] Cabriales, R. Luminiscencia en polímeros semiconductores. *Revista Ingenierías*, Vol. VII, No 23. Nuevo León, México. Abril-Junio 2004.
- [Cha05] Chang, S.K. Pavlova, I. Follen, M. Richards, R.K. *Fluorescence spectroscopy as a tool for detecting cervical pre-cancer*. *Science Direct*. EL SEVIER. 2005.
- [Dre02] Drezer, R. Follen, M. Richards, R.K. *Optical imaging for the detection of cervical cancers in vivo*. IEEE. 2002.
- [Dre03] Drezer, R. Et al. *Optical imaging of the cervix. Second international conference on cervical cancer. Supplement to cancer*. Abril 29 de 2003.
- [Duj94] Dujardin, B. Van, J. Van, A. Unger, J. Van, P. *Likelihood ratios: a real improvement for clinical decision making?*. 1994.
- [Ech02a] Echeverri, J.H. Trabajo de investigación sobre la sensibilidad y especificidad de la prueba de papanicolau para la detección de infección por VPH en hombres. Bucaramanga. 2002.
- [Ech02b] Echeverri, J.H. *Marco Teórico Cáncer*. Bucaramanga. 2002.
- [Fer03] Ferenczy, A. Faupel, M. *Optical Imaging of the Cervix. Second International Conference on Cervical Cancer*. Houston, Texas. American Cancer Society. 2003.
- [Fic98] Ficha de datos de seguridad. MERCK. Diciembre 15 de 1998.
- [Fin05] Finlay, J.C. Foster, T.H. *Recovery of hemoglobin oxygen saturation and intrinsic fluorescence with a forward-adjoint model*. *Appl Opt*. Abril de 2005.
- [Gom06] Gómez, A. *Página de Internet*. http://www.tuotromedico.com/temas/cancer_uterino.htm. Enero 27 de 2006.
- [Huu04] Huh, W. Et al. *Optical detection of high-grade cervical intraepithelial neoplasia in vivo: Results of a 604-patient study*. *American journal of obstetrics and gynecologic*. 2004.
- [Kon04] Kong, S. *"Analysis of hyperspectral fluorescent images for poultry skin tumor inspection"*. *Applied Optics*. Vol.43 No. 4. 1 de Feb. de 2004.
- [Lac05] Lacruz, P.C. Nomenclatura de las lesiones cervicales. 7º congreso hispanoamericano de anatomía patológica. http://www.conganat.org/7congreso/final/vistalmpresion.asp?id_trabajo=538. Septiembre de 2005.
- [Mac03] MacRae, M. *Solutions for materials analysis. Spectroscopy as an Emerging Tool for Cancer Diagnostics and Monitoring*. Review Spectroscopy, 2003.
- [Mee85] Meehan. *Optical Methods. Fundamentals of spectrophotometry*. 1985.
- [Med06] MedlinePLUS. Prueba Papanicolau. <http://www.nlm.nih.gov/medlineplus/spanish/ency/article/003911.htm>. Página internet. Septiembre 2006.
- [Mej97] Mejía, E. *Tesis de Grado. Implementación de un método de análisis cualitativo y cuantitativo por espectroscopía de fluorescencia inducida por láser*. Universidad Industrial de Santander. 1997.
- [Mie82] Mielenz, Klaus D. *Measurement of photoluminescence*. New York: academic press, 1982.
- [Mil05] Milbourne, A. Et al. *Result of pilot study of multispectral digital colposcopy for the in vivo detection of cervical intraepithelialneoplasia*. *Gynecologic oncology*. Octubre 3 de 2005.
- [Mil03] Millot, C. Bondza, P. Millot, J.M. Lallermand, A. Manfait, M. *Autofluorescence spectroscopy of malpighian epithelial cells, as a new tool for analysis of cervical cancer precursors*. *Histol Histopathol*. Abril de 2003.
- [Mül01] Müller, M. and Georgakoudi, Irene. *Intrinsic fluorescence spectroscopy in turbid media: disentangling effects of scattering and absorption*. *Applied Optics*. Vol 40. 2001.
- [Nai02] Nair, M.S. Ghosh, N. Raju, N.S. Pradhan, A. *Determination of optical parameters of human breast tissue from spatially resolved fluorescence: a diffusion theory model*. *Appl Opt*. Julio de 2002.
- [Ops05] Organización Panamericana de la Salud. *Cáncer Cervicouterino*. Hoja informativa. Página de Internet. Septiembre de 2005.
- [Pag04a] *Página de Internet*. <http://www.shu.ac.uk/schools/sci/chem/tutorials/molspec>. UV-Vis Luminescence Spectroscopy Theoretical principles. Agosto 2004.
- [Pag04b] *Página de Internet*. http://www.ub.es/biocol/wbc/tecnicas/principio_microscopia_fluorescencia.htm. Principios de la microscopía de fluorescencia. Agosto 2004.
- [Pal03] Palmer, G.M. Keely, P.J. Breslin, T.M. Ramanujam, N. *Autofluorescence spectroscopy of normal and malignant human breast cell lines*. *Photochem Photobiol*. Noviembre 2003.
- [Pav03] Pavlova, I. Sokolov, K. Drezek, R. Malpica, A. Follen, M. Richards-Kortum, R. *Microanatomical and biochemical origins of normal and precancerous cervical autofluorescence using laser-scanning fluorescence confocal microscopy*. *Photochem Photobiol*. Mayo de 2003.
- [Per00] Pérez, F. Tesis de grado. *Diseño y montaje de un sistema espectroscópico para espectroscopía Raman intensificada por efecto superficial*. Universidad Industrial de Santander. Bucaramanga. 2000.
- [Rah05] Rahman, M. Et al. *Optical imaging of cervical pre-cancer with structured illumination: An integrate approach*. *Gynecologic oncology*. Septiembre 9 de 2005.
- [Ric04] Richards-Kortum, Rebecca. Informe sobre estudios clínicos sobre una sonda de detección óptica de cáncer cervical. Universidad de Texas en Austin. *Página de Internet "Anderson Cancer Center home page"*. Marzo de 2004.
- [Rid06] Riddle, D. Revisión de producto: espectrómetros *Ocean Optics* y software. *Advanced Aquarist's Online magazine*. Página de Internet. Septiembre de 2006.
- [Rod04] Rodríguez, E. Klinger, J. Montaña, R. Castro, W. Bastidas, A. Modelo de construcción de un fotómetro de dispersión y fluorescencia aplicable al estudio de células patológicas. Universidad del Cauca. Popayán, Colombia. 2004.
- [Sac94] Sackett, D. Haynes, R. Guyatt, G. Tugwell P. *Epidemiología clínica. Ciencia básica para la medicina clínica*. 2ª ed. Madrid, España. Editorial médica panamericana. 1994.
- [San80] Sandler G. *The importance of the history in the medical clinic and the cost of unnecessary test*. 1980.
- [Sko94] Skoog, D.A. Leary, J. J. *Análisis instrumental*. Madrid, España. McGraw-Hill. 1994.
- [Tum98] Tumer, K. Ramanujam, N. Ghosh, J. Richards-Kourtum, R. *Ensembles of radial basis function networks for spectroscopic detection of cervical precancer*. *IEEE. Transaction on biomedical engineering*. Agosto, 1998.
- [Uni87] *Espectroscopía Raman y Estructura molecular*. Universidad de Concepción, Departamento de Química. Chile. 1987.
- [Wu04] Wu, Y.C. Xi, P. Qu, J. Cheung, T.H. Yu, M.Y. *Depth-resolved fluorescent spectroscopy reveals layered structure of tissue*. *Optics Express*. Julio de 2004.

Diego R. Mahecha. Magíster en Ingeniería área Electrónica, 2006– Universidad Industrial de Santander. Auditor Interno en sistemas de gestión de la calidad ISO9001:2000, 2006- Universidad Pontificia Bolivariana. Ingeniero Electrónico, 2003– Universidad Industrial de Santander.

Director de Calidad de la Empresa IONIX Ltda, dedicada al diseño de equipos electrónicos, director de grupo de Investigación CIR-IONIX, miembro del grupo de investigación UNITEL-USTA. Docente en el área de señales y sistemas, procesamiento digital, teoría electromagnética en las instituciones Santo Tomás, Universidad Industrial de Santander y Unidades Tecnológicas de Santander.

Reconocimiento de imágenes en una mesa XY mediante lógica difusa

Yonathan Eduardo Corredores Hurtado, Johnny Fernando Martínez Vera,
Iveth Liliana Holguín Carrillo y Enrique Estupiñán Escalante.

ABSTRACT. - *This project is an application of pattern recognition, in this case form, color, size and texture of different objects in two dimensions. The implementation is done using fuzzy logic. The information to classify includes number of sides, length of axes, area for the shape and average of the color of objects.*

Este proyecto es una aplicación de reconocimiento patrones, en este caso de forma, color, tamaño y textura de diferentes objetos en dos dimensiones. La implementación es hecha usando lógica difusa. La información a clasificar incluye número de lados, longitud de los ejes, área y promedio de color del objeto.

I. INTRODUCCION

En la sociedad moderna, la tecnología juega un papel fundamental en la disminución de los costos y el tiempo requerido para cumplir con las tareas asignadas; en la industria; procesos tales como la producción en masa de automóviles o de dispositivos electrónicos son los mejores ejemplos. Algunos procesos de automatización basan sus esfuerzos en la clasificación de objetos; es así como en una línea de producción se requiere manejar cada vez más tecnología, permitiendo que tareas en control de calidad evolucionen de inspección visual a inspección automática por diferentes métodos, uno de ellos es la visión artificial la cual tiene algoritmos que se ajustan a distintos niveles de precisión, muchas veces reduciendo el costo y el tiempo empleado en las diferentes tareas.

En el área de la seguridad también se realizan varios estudios de identificación de objetos, ya que se necesita tener un estricto control sobre diferentes elementos tales como armas, explosivos, y algunos otros que pueden atacar contra personas o contra la infraestructura. Los algoritmos usados para estas tareas son complicados de programar y se invierten grandes capitales en temas de seguridad que al cabo de algunos años benefician a las industrias.

La identificación de los objetos en la mesa XY, es un proceso de clasificación de acuerdo a su tamaño, color,

textura y forma, que se desarrolla con un software de procesamiento de imágenes. Para alcanzar este objetivo es necesario establecer parámetros característicos de los objetos a clasificar, para lo cual se usa un sistema de visión artificial que incluye la captura de imágenes de la mesa, el procesamiento de las mismas y luego por medio de lógica difusa se selecciona el objeto buscado.

II. OBJETIVOS

OBJETIVO PRINCIPAL

- Reconocer objetos según su forma, color, tamaño y textura en la mesa XY, usando lógica difusa.

OBJETIVOS SECUNDARIOS

- Usar técnicas de visión artificial para el procesamiento de imágenes.
- Comparar los métodos de clasificación de imágenes por lógica difusa y por los métodos tradicionales de visión artificial.

III. DESCRIPCIÓN

MESA X-Y

La parte mecánica y de control de la mesa x-y está formada por una mesa de acrílico transparente de 35cm x 50cm que posee tres carriles con tornillos sin fin (dos para el movimiento en X y uno para el movimiento en Y), dos interruptores de fin de carrera que detectan la posición origen de la mesa ($X=0, Y=0$) y dos motores paso a paso los cuales permiten el movimiento de los tornillos en el eje X y en el eje Y; a demás su parte de control está constituida por un sistema electrónico que sirve de puente entre el computador y los motores que son energizados con pulsos en valores fijos de 12 y 0 voltios DC. El Computador maneja la posición relativa de los motores por medio de un cable RS232 y una interfaz de procesamiento.

Sobre el sistema de movimiento se monta una base, en esta se fija la cámara para capturar las imágenes y así formar el mapa en el que se hace el reconocimiento de objetos.

Para el reconocimiento de las imágenes y su procesamiento se sigue el algoritmo presentado en la figura 1.

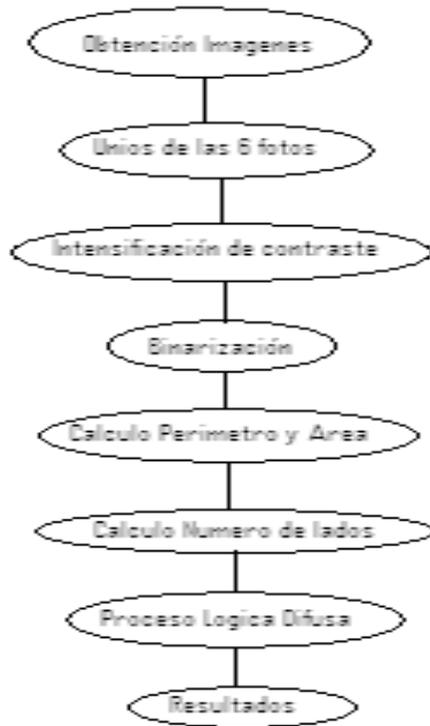


Figura1

IV. DESARROLLO

a) PARTE MECÁNICA Y CONTROL

Los motores paso a paso se mueven por medio de una secuencia de pulsos que llegan en grupos de cuatro bits de entrada, una secuencia para movimientos positivos, y otra secuencia para movimientos negativos. La mesa tiene un área efectiva de trabajo de 25cm x 26cm, con un fondo negro opaco de tela. Sobre esta área se colocan las diferentes figuras.

El micro-controlador es un PIC 16F877A. Su función es servir de interfaz entre el programa y los dos motores a través del puerto RS-232. Por medio de este puerto se envían datos al micro-controlador, que a su vez envía la posición relativa en el eje x y en eje y. El micro-controlador envía al computador una señal de retorno de dos bits que indican la posición cero en X y cero en Y.

b) PRE-PROCESO

En la figura 2 se muestra la interfaz del programa. En este se introducen las características de los objetos que buscan en la mesa x-y.



Figura 2 (IMAGEN COMPLETA DE LA MESA).

En primer lugar se captura la imagen recorriendo por completo la mesa y tomado seis fotos que son almacenadas y concatenadas para formar el mapa que puede divisarse la figura 2, la siguiente operación es pasar la imagen del formato RGB a una escala de grises.

A continuación se realiza una intensificación de contraste ya que las fotos tomadas no fueron de la mejor calidad, de esta forma ya no se tienen problemas con el reflejo de la luz sobre el fondo, como se ve en la figura 3. El proceso consiste en tomar cada bit del mapa, verificar si está por encima de 200 (en un rango de 1 a 255), se les asigna 0 y los que están por debajo los envía a 255. El nivel de 200 se establece con una serie de ensayos con el fin de establecer el rango donde no hay objetos falsos que compliquen el análisis.



Figura 3 (INTENSIFICACIÓN)

Como paso a seguir se toma la imagen intensificada y luego se pasa a la binarización (figura 4) que lleva los valores a cero, si el valor del bit está entre 0 y 199 o a 255 si este valor se encuentra entre 200 y 255^[7]. Dejando como resultado la imagen de la figura 5.

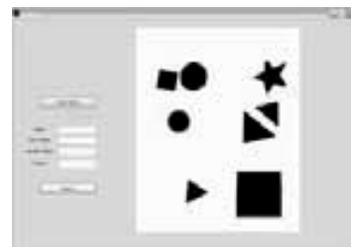


Figura 4 (BINARIZACIÓN)

Para más comodidad al trabajar los objetos se Trabaja con el negativo de esta imagen

Luego se filtra la imagen con un filtro de mediana y a demás elimina cualquier objeto que fuese menor en pixeles a un área de 5 cm^2 tomada como carácter mínimo para no tomar objetos falsos que no fueron eliminados pese a los tratamientos realizados a la imagen^[6].

c) ANALISIS DEL MAPA

Con el mapa ya libre de errores debidos a la calidad de la imagen tomada por la cámara se pueden extraer: el centroide (figura 5), el eje mayor, el eje menor de cada figura, el rectángulo circunscrito, la excentricidad y el área. Usando la resta de los ejes obtenemos una de las entradas al sistema difuso; con esto podemos diferenciar los rombos de los cuadrados y los círculos de las elipses.

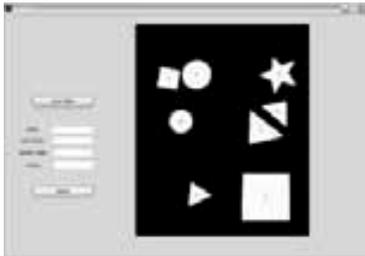


Figura 5 (CARACTERISTICAS OBJETOS)

Lo siguiente es extraer los bordes de los objetos para de esta manera poder buscar las formas que ellos describen, esto por medio del algoritmo *Canny*; que usa una máscara de convolución la primera derivada; ya que los puntos de contorno son como zonas de píxeles en las que existe un cambio brusco de nivel de gris^[7].

A partir de esto se calcula el perímetro de cada figura, esto se hace buscando los píxeles blancos en la figura, se divide el área total de la imagen en los sectores más pequeños obtenidos en la extracción de bordes, lo anterior para no incluir píxeles de otras figuras.

Luego se debe correr un algoritmo que suaviza el borde de las figuras. Esto se hace tomando una matriz de 3 píxeles por N píxeles (N es el numero de filas de la imagen), luego se operan sus filas usando la función OR y se asigna el resultado a cada una de las 3 columnas de la matriz, de esta forma todas las columnas tendrán el mismo valor por lo cual quedan los bordes de las figuras uniformes y con esta nueva imagen se puede correr el algoritmo de búsqueda de lados (ver figura 6). Este último algoritmo analiza cada columna de la figura y busca cambios de uno a cero o viceversa en toda la imagen, si encontramos un número impar de cambios podremos saber que en esta columna hay un ángulo que muestra el principio de uno o más lados.



Figura 6 (DETECCIÓN DE ESQUINAS)

IV. LOGICA DIFUSA

a) FUZZY FORMA

Para el análisis de la imagen por medio del sistema de inferencia difuso se decide trabajar con tres entradas: Relación perímetro-área^[1], número de lados y relación entre eje mayor y menor (ver figura 7). Se definen 7 reglas y se usa el método Mamdani y centroide para la defusificación.



Figura 7 (SISTEMA DE INFERENCIA DIFUSA PARA FORMA)

La primera entrada Perímetro-área, tiene un universo de discurso definido por los reales positivos con un rango de $[0 \ 30]$, ya que algunas figuras geométricas como los cuadrados, círculos y triángulos tienen esta relación bien definida. Se decidió tener 5 valores lingüísticos, el menor rango cubre las figuras que no tienen esta relación definida y el rango mayor por la relación de una estrella de 5 puntas. La primera función de pertenencia llamada "NADA", es una función Z con puntos de corte ($Z(0)=1$, $Z(12)=0$), la cual hace referencia a que ninguna figura tiene relación perímetro-área un estos valores. La segunda función de pertenencia llamada "CIRCULO", es una función triangulo con puntos de corte ($T(9)=0$, $T(12)=1$, $T(16)=0$) ya que la relación de un círculo es 12. La tercer función de pertenencia se llama "CUADRADO", esta es una triangulo con puntos de corte ($T(12)=0$, $T(16)=1$, $T(20)=0$) ya que la relación es de 16. La cuarta llamada "TRIANGULO", es una función triangulo también, con cortes en ($T(16)=0$, $T(20)=1$, $T(23)=0$). Por último la quinta función llamada "ESTRELLA", es una S con cortes en ($S(19)=0$, $S(26)=1$) ya que la relación en estas figuras es la mayor de todas.

Nuestra segunda entrada. Lados, con universo de discurso formado por los reales positivos, y un rango $[3 \ 5]$, está conformada por tres valores lingüísticos, ya que el

algoritmo de detección de lados entrega solo objetos de 3, 4 o 5 lados. La primera función es una función Z llamada "3" con puntos de corte en $(Z(0)=1, Z(4)=0)$ que va a ser la entrada correspondiente a triángulos. La segunda función es una función llamada "4" con puntos de corte en $(T(3)=0, T(4)=1, T(5)=0)$ la cual hace referencia a las figuras como: cuadrado, rectángulo y rombo. La tercera función es una S llamada "5", con puntos de corte en $(S(4)=0, S(10)=1)$ y que hace referencia a las estrellas, círculos y elipses, estos dos últimos se incluyeron en esta función de pertenencia ya que su número de lados tiende infinito.

Por último la entrada final es la Relación Ejes, el universo de discurso son los Reales Positivos con un rango de $[0, 4]$ esta entrada es necesaria ya que algunas figuras como los rectángulos no pueden ser definidos con las anteriores funciones de pertenencia ya que tiene 4 lados como los cuadrados y rombos pero la relación perímetro-área no es una constante, variando con el tamaño del objeto, pero en cambio si presenta una diferencia entre sus ejes, lo cual hace necesario una nueva función de pertenencia. La primera función es un triángulo llamada "CERO", con cortes en $(T(0)=0, T(1)=1, T(2)=0)$, esta entrada dice si los ejes son iguales, se da un holgura de 10 píxeles en un rango de $[0, 700]$ para tomar esta decisión lo cual muestra objetos como círculos y cuadrados. La segunda función es también un triángulo y se llama "UNO", con cortes en $(T(1)=0, T(2)=1, T(3)=0)$, esta se refiere a si la resta de los ejes fue menor a la mitad del eje mayor esto para seleccionar objetos como elipses y rombos. La última función también es un triángulo llamado "DOS", con cortes en $(T(2)=0, T(3)=1, T(4)=0)$ y esta se refiere a si la resta de los ejes es mayor a la mitad del eje mayor para seleccionar objetos como rectángulos.

Como salida se tienen 7 posibilidades: Triángulo (función Z [$Z(0)=1, Z(3)=0$]), Cuadrado (función Triángulo [$T(3)=0, T(4)=1, T(5)=0$]), Rombo (función Triángulo [$T(5.5)=0, T(6)=1, T(7)=0$]), Rectángulo (función Triángulo [$T(7)=0, T(8)=1, T(9)=0$]), Elipse (función Triángulo [$T(9)=0, T(10)=1, T(11)=0$]), Círculo (función Triángulo [$T(11)=0, T(12)=1, T(13)=0$]), Estrella (función S [$S(13)=0, S(14)=1$]). Esta segmentación se hace para que no se traslapen las salidas y sea más fácil tomar una decisión de la figura reconocida, en cuanto al orden de la segmentación no tiene ningún significado.

Se crean 7 reglas (ver figura 8) las cuales deciden según las entradas cual será la mejor respuesta, se escoge este número de reglas porque es el número de objetos que se pueden reconocer. Cada combinación de entradas tiene un peso en cada una de las reglas, y según el peso en cada una de estas reglas el programa decide cual sería la mejor respuesta, esto usando una "defuzificación" por centroide.



Figura 8 (REGLAS LÓGICA DIFUSA FORMA)

b) FUZZY COLOR.

En cuanto a la lógica difusa para el color encontramos 3 entradas: ROJO, AZUL, VERDE y una salida COLOR (ver figura 9). Aquí se usan 27 reglas, el método Mamdani y se defuzifica con el método del Centroide.

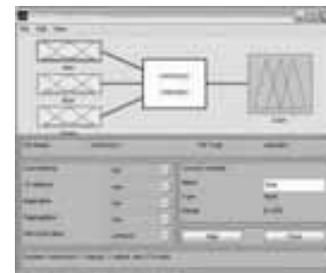


Figura 9 (LÓGICA DIFUSA PARA COLOR)

La primera entrada ROJO consta de 3 valores lingüísticos que nos muestra tres clases de rojo; "ROJO CLARO", representado con una función Z con cortes $Z(0)=1$ y $Z(95)=0$, "ROJO MEDIO", representado por una función triángulo con cortes $T(65)=0, T(127)=1$ y $T(180)=0$, y, "ROJO OSCURO", representado por una función S con cortes $S(160)=0$ y $S(255)=1$. Se define así porque los valores cercanos a 0 tiene un significado de Rojo claro, solo después de 50 se aproximan a rojo medio lo cual es extraordinariamente representado por un función tipo Z, similarmente el rojo oscuro es descrito por la función S. En el caso del rojo medio se implementa la función triángulo ya que es una transición entre las otras dos tonalidades.

Para el verde y el azul se usan las mismas funciones de pertenencia, azul claro con función de pertenencia Z, para el azul oscuro función de pertenencia S y el azul medio con función de pertenencia triángulo.

Se tienen 27 reglas las cuales deciden según las entradas cual será la mejor respuesta. Cada combinación de entradas tiene un peso en cada una de las reglas, y según el peso en cada una de estas reglas el programa decide cual sería la mejor respuesta, esto usando una "defuzificación" por centroide.

c) FUZZY TAMAÑO

En cuanto a la lógica difusa para el tamaño solo se tiene una entrada: "ENTRADA" y una salida "SALIDA" (ver figura 10). Aquí tan solo existen 3 reglas ya que teniendo el área en píxeles de la figura se sabe si es grande o pequeña según los requerimientos del proyecto. Se usa el método Mamdani y se "defuzifica" con el método del Centroide^[4].

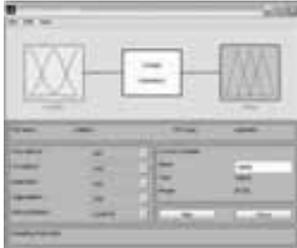


Figura 10 (LÓGICA DIFUSA PARA TAMAÑO)

Nuestra única entrada "ENTRADA" consta de 3 funciones que nos muestra tres tamaños (ver figura 11); "PEQUEÑO", representado con una señal Z con cortes en $(T(0)=1, Z(16)=0)$, "MEDIANO", representado por una función triángulo con cortes en $(T(4)=0, T(12)=1, T(20)=0)$, "GRANDE", representado por una función S con cortes en $(S(16)=0, S(25)=1)$.



Figura 11 (ENTRADA LÓGICA DIFUSA TAMAÑO)

Como salida tenemos igual tres funciones, al igual que la entrada, la diferencia aquí es que estas funciones no se traslapan, esto para hacer un poco más precisa la defuzificación (ver figura 12).

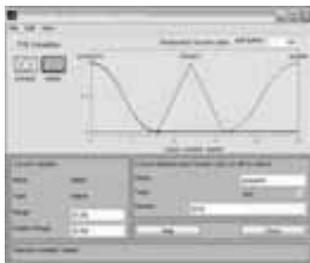


Figura 12 (SALIDA LÓGICA DIFUSA TAMAÑO)

V. TEXTURA

En las diferentes texturas se distinguen patrones que diferencian una tela, de papel o de plástico permitiendo

hacer una clasificación de acuerdo los resultados de un análisis por medio de histogramas de la transformada de Fourier. Previamente se realiza un binarizado en la imagen para poder establecer un criterio de agrupamiento, ya que con la transformada de Fourier se establecen la posición de los picos más altos respecto al centro y se pueden distinguir varios patrones que establecen los descriptores geométricos de la transformada de Fourier sin importar las variaciones de rotación, translación^[8].

Utilizando la transformada de Fourier para su visualización adecuada se implementa el algoritmo en Matlab, luego se determina el histograma para cada imagen ya que este da un umbral de binarización determinando, de esta forma se obtiene el vector característico para cada imagen según la textura de cada objeto, siendo posible a través de los descriptores geométricos (determinados a través de los picos máximos de la transformada de Fourier) obtener los correspondientes centros de cada grupo de frecuencias que componen el tejido principal.

VI. PROBLEMAS

La luz es el problema más grande, ya que si hay demasiada, la mesa brilla y las imágenes se confunden con el fondo, pero de igual manera si hay poca luz la cámara no alcanza a tomar una buena foto, obteniendo una imagen completamente negra

Aunque la mesa se cubre con una tela negra y se adaptan unos bombillos encima del área donde se toma la imagen, sigue existiendo un problema con el fondo, ya que la cámara no lo ve como negro y el software puede confundirlo con una nueva figura. Para solucionar esto se implementa un filtro más restrictivo lo cual dejó algunos colores por fuera del reconocimiento.

Cuando se extraen los bordes de la imagen estos no quedan parejos lo cual hace que en algunas posiciones de la imagen pueda aparecer abierta, esto se soluciona haciendo una redundancia en los bits en blanco de la figura.

Los colores de una misma imagen no son homogéneos, esto debido a que al tomar la foto, la calidad de la cámara influye mucho y un mismo color tiene diferentes tonalidades.

VII. RESULTADOS

TAMAÑO

Figura	Nº Pruebas	Nº Aciertos	% Error
Pequeño	11	8	27
Mediano	8	8	0
Grande	5	4	20

Tabla 1

Como se puede ver en la tabla 1, el algoritmo responde muy bien a la identificación de tamaños ya que se conoce la proporción de píxeles en cm^2 .

FORMA

Figura	Nº Pruebas	Nº Aciertos	# Errores	% Error
Triangulo	6	5	1	17
Cuadrado	4	3	1	25
Rectángulo	2	1	1	50
Rombo	5	5	0	0
Elipse	1	0	1	100
Circulo	5	4	1	20
Estrella	4	2	2	50

Tabla 2

Los resultados (tabla 2), muestran un gran error en el reconocimiento de elipse y rectángulos, esto porque su relación perímetro-área no esta bien definida y ésta es la entrada que más peso tiene en el sistema. Algo diferente pasa con los triángulos, círculos y cuadrados que tienen esa relación perfectamente definida.

COLOR

Figura	Nº Pruebas	Nº Aciertos	# Errores	% Error
Naranja	2	2	0	0
Amarillo	7	6	1	14
Rojo	2	2	0	0
Azul	4	2	2	50
Café	1	0	1	100
Verde Medio	1	0	0	100
Azul Claro	2	2	0	0
Verde oscuro	1	0	0	100
Purpura	2	1	1	50

Tabla 3

Los colores oscuros son difícilmente identificables por eso se ve un mayor error el colores como café y verde oscuro, esto ya que la luz distorsiona mayormente estos colores y con una menor proporción a los claros (tabla 3).

TEXTURA

Figura	Nº Pruebas	Nº Aciertos	% Error
Tela	4	2	50
Papel	18	16	11.11
Icopor	7	3	57

Tabla 4

Aun cuando la tela tiene un rango más amplio para su reconocimiento, la variedad de ellas hace que el error sea grande (tabla 4).

VIII. ANALISIS DE RESULTADOS

Durante todas las pruebas hechas, se observa que el proyecto tiene problemas en el reconocimiento de rectángulos y elipses, ya que su relación perímetro-área no está bien definida, conllevando a que el programa obtenga errores.

Con el círculo, cuadrado, estrella y triangulo se tiene un mejor reconocimiento, ya que la relaciones están bien marcadas y se facilita el reconocimiento.

El color tiene un problema que depende en su mayor parte por la calidad de la cámara, ya que esta no es de una buena resolución, los colores están algo distorsionados y al procesar la imagen esto no se puede arreglar, es más, al digitalizar las muestras se reduce la calidad.

IX. CONCLUSIONES

- Se conocieron algunos métodos y técnicas para reconocer imágenes, que llevaron a utilizar mejor los recursos disponibles para obtener mejores resultados, siendo estos fácilmente comprensibles.
- Los resultados obtenidos fueron aceptables para este proyecto de acuerdo a los algoritmos y técnicas usadas, con mejores herramientas la cantidad de acierto será mayor.

X. BIBLIOGRAFIA

- [1] GONZÁLES JIMÉNES, Javier. Visión por computador.
- [2] K. S Fu, GONZÁLES R. C.; C. S. G. Lee. Robótica: control, detección, visión e inteligencia.
- [3] MARAVALL, Darío. Reconocimiento de formas y visión artificial.
- [4] KLIR, George; ST. CLAIR, Ute; YUAN Bo. Fuzzy set theory. Foundations and applications.
- [5] ROSS, Timothy J. Fuzzy logic with engineering applications.
- [6] [http://www.scribd.com/doc/23371/ Procesamiento de imagenes-con-Matlab](http://www.scribd.com/doc/23371/Procesamiento-de-imagenes-con-Matlab)
- [7] [http://iee.uach.cl/downloads/ingelectora2006/ 10_Ordenez.pdf](http://iee.uach.cl/downloads/ingelectora2006/10_Ordenez.pdf)
- [8] [http://www.mathworks.com/products/ sysid/](http://www.mathworks.com/products/sysid/)

Detección e identificación de rostros utilizando SVM y color

M. Duarte, M. Hidalgo and O. Pinilla

Abstract— En este trabajo se presenta una propuesta para la detección e identificación de rostros aplicado en imágenes donde se encuentre un grupo de personas. De manera general, el sistema se compone de 2 etapas principales: una etapa de preprocesamiento y detección y una etapa de identificación. En la detección se utilizan técnicas de color y Convex Hull para determinar las posibles regiones que pueden contener rostros y, algoritmos de Adaboost en conjunto con Clasificadores Sencillos de Haar para validar dichas regiones. Para la identificación se aplican técnicas de Eigenfaces y Fisherfaces con el fin de extraer las características con las que se entrenan las Máquinas de Vectores de Soporte. (CATEGORÍA 2)

Palabras claves—Detección de Rostros, Reconocimiento de Rostros, Espacio de color HSV, SVM, Características de Haar, Adaboost, Convex Hull, Eigenfaces y Fisherfaces.

I. INTRODUCCION

Es común a todo ser humano, como primer acercamiento a un posible reconocimiento, identificar el rostro de un individuo. Siendo este un proceso diario y “sencillo” para el cerebro humano que permite consolidar criterios de diferenciación entre una y otra persona, en el ambiente computacional resulta ser un proceso altamente complejo que ha requerido de extensos estudios por parte de diversos grupos de investigación.

El reconocimiento de rostros se ha trabajado desde 1966, sin embargo, fue hasta 1991 cuando se dio un primer paso sobre el tema, pues el documento de Matthew Turk y Alex Pentland [12] sirvió como base para la mayoría de artículos que se han escrito hasta el momento. Desde entonces, el interés por buscar soluciones cada vez más eficientes a la idea de detectar e identificar rostros contenidos en una imagen, incluso en medios variantes y bajo circunstancias poco controladas (intensidad, contraste, posición del rostro, ángulo de rotación, uso de accesorios, entre otros), motiva a muchos investigadores a mejorar técnicas ya existentes ó, ha implementar nuevas con el fin de obtener resultados cada vez más óptimos. Es por esto, que actualmente existe un grupo amplio y variado de técnicas que son útiles para resolver el complejo problema de detectar e identificar rostros, cada vez de una forma más eficiente, entre las que se resaltan: Análisis de Componentes Principales (PCA) [1], [25], Análisis de Propiedades Locales (LFA) [25], Análisis de Propiedades Geométricas Basado en Componentes de Posición de Rostros (GFA) y diversos métodos de clasificación como Redes Neuronales y Máquinas de Vectores de Soporte (SVM).

La detección e identificación del rostro humano se han convertido en un recurso alternativo para dar solución a gran variedad de aplicaciones adicionales a los fines investigativos,

como por ejemplo: control de acceso, animación, manejo y clasificación de bases de datos, pasaportes biométricos y voz visual.

La propuesta que se presenta a continuación, tiene como enfoque principal, implementar tecnologías de software que no solo faciliten el proceso de detección e identificación de los rostros de las personas que se encuentren en una imagen, sino que adicionalmente consigan la mayor independencia posible del entorno en el que se adquiera la imagen. Para cumplir con tal objetivo, se utiliza una combinación de técnicas, entre las que se destacan: detección por color [1-3], detección de bordes, Convex Hull [12], Clasificador Sencillo de Haar, [13], [16], Adaboost [13-16], Eigenfaces [9-10], Fisherfaces [10], [17] y Máquinas de Vectores de Soporte [4-8].

En los párrafos que se muestran a continuación, se encuentra una descripción detallada tanto de las técnicas utilizadas (Sección II) como del procedimiento (Sección III), adicionalmente, se presenta un paralelo entre la propuesta que se planteó al iniciar el trabajo y los cambios que se llevaron a cabo en el transcurso del desarrollo del proyecto (Sección IV). Al finalizar el documento, se dan a conocer las conclusiones del trabajo expuesto.

II. MARCO TEÓRICO

A. EXTRACCIÓN DE CARACTERÍSTICAS Y CLASIFICADOR SENCILLO HAAR.

Los Filtros de Haar se usan para extraer, de forma eficiente y rápida, características de una imagen y representarlas en los $\{R\}$. Por lo que será posible establecer un umbral de decisión y así, clasificar las muestras entre válidas y no válidas (-1 y 1). Para extraer las características se divide la ventana en dos, tres o cuatro rectángulos (ver Figura 1) y se calcula la diferencia entre la suma de los píxeles que se encuentran en el rectángulo blanco y negro. Dependiendo de la forma escogida se resaltarán características de la imagen que pueden ser importantes para la detección, como bordes, textura, forma, etc.

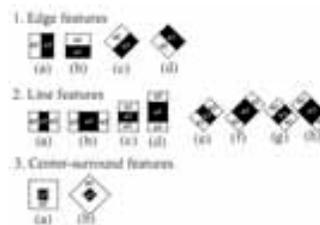


Figura 1. División en rectángulos de la imagen

La representación matemática para el cálculo de las características Haar se define:

$$f(x) = W_0 \times \text{sum}_{\text{white rec tan gle}}(\text{pixel gray level}) - W_1 \times \text{sum}_{\text{black rec tan gle}}(\text{pixel gray level}) \quad (1)$$

En donde W_1 y W_0 son constantes que se usan para compensar el área de los rectángulos, que en general se pueden establecer de la siguiente manera:

$$W_0 = 1 \text{ y } W_1 = \frac{a_0}{a_1} \quad (2)$$

La sumatoria de píxeles se representa por una integral de imagen:

$$P(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (3)$$



Figura 2. Representación de P(x,y)

Y por último, el clasificador sencillo (tipo stump) se resume en la siguiente ecuación:

$$h_j(x) = \begin{cases} 1 & \text{si } p_j f_j(x) < p_j \theta_j \\ 0 & \text{para otros} \end{cases} \quad (4)$$

B. ALGORITMOS DE BOOSTING Y ADABOOST

Un algoritmo de Boosting toma reglas de baja complejidad para luego formar una regla compleja. Es decir, toma clasificadores débiles cuyo error sea al menos ligeramente menor que escogerlo aleatoriamente entre clases y los fusiona como una combinación lineal donde la solución se expresa como sigue:

$$f(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right) \quad (5)$$

Para poder fusionar estos clasificadores en un solo criterio, se deben resolver los siguientes dilemas: ¿Cuáles de las reglas sencillas son las más influyentes para la decisión? (Cálculo de los coeficientes) Y, ¿cómo se pueden combinar todas las reglas de la forma más precisa y apropiada?

Para resolver lo anterior es necesario revisar el desarrollo teórico propuesto por Kearns y Valiant [11] y el algoritmo mostrado en la Figura 3, que encuentra la solución después de un intenso cálculo repetitivo.

C. FISHERFACES

El algoritmo de Fisherfaces se basa en el reconocimiento de patrones para la clasificación taxonómica planteada por Robert Fisher en 1936. Éste método es una derivación del Discriminante Lineal de Fisher (Fisher Linear Discriminant FLD).

Al aplicar el discriminante de Fisher se busca encontrar la matriz de transformación similar que exprese los datos de entrada (PCA II) en un espacio N-dimensional que intentará minimizar la varianza intra-clase y maximizar la varianza

entre-clase, utilizando la función de costo de Rayleigh.

Given: $\{x_1, y_1, \dots, x_m, y_m\}$ where $x_i \in X, y_i \in Y = \{-1, +1\}$
Initialize $D_1(i) = 1/n$.
For $t = 1, \dots, T$:

- Train weak learner using distribution D_t .
- Get weak hypothesis $h_t : X \rightarrow \{-1, +1\}$ with error $\epsilon_t = \text{Pr}_{D_t} [h_t(x_i) \neq y_i]$.

- Choose $\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right)$.
- Update:

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{if } h_t(x_i) = y_i \\ e^{+\alpha_t} & \text{if } h_t(x_i) \neq y_i \end{cases} \\ = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

where Z_t is a normalization factor (chosen so that D_{t+1} will be a distribution).

Output the final hypothesis:

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right)$$

Figura 3. Algoritmo de Adaboost

$$W_{OPT} = \arg \max \frac{|W^T S_B W|}{|W^T S_W W|} \quad (6)$$

Dado que la solución del problema de Rayleigh se limita a distribuciones convexas, la variación de las condiciones de iluminación o gestos que se quieren controlar deben tener una distribución lineal. Para calcular los Fisherfaces se definen las siguientes ecuaciones y matrices de dispersión:

Dispersión entre-clases:

$$S_B = \sum_{i=1}^C N_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (7)$$

Dispersión intra-clase:

$$S_W = \sum_{i=1}^C \sum_{x_k \in X_i} N_i (x_k - \mu_i)(x_k - \mu_i)^T \quad (8)$$

Donde μ_i es la media de la clase, Y_i es la clase, N_i es el número de muestras en la clase Y_i y $\{x_1, x_2, \dots, x_N\}$ es el conjunto de imágenes de muestra con su respectiva representación en un espacio de imágenes N-dimensional, donde cada imagen pertenece a una de las C clases $\{Y_1, Y_2, \dots, Y_C\}$.

La solución del planteamiento de Rayleigh se obtiene al calcular los valores y vectores propios de la matriz $S_W^{-1} S_B$. Sin embargo, puede presentarse el caso que S_W sea singular, por lo que S_W y S_B deben diagonalizarse con ayuda de la matriz ortonormal W_{PCA} , obtenida para el cálculo de los eigenfaces. Las ecuaciones que describen este procedimiento son:

$$W_{OPT}^T = W_{FLD}^T W_{PCA}^T = [w_1 w_2 \dots w_m]^T \quad (9)$$

Con,

$$W_{FLD} = \arg \max \frac{|W^T W_{PCA}^T S_B W_{PCA} W|}{|W^T W_{PCA}^T S_W W_{PCA} W|} \quad (10)$$

Se debe aclarar, que se podrán obtener los $c-1$ vectores propios generalizados cuyos valores propios sean no nulos y de mayor energía.

D. SVM (SUPPORT VECTOR MACHINE)

Las Máquinas de Vectores de Soporte son métodos de clasificación binaria que encuentran la superficie lineal óptima de decisión, la cual se obtiene al calcular los vectores de soporte, que son los pesos de cada uno de los elementos desde la base de entrenamiento a la superficie mencionada.

Para el entrenamiento de las máquinas de vectores de soporte se tendrán un conjunto de muestras:

$$S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\} \quad (11)$$

donde $\mathbf{x}_i \in \mathbf{R}^N$ y $y_i \in \{-1, 1\}$

Si el conjunto es linealmente separable, se puede obtener un hiperplano óptimo de separación, OSH (Optimal Separating Hyperplane), que maximiza la magnitud de los vectores de soporte. Este plano se expresa:

$$f(\mathbf{x}) = \sum_{i=1}^N \alpha_i y_i (\mathbf{x}_i^T \mathbf{x}) + b \quad (12)$$

En la Figura 4 observa que el algoritmo escoge el mayor margen posible usando los vectores de soporte.

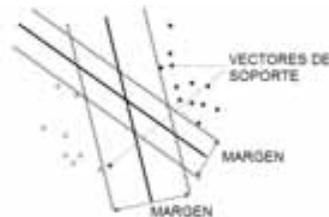


Figura 4. Ejemplo de maximización del margen

Para el caso en que el conjunto no pueda ser linealmente separable se realiza un mapeo de los vectores de entrada a un espacio dimensional mayor (Se reemplaza el producto punto por estructuras Kernel). Este nuevo espacio es llamado *espacio de características* y es allí donde el algoritmo anterior será aplicado, con la gran diferencia que el hiperplano de decisión será formado en el espacio de características.

Las estructuras Kernel escogidas deben cumplir con el teorema de Mercer. Entre algunas Estructuras de Kernel (productos punto) se encuentran:

- Kernel Polinomial: $(\mathbf{x}_i^T \mathbf{x} + 1)$ (13)

- Kernel Gaussiano: $\exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x}_j - \mathbf{x}\|^2\right)$ (14)

- Kernel Tangente Hiperbólico: $\tanh(\alpha \mathbf{x}_j^T \mathbf{x} + \beta)$ (15)

Por lo tanto, la función que define la superficie de decisión es:

$$f(\mathbf{x}) = \sum_{i=1}^N \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b \quad (16)$$

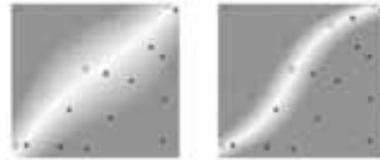


Figura 5. Comparación SVM Lineal y No lineal.

Las máquinas mencionadas anteriormente encuentran su aplicación únicamente para dos clases, por lo que se debe extender la metodología al caso multiclase, en el cual se resaltan las siguientes alternativas:

- OAA (One-against-all): Esta metodología sugiere entrenar C clasificadores que obtengan la región de separación óptima entre cada clase y el resto. Esto causa que ciertas muestras en la fase de prueba no sean clasificadas en ninguna de las clases o que la muestra pertenezca a varias clases.
- OAA continuo: Esta metodología elimina el caso en el que la muestra es clasificada como si no perteneciera a ninguna de las clases al cambiar el criterio de decisión de una función *signo* por una función continua.
- PAIRWISE: Se subdivide en dos categorías. La primera metodología emplea un esquema tipo torneo, donde la muestra pertenece a la clase que llega al final. La segunda sugiere una comparación de “una contra cada una de las clases”, al final la muestra pertenecerá a aquella clase que acumule mayor cantidad de votos, si existe el caso en que dos clasificadores obtengan igual cantidad de votos, se elegirá por medio del clasificador que diferencie a las dos clases.

III. PROCEDIMIENTO

En el diagrama de bloques de la Figura 6 se observa el procedimiento necesario para que al ingresar una imagen al sistema, éste detecte la totalidad de las personas presentes en la imagen e indique la posición y el nombre del rostro de quienes se encuentran en ella, siempre y cuando pertenezcan a la base de datos; si ocurre lo contrario, el sistema informará su posición y lo etiquetará como desconocido.

Tal como se observa en la Figura 6, el diagrama de bloques consta en términos generales, principalmente de 2 etapas: Preprocesamiento y Detección e Identificación. Estas etapas se explicarán en las subsecciones siguientes.

Sin embargo, antes de seguir con la explicación de cada uno de los bloques, es de vital importancia hacer énfasis en primer lugar en la adquisición y agrupación de la base de datos.

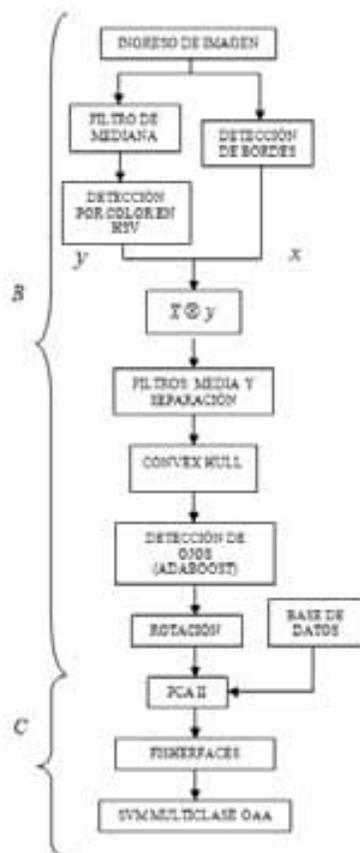


Figura 6. Diagrama de bloques sistema de detección e identificación

A. BASE DE DATOS: ADQUISICIÓN Y AGRUPACIÓN

Uno de los aspectos más importantes que se debe tener en cuenta antes de iniciar el procedimiento de detección e identificación, es la recolección de imágenes con las que se va a crear la base de datos completa, ya que a partir de ésta se van a realizar las pruebas y experimentos de evaluación del sistema. La base de datos del proyecto se divide en: base de datos individuales y base de datos grupales; ya que el sistema reconoce rostros en cualquiera de los dos casos. A su vez, dentro de los dos grupos, se dividen en imágenes de entrenamiento e imágenes de prueba.

Para la adquisición de imágenes se utilizaron 2 cámaras digitales de 2M píxeles de resolución. Adicionalmente, se establecieron especificaciones de adquisición, entre las que se resaltan las que se mencionan a continuación y las que se muestran en la Figura 7:

- Fotos con diferentes condiciones de iluminación y en ambientes parcialmente controlados, teniendo cuidado en que las imágenes adquiridas, en la región de interés (alrededor del rostro), no deben contener fondos cuyos píxeles pertenezcan al intervalo de color de la piel.
- La zona de los ojos debe estar completamente visible, sin importar si estos están total o parcialmente abiertos.

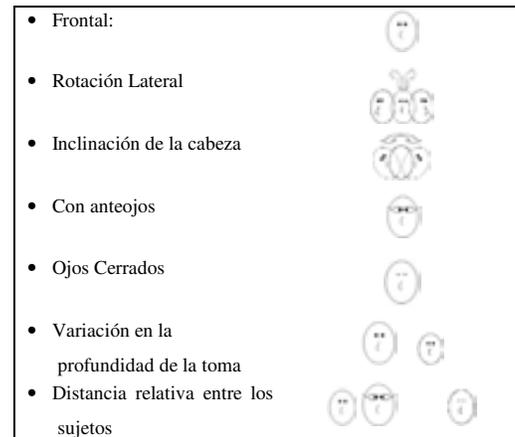


Figura 7. Condiciones de adquisición base de datos individuales y grupales

B. PREPROCESAMIENTO Y DETECCIÓN

Con el objetivo de reducir el ruido en la imagen que ingresa al sistema, el cual se genera debido a las variaciones en la intensidad y a los diferentes problemas que pueden presentarse en la cámara al capturar la imagen, es necesario aplicar un filtro de mediana de ventana cuadrada de 3x3. La ventana se escogió cuadrada porque esta reduce el ruido sin afectar mayormente los bordes y su tamaño de 3 x 3, evita que al ser muy grande llegue a distorsionar la imagen.

Para la detección por color se necesitó tomar algunas muestras de piel y agruparlas en el espacio HSV, con el fin de encontrar el intervalo de color de aceptación de diversos tipos y tonos de piel (nube de color). Entonces, cuando se ingrese una imagen nueva al sistema, se filtrará de acuerdo al intervalo en donde los píxeles válidos serán aquellos que caigan dentro de la nube de color.

Debido a que algunos fondos pueden tener un color similar al de la piel y por lo tanto ser erróneamente aceptados, se emplea la detección de bordes para disminuir el efecto de este inconveniente, ya que con ésta técnica es posible detectar un cambio brusco en los límites del rostro.

Como es necesario mantener las características que se obtienen al aplicar los dos procedimientos (bordes y color) se debe multiplicar píxel a píxel la imagen obtenida por HSV y la negada de los bordes. Al realizar la multiplicación mencionada, en la imagen resultante puede aparecer gran cantidad de ruido, por lo que se debe implementar un filtro de media con una ventana cuadrada de 3x3 y un filtro que elimine la gran cantidad de bifurcaciones que aparecen en los bordes de la imagen binarizada. La labor del filtro de eliminación es cambiar uno o máximo tres píxeles blancos (1's), que se encuentren entre dos negros (0's), por píxeles negros (0's). Para obtener resultados óptimos al aplicar el filtro de eliminación, se utiliza una búsqueda vertical, horizontal y diagonal (Figura 8).



Figura 8. Ejemplo de pre-procesamiento y detección por color.



Figura 10. Detección de las coordenadas de los ojos por Adaboost.

Desafortunadamente, con el procesamiento desarrollado hasta el momento se tiene un inconveniente con la región resultante, ya que ésta puede no tener forma de rostro y adicionalmente, no todas sus regiones internas se encuentran en el intervalo de color, lo que genera espacios vacíos dentro de la misma. Por esta razón, se implementa la técnica de Convex Hull [12] con la que se busca unir los puntos más externos teniendo en cuenta el ángulo que forman con los puntos vecinos. Al finalizar este procedimiento, se obtienen las posibles regiones que pueden contener rostros. Dichas regiones, solamente serán candidatas si cumplen la condición de que su perímetro supere un umbral establecido, de lo contrario, serán rechazadas (Figura 9).

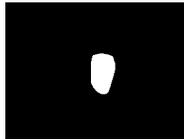


Figura 9. Región final Convex Hull.

El bloque de detección de ojos desempeña un papel vital en la detección, ya que una vez que se obtengan las posibles regiones para validar la presencia de un rostro en ellas, se evalúa la existencia de exactamente dos ojos en cada una. Además, dado que en la base de datos existen imágenes con rostros rotados, su detección sería más compleja, por lo que se busca reorientar los rostros a tal punto que su ángulo de inclinación sea cero. La rotación se realiza a partir de las coordenadas exactas de los ojos.

Para la detección de los ojos se emplea las características de Haar y de Lienhart en adición a una serie de clasificadores simples (Adaboost) [13-14], cuyo umbral de decisión se obtiene después del respectivo entrenamiento. Al detectar los ojos se encuentran las coordenadas exactas de su ubicación, a partir de las cuales se obtiene el ángulo y la distancia de separación entre ellos, lo que permite rotar y encontrar un tamaño exacto para la región de interés basado en las medidas del hombre de Vitrubio [19], respectivamente.

Los clasificadores se entrenaron con las siguientes características:

- 2360 ojos (muestras positivas)
- 3000 no ojos (muestras negativas)
- Ventana de 40 x 20
- Mínima tasa de efectividad: 0.995000

C. IDENTIFICACIÓN

Una vez se escalize y se rote el rostro, se continúa con el proceso de extracción de características al utilizar las técnicas de Fisherfaces y Eigenfaces, a partir de las cuales, se entrenan las Máquinas de Vectores de Soporte.

Con PCA II (Eigenfaces) se representa cada rostro en un espacio de características multidimensional, en donde el número de dimensiones es igual al número total de imágenes de la base de datos.

La ventaja que ofrece Fisherfaces sobre Eigenfaces es que el nuevo espacio de características reduce la varianza intra-clase y maximiza la varianza extra-clase.

Finalmente, para el proceso de identificación se emplean Máquinas de Vectores de Soporte [5], que son métodos de clasificación binaria que permiten maximizar la varianza extra-clase (encontrar la superficie lineal óptima de decisión).

Se debe tener presente que la base de datos está conformada por muchas clases (por lo menos mayor a dos), por lo cual, para realizar el análisis es necesario extender la metodología de SVM al caso multiclase, en donde se escogió trabajar con la técnica denominada OAA (One Against All). Esta permite obtener muestras que no pertenezcan a ninguna de las clases, lo que significa que esa persona no pertenece a la base de datos. Por otra parte, cuando una muestra es clasificada en más de uno de los clasificadores, entonces se escogerá la clase donde la distancia entre la muestra y el margen sea máxima.

Las Máquinas de Vectores de Soporte fueron entrenadas con los siguientes parámetros:

- Kernel Gaussiano
- Gamma: 0.0025
- Épsilon: 0.001

Donde el valor de épsilon es la condición de parada del algoritmo y fue escogido pequeño para obtener un error mínimo, el valor de gamma y el kernel, se escogieron según el desempeño y la documentación sugerida en [5] (Figura 11).

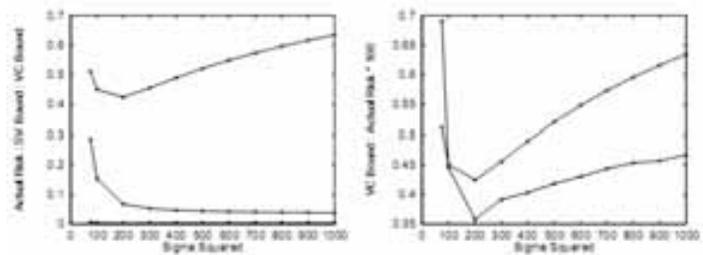


Figura 11. Resultados reportados sobre el kernel Gaussiano [5].

III. POSIBLES VARIACIONES EN LA METODOLOGÍA

Al plantear inicialmente el desarrollo del proyecto se realizó una propuesta que con el tiempo tuvo que reconsiderarse debido a que tanto el comportamiento del sistema, como los resultados no eran acordes a lo que se planteó. Las diferencias radican y se pueden agrupar en los siguientes puntos principalmente:

- El fondo inicial que se utilizó para adquirir las imágenes fue una pared de ladrillos, este lugar tuvo que cambiarse

debido a que para realizar las diferentes pruebas es necesario contar con más de un fondo sin control ni manipulación. Adicionalmente, y la explicación más importante, es que el fondo debe cumplir con la característica de no contener píxeles junto a la región de interés (rostro) que pertenezcan al intervalo o se encuentren demasiado cerca del intervalo del color de la piel.

- A pesar que el sistema funciona correctamente con imágenes de cualquier tamaño, se realizó un submuestreo para ajustarlo a la medida de 800x600, con el fin de mejorar el desempeño del sistema tanto a nivel de memoria como de velocidad de procesamiento.

- Se planteó inicialmente reconocer los ojos por técnicas que utilizan histograma o biometría, pero debido a que dentro de la base de datos existen imágenes con rostros rotados y/o ojos semicerrados, el uso de éstos procedimientos fueron tediosos por la necesidad de rotar en diferentes valores de ángulo el rostro hasta encontrar el valor aproximado de la inclinación. Se buscaron alternativas de evaluación que presenten mejor comportamiento y mayor robuztes antes las variaciones mencionadas. Debido a lo anterior, se decidió emplear Clasificadores de Haar entrenados con diferentes condiciones de apertura y rotación de ojos y rostro respectivamente.

- Para validar si una región es un rostro o si no lo es, inicialmente se planteó localizar tanto los ojos como la boca, pero al presentarse el cambio en la técnica de detección de ojos, la evaluación resulta ser más eficiente por lo que no es necesario encontrar la ubicación de la boca dentro de la región.

- Para detectar el color de la piel, se planteó el uso del espacio de color YCbCr, pero se determinó que no es el más adecuado ya que la representación de la piel y de ciertos fondos llegan a ser muy similares. Por esta razón, fue necesario buscar otro espacio de color en el que se agrupará mejor los píxeles de piel y por lo tanto, obtener un mayor porcentaje de diferenciación entre el rostro y los objetos pertenecientes a la imagen. El espacio de color HSV presenta una mejor representación del color ya que es una aproximación bastante cercana a la percepción humana.

IV. CONCLUSIONES

El espacio HSV tiene un rango de diferenciación más amplio que el encontrado en YCbCr, a pesar de presentar traslape de ciertos fondos con la representación de la piel, lo que permite realizar ciertos ajustes para compensar este efecto.

La detección del rostro basada únicamente en algoritmos de color no es suficiente para garantizar el obtener la totalidad del rostro, ya que existen regiones de piel que no se encuentran en el intervalo de color. Esto hace que aparezcan regiones con formas no similares a las de una cara y que adicionalmente pueden contener espacios vacíos. Para atenuar estos inconvenientes, se decidió emplear un algoritmo de Convex Hull.

Los resultados del proceso de entrenamiento de Adaboost para la detección de ojos son eficientes y óptimos, siempre y cuando la resolución de las muestras de entrenamiento sean visibles y con mínima distorsión. El inconveniente de utilizar

el método de Adaboost es su extenso tiempo de procesamiento debido al número de características seleccionadas para el mismo.

Actualmente, se continúa con las pruebas del sistema.

BIBLIOGRAFÍA

- [1] Berbar, M.A., Kelash, H.M. y Kandeel, A.A., "Faces and Facial Features Detection in Color Images", Geometric Modeling and Imaging--New Trends, 2006 05-06 July 2006 Page(s):209 - 214, Digital Object Identifier 10.1109/GMAL2006.18
- [2] Abdel-Mottaleb, M. Rein-Lien Hsu y Jain, A.K., "Face detection in color images", Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 24, Issue 5, May 2002 Page(s):696 - 706, Digital Object Identifier 10.1109/34.1000242
- [3] Yu-Ting Pai, Shang-Jang Ruan, Mon-Chau Shie y Yi-Chi Liu, "A Simple and Accurate Color Face Detection Algorithm in Complex Background", Multimedia and Expo, 2006 IEEE International Conference on 9-12 July 2006 Page(s):1545 - 1548, Digital Object Identifier 10.1109/ICME.2006.262838
- [4] Shavers, C., Li, R. y Leiby, G., "An SVM-based approach to face detection", System Theory, 2006 Proceeding of the Thirty-Eighth Southeastern Symposium on 5-7 March 2006 Page(s):362 - 366, Digital Object Identifier 10.1109/SSST.2006.1619082
- [5] Burges C., "A Tutorial on Support Vector Machines for Pattern Recognition", Bell Laboratories, Lucent Technologies 1998. Page(s):1 - 43
- [6] Phillips P., "Support Vector Machines Applied to Face Recognition", National Institute of Standards and Technology 1999. Page(s):1 - 7
- [7] Anh Tran Quang, Qian-Li Zhang y Xing Li, "Evolving support vector machine parameters", Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on Volume 1, 4-5 Nov. 2002 Page(s):548 - 551 vol.1 Digital Object Identifier 10.1109/ICMLC.2002.1176817
- [8] Ming - Hsuan Yang, "Face Recognition Using Kernel Methods", Honda Fundamental Research Labs. 2002 e(s):1 - 8
- [9] Turk M. y Pentland A., "Eigenfaces for Recognition", Journal of Cognitive Neuroscience Volume 3, 1. 1991 Page(s):71 - 86 vol.3
- [10] Belhumeur, P.N., Hespanha, J.P. y Kriegman, D.J., "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection", Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 19, Issue 7, July 1997 Page(s):711 - 720 Digital Object Identifier 10.1109/34.598228
- [11] Michael J. Kearns y Umesh V. Vazirani, "An introduction to computational learning theory", MIT press, Cambridge, Massachusetts, 1994.
- [12] Abellanas M., "Envolvente convexa", triangulación de Delaunay y diagrama de Voronoi: Tres estructuras geométricas en una, con muchas aplicaciones. Universidad Politécnica de Madrid. Disponible en: <http://divulgamat.ehu.es/weborriak/TestuakOnLine/06-07/PG-06-07-Abellanas.pdf>
- [13] Romero C., "Aplicación de algoritmos de boosting para la identificación de objetos móviles en entornos complejos". Tesis Maestría. Universidad de los Andes. Enero 2007.
- [14] Yoav Freund y Robert E. Schapire, "A Short Introduction to Boosting", AT&T Labs. Journal of Japanese Society for Artificial Intelligence, 14(5):771-780, September, 1999.
- [15] Yoav Freund y Robert E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting", AT&T Labs. Journal of computer and system sciences, 55, 119139 (1997). Article no. SS971504
- [16] Rainer Lienhart y Jochen Maydt, "An Extended Set of Haar-like Features for Rapid Object Detection", Intel Labs, Intel Corporation, Santa Clara.
- [17] Jian Yang, Avid Zhang y Jing-yu Yang, "Is ICA Significantly Better than PCA for Face Recognition?", Biometrics Centre, Department of Computing, Hong Kong Polytechnic University, Department of Computer Science, Nanjing University of Science and Technology. Proceedings of the Tenth IEEE International Conference on Computer Vision, 2005.
- [18] Duarte M., Hidalgo M., "Evaluación de la Técnica eigenfaces para Reconocimiento de Rostros", Pontificia Universidad Javeriana. Simposio en Tratamiento de Señales, Imágenes y Visión Artificial STSIVA 2007.
- [19] Arribas C., "Los misterios de Da Vinci, o las medidas del hombre" Disponible en: <http://www.extremaduraaldia.com/reportajes/los-misterios-de-da-vinci-o-las-medidas-del-hombre/27379.html>

VHDL open core for implementation of IEEE-754 single precision multiplication standard on FPGAs

Abraham H. Días - Pérez Popular University of the Cesar/Electronic Department, Sabanas Campus, Valledupar - Cesar, Colombia
Email: abraham.diaz@gmail.com

Abstract— In this document is presented an original VHDL (Very High Speed Integrated Circuits, Hardware description Language) Open Core to obtain the multiplication of numbers using the standard IEEE-754 single precision (32-bit) by means of the design software Quartus II® 7.2 Web Edition and the FPGA Cyclone II® EP2C5Q208C8 of the Altera® Company. Multiplications of Floating Points numbers are hard to implement on FPGAs due to the complexity of their algorithms, however many scientific operations as real-time filtering, convolution and Fast Fourier Transform require floating point operations with high levels of accuracy in their calculations. Recent advances in software design tools as Quartus II from Altera or ISE from Xilinx allows to the researcher focus on the algorithm problem more than the low level design (gates interconnections and addressing), giving the best optimization for analysis and synthesis, it is maximizing performance, minimizing logic usage or a balance between high performance with minimal logic usage.

Index Terms—VHDL, FPGAs, IEEE-754 floating point arithmetic, software design tools.

I. INTRODUCTION

Many scientific problems require floating point arithmetic with high level of accuracy in their calculations, between them, the Fast Fourier Transform and Cyclic Convolution algorithms for digital filters. Furthermore, these problems have a high degree of regularity that makes them candidates for hardware accelerated implementations [1]. In this work is presented a VHDL open-core for the realization of the multiplication of binary numbers in the IEEE 754 format, single precision (32 bits), with the objective to have a functions library to be used in the FPGAs implementation of algorithms for the cyclic convolution in one and two dimensions and also for the application of the one and two dimensional FFT in this kind of dispositive.

Manuscript accepted for publication August 21, 2008. This work was supported in part by the Electronic Department and the Spectroscopy and Laser Group (GEL) of the Physics Department of the Popular University of the Cesar.

A. H. Diaz-Perez is with the National Institute for advance of sciences COLCIENCIAS, cell phone: 316-269-1983; e-mail: abraham.diaz@gmail.com.

II. THEORETICAL FRAMEWORK

A. IP core

An IP (intellectual property) core is a block of logic or data that is used in making a field programmable gate array (FPGA) or application-specific integrated circuit (ASIC) for a product. As essential elements of design reuse, IP cores are part of the growing electronic design automation (EDA) industry trend towards repeated use of previously designed components. Ideally, an IP core should be entirely portable - that is, able to easily be inserted into any vendor technology or design methodology.

Universal Asynchronous Receiver/Transmitter (UARTs), central processing units (CPUs), Ethernet controllers, and PCI interfaces are all examples of IP cores.

IP cores fall into one of three categories: *hard cores*, *firm cores*, or *soft cores*. Hard cores are physical manifestations of the IP design. These are best for plug-and-play applications, and are less portable and flexible than the other two types of cores. Like the hard cores, firm (sometimes called *semi-hard*) cores also carry placement data but are configurable to various applications. The most flexible of the three, soft cores exist either as a *netlist* (a list of the logic gates and associated interconnections making up an integrated circuit) or hardware description language (HDL) code [2].

B. Floating Point Format

As mentioned above, the IEEE standard for Binary Floating Point Arithmetic (ANSI/IEEE Std 754-1985) [3] will be used throughout the present work. The single precision format is shown in Figure 1. The numbers in this format are composed of the following three fields [4]:

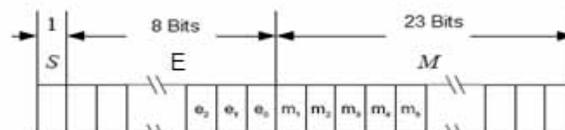


Figure 1. IEEE754 floating point single precision representation. Image Credit IEEE Standards Board.

Field	Size/bits	Meaning
Sign Bit	1	0: Indicates a Positive number, 1 Indicates a Negative number.
Exponent	8	Unsigned biased representation where the binary value is the sum of the original signed exponent value and the bias number. Zero is represented by a binary value of 0.
Mantissa	23	Unsigned fixed point fraction, $1.0 \leq \text{mantissa} < 2.0$. The MSB is always 1, and is omitted, so only $mw-1$ bits are stored.

Table 1 – IEEE754 Single Precision Floating point definitions

C. Floating Point Conversion

For a normalized number the IEEE-754 floating point value, f , is constructed from the fields as follows:

$$f = (-1)^s \times 1.M \times 2^{E-127}$$

Where M is the fractional mantissa number constructed from the binary value as follows:

$$1.M = 1 + \sum_{i=1}^{23} m_i 2^{-i}$$

(m_i is the mantissa bit relative to the maximum significant bit.) And where E is the exponent value constructed from the binary value as follows:

$$E = \sum_{i=0}^7 e_i 2^i$$

(e_i is the exponent bit relative to the lower significant bit.) And S is the value of the sign bit.

III ALGORITHM DEVELOPMENT

We start the algorithm development with the base that numbers to multiply be already in IEEE74 single precision format, in the following lines the VHDL code is showed and explained (the platform used for simulation was Quartus II V. 8.0sp1web edition, and the FPGA used is the):

CODE	DETAILS
library IEEE; use IEEE.STD_LOGIC_1164.ALL; use IEEE.STD_LOGIC_ARITH.ALL; use IEEE.STD_LOGIC_UNSIGNED.ALL;	Calls to de IEEE standard libraries. The VHDL words are in blue.
entity mult_ieee is Port (clk : in std_logic; Dato1 : in std_logic_vector(31 downto 0); Dato2 : in std_logic_vector(31 downto 0); man_s : out std_logic_vector(22	Setting ports, clock, inputs (Dato1, Dato2) and outputs (man_s; output mantissa, exp_s; output

<pre> downto 0); exp_s : out std_logic_vector(7 downto 0); mul_s : out std_logic; Sal : out std_logic_vector(31 downto 0)); end mult_ieee;</pre>	exponent, mul_s; sign of the output, Sal; Output)
<pre> architecture proceso of mult_ieee is signal mult_mant : std_logic_vector(47 downto 0); signal mant_sal : std_logic_vector(22 downto 0); signal exp_sal : std_logic_vector(7 downto 0); signal signo_sal : std_logic; signal conc_mantiza1 : std_logic_vector(23 downto 0); signal conc_mantiza2 : std_logic_vector(23 downto 0); signal exp_extra1 : std_logic_vector(7 downto 0); signal exp_extra2 : std_logic_vector(7 downto 0); signal signo1 : std_logic; signal signo2 : std_logic;</pre>	Here the mantissa of the multiplication is declared. See the size of the array 48 bits, then the mantissa of the output is reduced, to only 23 bits, the exponent of the output, and the sign of the output are also declared. Other auxiliary signals are used to facilitate the computation.
<pre> begin conc_mantiza1 <= '1'&Dato1(22 downto 0); conc_mantiza2 <= '1'&Dato2(22 downto 0); exp_extra1 <= Dato1(30 downto 23); exp_extra2 <= Dato2(30 downto 23); signo1 <= Dato1(31); signo2 <= Dato2(31);</pre>	The assignments of the signals.
<pre> process (clk) begin if (clk'event and clk = '1') then signo_sal <= signo1 xor signo2; mult_mant<=conc_mantiza1*conc_mant iza2; case mult_mant(47) is when '1' => mant_sal <= mult_mant(46 downto 24); exp_sal <= exp_extra1 +exp_extra2 - 126; Sal<= signo_sal&exp_sal&mant_sal; when '0' => mant_sal <= mult_mant(45 downto 23); exp_sal <= exp_extra1+exp_extra2 - 127;</pre>	The process with the clock begins. The sign of the resultin multiplication is assigned with the XOR command and the mantissas multiplication is done. If this multiplication fills the bit number 47 (worse case) then the exponent is compressed if not (bit 47 is zero), the

```

Sal<=signo_sal&exp_sal&mant_sal;
when others => null;
end case;
end if;
end process;
end proceso;
    
```

exponent remain uncompressed. The final stage is the concatenation of the sign and exponent and mantissa. An all the processes are ended.

Table 2– VHDL code for IEEE754 Single Precision Floating point multiplication.

The following step consists in start the compilation of the code, the following graph show the flow summary after the VHDL compilation:

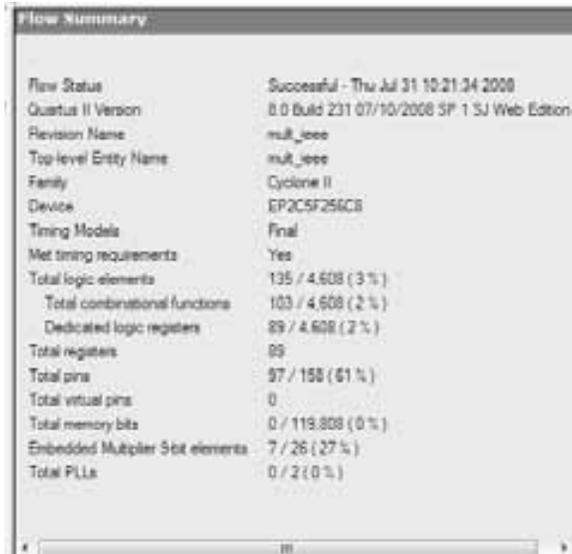


Figure 2. Flow summary of the multiplication algorithm

It is possible to appreciate that a high quantity of the available pin of the FPGA Cyclone II are used, 68% of them, which is not a desired feature, and the memory is unused.

The next step consist in validate the VHDL code by simulation, for space reason it is shown with only one example, corresponding to the multiplication between the real numbers 4 and -2. Theses numbers in IEEE754 single precision format are represented as:

01000000100000000000000000000000 for 4 and,

11000000000000000000000000000000 for -2.

Their multiplication (4x-2=-8) is performed in IEEE754 single precision as:

11000001000000000000000000000000.

When the platform Quartus II is used, it allow the creation of waveform files to simulate the algorithms in VHDL, for the case at hand we use the Cyclone II EP2C5F256C8, that permit a 260-MHz operation [5], it is approximately a delay of 7,7 ns for a complete clock cycle. The inputs and

outputs of the example at hand, handle with a clock of 100ns of period are shown in the following figure for 2 completes clock cycles:





Figure 3. Inputs and Output of the multiplication of real numbers 4 and -2, with a clock period of 100ns on the Cyclone II EP2C5F256C8 FPGA.

For space reason the windows corresponding to the bits 13 to 19 for Dato1, Dato2, and Sal were omitted (all are zero). In the 5th window of the figure above it is possible to appreciate the delay in the answer of the FPGA Cyclone II that was calculated before, approximately 7,7 ns in a complete cycle. It is important to note that the results of the simulation were consistent with the calculation made earlier in the IEEE754 single precision format.

IV. FUTURE WORK

The present work open the way to a more realistic implementation of multiplication algorithm on FPGA, the actual algorithm uses almost the 67 % of the available pins of the FPGA (input-output) only for one multiplication of numbers. In FFT or cyclic convolution problems the array of vectors are very large and a serial computation of every data will could be inappropriate. The following step is then to study the several open cores to FFT and try to implement a variation of the present algorithm for the FFT and Cyclic Convolution problems.

V. CONCLUSIONS

This work is our first approach to IEEE754 single precision multiplication FPGA development; we can not pretend to be competitive with the several cores available in the market as the Quixilica® Floating Point FPGA Cores. As we say before, it is the initial step to construct a good model for floating point multiplication to be applied on the Digital Signal Processing problems as FFT and Cyclic Convolution. The algorithm showed in the present work although is fully parallel has the disadvantage that could be impractical to apply for the FFT or Cyclic Convolution problems, because almost all the inputs of the FPGA are used by the two numbers to be multiplied, and then the in-place characteristic of the FFT will be impossible to reach in this kind of model. The models showed by the cores in the market present better performance when are applied to the FFT problem due that they realize a serial operation of delay and unpack the inputs before the multiplication step.

VI. REFERENCES

- [1] Loucas Louca, Todd A. Cook, William H. Johnson, *Implementation of IEEE Single Precision Floating Point Addition and Multiplication on FPGAs*. Rutgers University NJ, 1998
- [2] From Wikipedia, the free encyclopedia
- [3] IEEE Standards Board. *“IEEE Standard for Binary Floating-point Arithmetic”*. Technical Report ANSI/IEEE Std 754-1985, The Institute of Electrical and Electronics Engineers, New York, 1985.
- [4] QuineticQ *Datasheet Quixilica® Floating Point FPGA Cores*. December 2002.
- [5] Altera. *Cyclone II Device Family Data Sheet*. February 2008.



Abraham H. Diaz-Perez Was born in Barranquilla Colombia, 9 July 1972. Obtain the Bachelor of Electric Engineering from the Universidad Tecnológica de Bolivar at Cartagena-Colombia 1996, and the Master of Sciences degree from the University of Puerto Rico at Mayaguez-PR 2004.

Professor Diaz-Perez is a member of the International Institute of Systemics (IIS) and the Colombian institute for advance of sciences COLCIENCIAS.

He is currently working at the University Popular del Cesar as Professor-Researcher in the Group of Spectroscopy and Laser (GEL) at Valledupar Cesar. His principal areas of interest cover Digital Signal Processing, Image and Video, Fast Algorithms to Cyclic Convolution and Adaptive filters.

