

ARTÍCULO DE INVESTIGACIÓN / RESEARCH ARTICLE

Clasificador para la detección de personas desnudas en imágenes

Classifier for naked people
detection in images

Siovel Rodríguez Morales*

Universidad de Ciencias Informáticas –UCI (Cuba)

Walter Baluja García**

Instituto Superior Politécnico

José Antonio Echeverría –ISPJAE (Cuba)

* Universidad de Ciencias Informáticas (UCI), Facultad 1 (Cuba).
Ingeniero en Automática. Máster en Ciencias. siovelrm@gmail.com

**Instituto Superior Politécnico José Antonio Echeverría (ISPJAE),
Facultad de Ingeniería Eléctrica (Cuba). Ingeniero en Telecomunicaciones y Electrónica. Doctor en Ciencias. walter@tesla.cujae.edu.cu

Correspondencia: Walter Baluja García, Calle 10 n°. 11614, % 11 y Canal, Reparto Antonio Maceo, Municipio Cerro, La Habana (Cuba).
Tel.: + 537 2663206.

Resumen

Las imágenes constituyen un componente esencial de los contenidos digitales que se generan e intercambian a través de las redes de telecomunicaciones. Desde hace años la comunidad internacional ha trabajado en diversos sistemas de clasificación y filtrado de estos contenidos, dentro de los cuales la evaluación de las imágenes tiene un papel fundamental.

En el presente trabajo se exponen los principales resultados del diseño, la implementación y prueba de un módulo clasificador de imágenes, cuya tarea fundamental consiste en la detección de personas desnudas.

Este módulo, basado en el empleo de redes neuronales del tipo MLP (*MultiLayer Perceptron*), logra elevados índices de detección, lo cual, unido a su velocidad y fácil desarrollo, lo convierte en una seria alternativa para su inclusión en *softwares* destinado al filtrado de contenidos.

Palabras clave: Clasificación de imágenes, filtrado de contenidos, redes neuronales, seguridad de la información.

Abstract

Images constitute an essential component within the digital contents generated and interchanged through telecommunication networks. Since some years ago, the international community has been working in various filtering and contents classification systems. Within these systems, the evaluation of the images plays a crucial role. In this work, the main results are presented in relation to the design, implementation and testing of an image classification module, whose first task is to detect naked people. This module, MLP (*MultiLayer Perceptron*) neural networks based, achieves elevated detection levels, which along to its velocity and easy development, turns it into a serious alternative for its inclusion in content filtering software.

Key words: Content filtering, images classification, information security, neural networks.

Fecha de recepción: 11 de abril de 2012
Fecha de aceptación: 4 de agosto de 2012

1. INTRODUCCIÓN

Actualmente, Internet constituye la mayor fuente de información a nivel mundial. El contenido que almacena se considera equivalente a tres millones de veces el total de libros escritos en la historia [1]. Su tamaño se calcula en el orden de los hexabytes y su cantidad de sitios Web supera los 200 millones [2].

En Internet se alberga una enorme variedad de contenidos científicos, educativos, deportivos, culturales y más, los cuales han contribuido sustancialmente con el desarrollo de la humanidad. Sin embargo, también ofrece acceso a mucho contenido indeseado, tanto para el sector empresarial como el privado. En este sentido se destacan negativamente los sitios con contenidos pornográficos (12% del total de sitios Web [3]), pedófilos, xenofóbicos, terroristas, extremistas, entre otros). En ese contexto las técnicas y los sistemas de filtrado de contenidos se han convertido en importantes herramientas para discriminar el flujo de datos en cualquier entorno, entiéndase escuelas, ministerios, hogares y otros. Se aplican para lograr que el acceso a la información esté acorde con los intereses de las instituciones y las familias, mejorar la explotación del ancho de banda, evitar los perjuicios morales y culturales de la influencia de los contenidos dañinos, y proteger la integridad de personas y de organizaciones.

Dentro de las técnicas de filtrado de contenidos, cualquiera que sea su aplicación, la clasificación de imágenes resulta esencial. En múltiples trabajos se aborda este tipo de clasificación, y especialmente aquella que trata de identificar la piel, las personas desnudas o vestidas, los rostros, las partes del cuerpo y otros [4]-[8], sin que existan soluciones universales o evidentemente superiores hasta el momento.

Antecedentes de la clasificación de imágenes y de la detección de personas desnudas

La clasificación de contenidos, y especialmente la de imágenes, emplea sistemas de diversa naturaleza. La detección de la piel, los rostros y otros han sido tareas llevadas a cabo por especialistas que aplican, generalmente, técnicas de inteligencia artificial.

Tal es el caso de Ming-Ji, et al. [9], quienes proponen un método de detección de la piel basado en umbrales adaptables, el cual es capaz de encontrar valores óptimos en fondos complejos y diversos. Para lograr este objetivo se utilizan cuatro premisas basadas en el empleo de las propiedades de histogramas de probabilidades de aparición de la piel en la imagen (SPDH, *Skin Probability Distribution Histogram*). Este método tiene un comportamiento superior al del histograma con umbral fijo, empleado ampliamente hasta ese momento.

Por su parte, Jones y Rehg [10] describen la construcción de modelos de color para las clases *piel* y *no piel*, a partir de un conjunto de cerca de un billón de píxeles etiquetados. Utilizando estas clases se construyó un detector de píxeles con piel. Este sistema permitió obtener un clasificador para determinar si en la imagen aparecen personas desnudas con una eficacia superior al 85 %.

En [5] se describe un método para detectar piel en una imagen, usando redes neuronales del tipo SOM (*Self-Organizing Map* o mapas auto organizados), el cual alcanza una efectividad de un 94% en imágenes faciales (rostros). Además, se describe un procesador particular (AXEON) como base para la construcción de un detector de piel por *hardware*, y se muestran los beneficios de utilizar este sistema en un entorno de grandes dimensiones.

Adicionalmente, existen varios sistemas que tratan de determinar si un sitio es pornográfico, aplicando las técnicas de clasificación de imágenes. Uno de estos casos se muestra en [6], donde se presenta una aplicación que se basa en tres procesos para clasificar cada imagen:

- Análisis de la relación R (*Red*)/ G (*Green*) normalizada (a nivel de pixel).
- Análisis del histograma.
- Detección de la piel con el empleo de la Matriz de Composición Humana (HCM, por sus siglas en inglés).

Revisando los resultados y las aplicaciones de los trabajos comentados arriba y de otros, puede concluirse que en la actualidad no existe un sistema

o solución única, o reconocidamente superior, en el procesamiento de imágenes para la detección de personas desnudas. Los trabajos analizados aplican diferentes técnicas, generalmente de elevada complejidad, y algunos llegan a resultados que pueden catalogarse de buenos, pero ninguna alternativa resulta universal.

Por otra parte, aun cuando en la literatura aparecen diversas aplicaciones de las redes neuronales artificiales (RNA), en el caso particular de las MLP (*MultiLayer Perceptron*), caracterizadas por su flexibilidad y sencillez, se encuentran muy pocas referencias de su explotación como herramienta para la detección de piel o de personas desnudas en imágenes.

Por tales motivos, el presente trabajo explora el empleo de las redes neuronales MLP en la clasificación de imágenes para detectar la presencia de personas desnudas, con el objetivo de obtener un clasificador simple y efectivo que pueda ser empleado en sistemas de filtrado de contenidos (desarrollado totalmente en software).

2. METODOLOGÍA

Desde hace algunos años se desarrollan y despliegan una amplia variedad de sistemas de filtrado de contenidos. La mayor parte de ellos están dirigidos al control de la navegación Web, aunque existen algunos enfocados a otros servicios.

En Cuba se destaca FILPACON [11], el cual puede ser catalogado como un sistema de filtrado de contenidos contextual. En general, este tipo de sistemas utiliza mecanismos de clasificación automatizada de contenidos, apoyados en técnicas de Inteligencia Artificial.

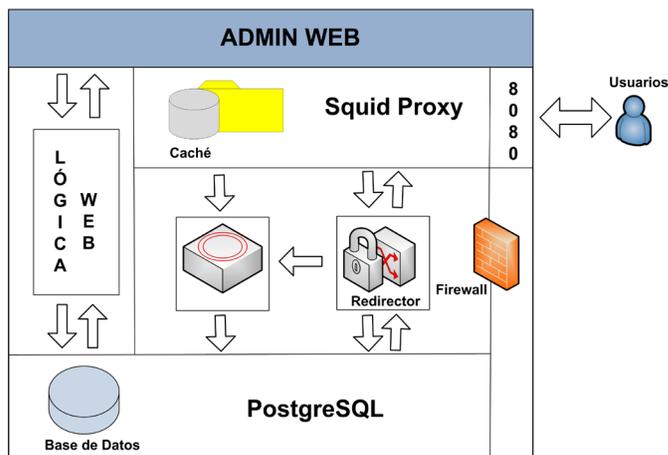


Figura 1. Arquitectura de FILPACON [11]

En la figura 1 se representa la arquitectura básica de FILPACON. En ella se pueden apreciar cuatro componentes esenciales: un Proxy Squid, una Base de Datos (BD) con millones de URLs preclasificadas, un programa Redirector y un Motor de Clasificación Dinámica (MCD). El MCD se encuentra en fase de desarrollo, y es donde aportan directamente los resultados del presente trabajo.

El mecanismo básico de operación del sistema es el siguiente [11]:

1. Los clientes solicitan una URL al Proxy Squid.
2. El Squid, a través del *software* redirector, hace una consulta a la BD.
3. En dependencia de esta consulta, donde se chequea quién es el usuario (política de navegación asociada) y a qué recurso pretende acceder (los contenidos pueden ser Ilícitos, Nocivos o Adecuados), se toma una decisión sobre si el recurso solicitado se le entrega al usuario o si, en su lugar, se devuelve una página de error.

El trabajo de análisis que realiza el MCD para la preclasificación de los sitios está enfocado en tres elementos fundamentales que componen las páginas Web: enlaces, textos e imágenes, como se muestra en la figura 2.



Figura 2. Análisis de la Web por el MCD

Cada uno de estos elementos es atendido por módulos diferentes del MCD, que aportan un conjunto de descriptores útiles para determinar la o las categorías a las que puede pertenecer el sitio. En particular, el análisis de las imágenes resulta extremadamente importante pues en muchas ocasiones son decisivas en la clasificación de los contenidos. Específicamente, la presencia de desnudez resulta vital en la identificación de pornografía o de contenidos similares.

Análisis de imágenes en el MCD

En el contexto de FILPACON, y como parte del MCD, se desarrolla el detector de personas desnudas en imágenes, cuyo componente clasificador se describe en este trabajo. Existen básicamente dos formas de asumir esta tarea, una es la comparación de la imagen con una base de datos de imágenes preclasificadas, y la otra es mediante el análisis inteligente de la imagen, la cual es la aplicada en FILPACON. En este último caso se extraen las características de la imagen, se analizan, y después se clasifican de acuerdo con parámetros como el color, la textura, las dimensiones, la forma y la cantidad de los objetos relevantes [12].

Una de las características más usadas en estos procesos de clasificación es el color de la piel, elemento predominante en las imágenes con personas desnudas. En esta línea Jones y Rehg proponen técnicas para la detección

del color de la piel mediante un modelo no paramétrico [10]. Esto permite la segmentación de la imagen para extraer una serie de datos que luego son utilizados en la clasificación [12].

Etapas de la clasificación de imágenes en el MCD

Aunque el esquema general del procesamiento de imágenes y de la visión artificial puede ser más complejo [13], en este trabajo se ha realizado una simplificación de ese proceso, eliminando fundamentalmente la etapa de Preprocesado, pues las mejoras que pudiera aportar son despreciables en comparación con el serio deterioro de la velocidad de clasificación que ocasiona su empleo. Además, se añade la etapa de Entrenamiento, debido a la utilización de las redes neuronales artificiales (MLP) en la Clasificación, lo cual permite volver a entrenar la red para buscar una mejor configuración de sus parámetros. Las cinco etapas utilizadas en el clasificador de imágenes propuesto en este trabajo se muestran en la siguiente figura y serán descritas a continuación.



Figura 3. Etapas definidas para la clasificación de imágenes en el MCD

Adquisición y filtrado

La adquisición de las imágenes es muy sencilla una vez descargadas las páginas de Internet. Sin embargo, muchas de las imágenes obtenidas no aportan información desde el punto de vista de la clasificación de un determinado sitio, por lo que se hace necesario filtrarlas. En el caso del clasificador que aquí se expone, se determinó que las imágenes con una resolución menor a 30 píxeles, ya sea por las filas o por las columnas, no sean contempladas.

Segmentación

La segmentación consiste en la división de la imagen en regiones de atributos similares, con el objetivo de distinguir algunos objetos del resto de los

que aparecen [14], [15]. En este caso, se propone realizar la segmentación a partir del color de la piel humana. A continuación se describe esta tarea.

Modelado del color de la piel

El objetivo del modelado del color de la piel es construir una regla de decisión que discrimine entre los píxeles de una imagen que corresponden al color de la piel y aquellos que no [16]. Para identificar el color de la piel en el espectro visible se depende, fundamentalmente, de la concentración de melanina y de hemoglobina en el cuerpo humano [17].

En particular, la idea clave de los métodos de modelado no paramétricos consiste en estimar la distribución del color de la piel por medio del entrenamiento de datos, sin derivar un modelo explícito. Generalmente, como resultado de la aplicación de estos métodos, se crea un mapa de probabilidad de presencia de piel que consiste en la asignación de una probabilidad a cada punto de los distintos espacios de color.

Según Jones y Rehg, para la construcción de los dos modelos (*piel* y *no piel*) se crean dos histogramas en 3D (tres dimensiones), donde los espacios de color se cuantifican con un número determinado de intervalos y cada intervalo almacena el número de veces que un color aparece en las imágenes de entrenamiento. Luego se puede calcular la probabilidad que un determinado color tenga en las clases *piel* y *no piel*, a través de las siguientes expresiones [10]:

$$P(\text{rgb} / \text{piel}) = \frac{S[\text{rgb}]}{T_s} \quad (1)$$

$$P(\text{rgb} / \text{no piel}) = \frac{N[\text{rgb}]}{T_n} \quad (2)$$

Donde $S[\text{rgb}]$ es la cantidad de píxeles obtenida del histograma referente a la piel y $N[\text{rgb}]$ es el equivalente referente a la ausencia de piel. T_s y T_n son la sumatoria de todos los valores en cada uno de los histogramas.

Luego se aplica un clasificador bayesiano [10]:

$$P(\text{piel} / \text{rgb}) = \frac{P(\text{rgb} / \text{piel})P(\text{piel})}{P(\text{rgb} / \text{piel})P(\text{piel}) + P(\text{rgb} / \text{npiel})P(\text{npiel})} \quad (3)$$

Donde $P(\text{piel})$ y $P(\text{npiel})$ son las probabilidades *a priori* de las clases *piel* y *no piel* respectivamente.

De estas tareas se obtiene un mapa de probabilidades, que no es más que una imagen en escala de grises, en la que cada píxel posee un valor entre 0 y 1, que indica su grado de pertenencia a la clase *piel*.

La fortaleza de este método radica en que se pueden tener imágenes con personas desnudas de diferente color de piel y sin embargo, al aplicarles el método, el resultado serán mapas de probabilidades con bastante similitud.

Selección del umbral para la conversión binaria

Luego de haber aplicado el método de Jones y Rehg a la imagen, y de haber obtenido un mapa de probabilidades (imagen en escala de grises), el siguiente paso es convertir esa imagen en binaria. Para ello se necesita un valor umbral a partir del cual todo valor menor, será etiquetado como “cero” (negro, no piel), y todo lo que sea mayor o igual será etiquetado como “uno” (blanco, piel). Con este objetivo se ensayaron, con cientos de imágenes, los métodos de Selección Iterativa [18], Otsu [19] e Inspección Visual (se seleccionan cientos de umbrales fijos considerados óptimos de forma visual y luego se aplica una media aritmética).

Extracción de características

En esta etapa se extrae, de cada imagen, un grupo de características que hacen posible su clasificación. Para este trabajo fueron seleccionadas las siguientes:

- Total de píxeles (filas).
- Total de píxeles (columnas).

- Cantidad de píxeles detectados con color de piel.
- Cantidad de regiones con color de piel.
- Cantidad de píxeles detectados con color de piel en la región más grande.
- Perímetro de la región más grande.
- Compacidad de la región más grande.

Este vector de características se utiliza como entrada a una red neuronal artificial (7 neuronas) del tipo MLP.

Entrenamiento

En todo sistema de reconocimiento de patrones es muy importante el proceso de entrenamiento y el algoritmo o técnica que se escoja para llevarlo a cabo. Básicamente, lo que se hace es seleccionar un grupo de muestras del dominio con el que se trabaja y separarlos en clases, donde cada clase engloba un conjunto de entidades que tienen características similares. De esta forma, se pueden tener tantas clases como clasificaciones se deseen emitir. En este caso solo se emplean dos clases: imágenes con personas desnudas y el resto de las imágenes.

Para seleccionar la cantidad de muestras para el entrenamiento de la red neuronal se cumplió con un requisito simple como $M > N/E$, donde M es la cantidad de muestras de entrenamiento, N la cantidad de nodos ocultos y E el máximo error permitido. Las pruebas de entrenamiento para conformar la red neuronal definitiva comenzaron con $N=4$ y $E=0.01$, por lo que debía cumplirse $M > 400$, para lo cual se seleccionaron inicialmente 420 muestras, 210 para cada una de las clases.

En esta etapa del trabajo se utilizó MATLAB para seleccionar las características de la red MLP y el algoritmo de entrenamiento a emplear. Se exportó el vector de características de la aplicación original y se emularon todas las condiciones del problema a resolver. Entre otros, se emplearon los métodos de aprendizaje Retropropagación del Error (*SteepestDescent* en LTI-Lib) y Gradiente Conjugado (*ConjugateGradient* en LTI-Lib).

La utilización del MATLAB, como paso previo, permite mostrar gráficamente el progreso del entrenamiento, y a partir de ahí desechar algunos

algoritmos y comprobar las configuraciones que pudieran ser las más adecuadas.

La definición de la red neuronal para el algoritmo clásico de Retropropagación del Error, se muestra en la siguiente expresión:

```
net = newff(minmax(P),[N1 NS],{' tansig ' ' tansig '},'traingdm');
```

donde:

- P*: matriz con los vectores de características de las imágenes.
- minmax*: función que devuelve el menor y mayor valor para cada una de las filas de la matriz *P*.
- N1*: cantidad de neuronas en la capa oculta. Cantidad variable a comprobar.
- NS*: cantidad de neuronas en la capa de salida. Dos en este caso (respuesta afirmativa y respuesta negativa).
- tansig*: función de activación en la capa oculta y en la de salida. En este caso, sigmoideal.
- traingdm* algoritmo de entrenamiento clásico de Retropropagación del Error con momento.

Además, se definieron los siguientes parámetros que condicionaron el entrenamiento:

- net.trainParam.epochs* = 800 máximo número de iteraciones en el entrenamiento.
- net.trainParam.goa* = 0.01 error final de entrenamiento.
- net.trainParam.lr* = 0.01 tasa de aprendizaje.
- net.trainParam.mc* = 0.9 momento.

Finalmente, se procede al entrenamiento de la red, donde *T* es la salida esperada:

```
net = train(net,P,T);
```

Por otra parte, para la utilización del método del Gradiente Conjugado, lo único que varía en la expresión es el algoritmo de entrenamiento, en este caso el 'traincgf':

```
net = newff(minmax(P),[N1 N2],{tansig' 'tansig'},'traincgf');
```

Los demás parámetros se mantienen igual, excepto que en este caso no se emplea el parámetro *net.trainParam.mc* (momento), propio del algoritmo clásico de Retropropagación.

Clasificación

La etapa de clasificación consiste en seleccionar un conjunto de imágenes de cualquier índole y comprobar si el clasificador trabaja correctamente, en dependencia de las clases que se han definido (imágenes con personas desnudas e imágenes sin personas desnudas). En esta etapa se determina la eficacia de clasificación.

Instrumentos

Para la implementación y prueba de la red neuronal se utilizó un conjunto de herramientas de *software* libre, bajo la Licencia Pública General de GNU (GNU), GPL (*General Public License*) y la Licencia Pública General Reducida de GNU (LGPL, por sus siglas en inglés). Adicionalmente, se empleó MATLAB para facilitar la realización de algunas pruebas con la función *graythresh*, la cual implementa el método de Otsu, así como para ayudar en la selección de los algoritmos y de la topología de la red neuronal.

Se trabajó sobre plataforma Debian GNU/Linux Etch 4.0 con el Entorno de Desarrollo Integrado Anjuta (<http://www.anjuta.org/>), el cual se caracteriza por ser bajo consumidor de recursos de cómputo.

Durante la programación del clasificador se empleó C++ y la biblioteca de clases LTI-Lib (*Lehrstuhl fuer Technische Informatik*, <http://ltilib.sourceforge.net/doc/homepage/index.shtml>), orientada a objetos, con algoritmos y estructuras de datos frecuentemente usados en el procesamiento de imágenes y en la visión por computadora.

3. RESULTADOS

A continuación se describirán los principales resultados obtenidos durante el diseño, la implementación y prueba del clasificador.

Etapas de segmentación

En la figura 4, se muestran algunas imágenes a las cuales se le aplicó el método de Jones y Rehg descrito anteriormente, y se obtuvo en cada caso una imagen en escala de grises (mapa de probabilidades).



Figura 4. Imágenes a las que se les aplicó el método de Jones y Rehg. Izquierda: imagen original, derecha: imagen en escala de grises

A lo largo de los experimentos de segmentación, realizados con un amplio conjunto de imágenes (más de 1000), de las cuales aproximadamente en la mitad aparecían personas desnudas, pudo comprobarse la exactitud del algoritmo (o al menos en su implementación con la librería utilizada). Solo se presentaron problemas con algunas imágenes de fondo amarillo-rojizo (como la que se muestra en la figura 5).

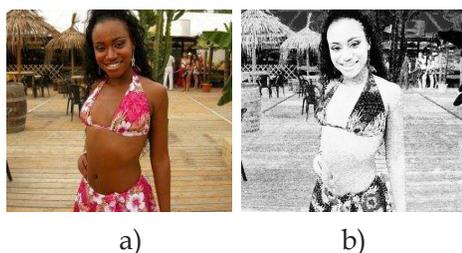


Figura 5. Imagen con fondo amarillo-rojizo que presenta problemas con el algoritmo Jones y Rehg. a) Imagen original b) Imagen en escala de grises

- Selección del umbral para la conversión binaria

Para determinar el algoritmo de selección del umbral, primero se probó el de Selección Iterativa y los resultados obtenidos fueron aceptables (figura 6c). No obstante, varios píxeles del fondo de las imágenes fueron incluidos como *piel* (color blanco) cuando realmente no lo son. En la biblioteca de desarrollo utilizada, LTI-Lib, este método recibe el nombre de *OptimalThresholding*.

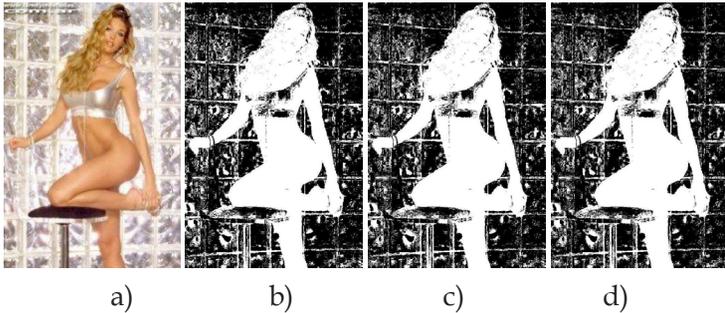


Figura 6. Proceso de segmentación. a) Imagen original b) Imagen en escala de grises c) Imagen binaria por el método de Selección Iterativa d) Imagen binaria por el método de Otsu

El método de Otsu fue implementado con MATLAB. En la figura 6d, se muestra un ejemplo de imagen binaria obtenida. Como se puede apreciar, los resultados obtenidos con este método son prácticamente iguales a los arrojados por el método anterior.

Para probar el método de Inspección Visual se seleccionaron 1000 imágenes, a las que de forma manual se les fue variando el valor de umbral y se fue seleccionando visualmente el valor óptimo para cada una de ellas. Algunos ejemplos se muestra en la figura 7.

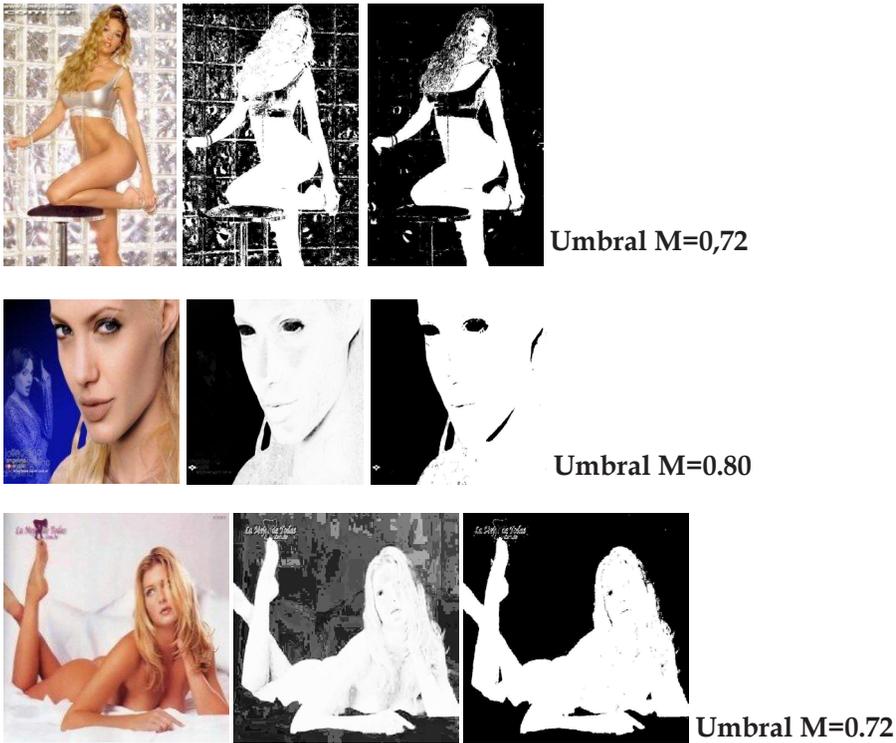


Figura 7. Proceso de segmentación. Izquierda: imagen original, centro: imagen en escala de grises, derecha: imagen binaria obtenida tras el método de Inspección Visual

Posteriormente, se calculó el valor medio de umbral para las 1000 imágenes: 0,81. En la figura 8 se muestran las mismas imágenes de la figura 7, esta vez segmentadas con el valor medio de umbral. Los resultados son buenos; la mayoría de los píxeles pertenecientes al fondo, que en los métodos anteriores eran incluidos como *piel*, son clasificados como *no piel*. Además, como el valor de umbral es fijo, no tiene que calcularse para cada imagen analizada, aspecto relevante para los sistemas que deben trabajar en línea. Este valor de umbral de 0,81 fue utilizado en el proceso de conversión a imágenes binarias en la primera versión del clasificador de imágenes del sistema FILPACON.



Figura 8. Imágenes binarias obtenidas con el umbral fijo (0,81) calculado por el método de Inspección Visual

Entrenamiento

Para el entrenamiento de la red neuronal MLP primero se ensayó el algoritmo clásico de Retropropagación del Error. Al probar con diferentes cantidades de neuronas en la capa oculta (de 4 a 20), se comprobó que el error no se acercaba a los valores deseados, siendo el mejor valor obtenido de 0,221526 (figura 9), logrado con 15 neuronas.

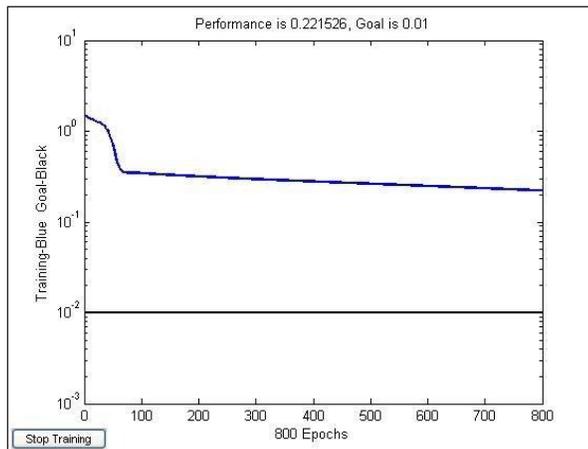


Figura 9. Mejor error de entrenamiento para el algoritmo clásico de Retropropagación del Error, topología de red 7-15-2

Cuando se procedió a realizar algunas pruebas de clasificación (con 100 imágenes), se obtuvo una eficacia de clasificación de un 42%, lo que condujo a desechar este algoritmo.

Posteriormente, se realizaron pruebas de entrenamiento con el algoritmo de Gradiente Conjugado. Los resultados obtenidos para las diferentes cantidades de neuronas en la capa oculta (de 4 a 20) fueron bastante buenos. El valor de error casi siempre disminuyó con tendencia exponencial y con valores inferiores a 0,08, correspondiendo a la configuración de 14 neuronas el mejor valor obtenido: 0,047011 (figura 10).

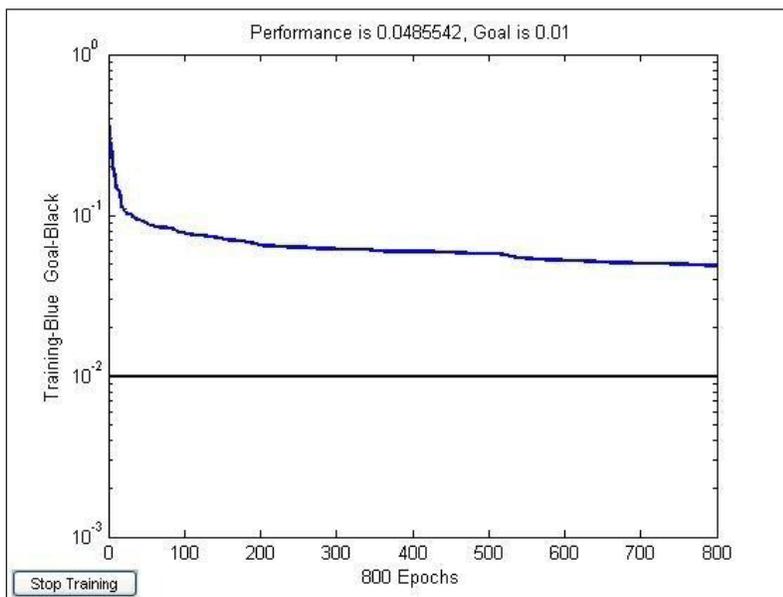


Figura 10. Mejor error de entrenamiento para el algoritmo de Gradiente Conjugado, topología de red 7-14-2

Una vez desechado el algoritmo clásico de Retropropagación del Error, se procede a realizar el entrenamiento real, implementando la red con la biblioteca LTI-Lib, y usando el algoritmo del Gradiente Conjugado. En la tabla 1 se muestran los errores de entrenamiento alcanzados para los distintos valores de neuronas en la capa oculta, así como el tiempo de entrenamiento correspondiente.

Tabla 1
Error y tiempo de entrenamiento para las diferentes cantidades
de neuronas en la capa oculta

Cantidad de Neuronas	Error	Tiempo (800 épocas)
4	0.146534	1m 35.334s
5	0.136398	1m 39.779s
6	0.179557	1m 40.426s
7	0.076511	1m 44.979s
8	0.076663	1m 46.869s
9	0.051415	1m 48.894s
10	0.089369	1m 50.408s
11	0.058511	1m 53.841s
12	0.048200	1m 55.084s
13	0.115951	1m 57.500s
14	0.047011	2m 2.156s
15	0.057927	2m 3.318s
16	0.058319	2m 4.534s
17	0.097419	2m 6.883s
18	0.067224	2m 8.275s
19	0.049027	2m 12.827s
20	0.058863	2m 13.434s

Tener un error de entrenamiento bajo, aunque generalmente constituye un buen indicador, no implica que necesariamente el sistema real clasificará mejor [20], por lo que se decidió que todas las topologías que arrojaran errores de entrenamiento menores que 0.06 fueran utilizadas en la etapa de Clasificación. En ese caso, se seleccionaron las redes con 9, 11, 12, 14, 15, 16, 19 y 20 neuronas en la capa oculta.

Clasificación

Una vez entrenada, la red neuronal se encuentra lista para clasificar. En un experimento inicial de clasificación se utilizaron 386 nuevas imágenes con personas desnudas y otras 287 nuevas imágenes sin personas desnudas, para un total de 673. Los resultados obtenidos en la clasificación se muestran en la tabla 2.

Tabla 2
Resultados de la clasificación empleando la RNA entrenada

Neuronas en la capa oculta	Imágenes bien clasificadas				% General	Tiempo de trabajo
	Con personas desnudas	%	Sin personas desnudas	%		
9	308	79,79	259	90,24	84,25	1m 57.021s
11	326	84,46	202	70,38	78,45	1m 56.518s
12	321	83,16	208	72,47	78,6	1m 57.581s
14	318	82,38	259	90,24	85,74	1m 57.716s
15	354	91,71	264	91,99	91,83	1m 57.587s
16	333	86,27	201	70,03	79,35	1m 56.875s
19	309	80,05	262	91,29	84,84	1m 57.584s
20	329	85,23	268	93,38	88,71	1m 56.575s

Como se puede apreciar en la tabla, los mejores resultados obtenidos fueron para la red de 15 neuronas en la capa oculta (topología 7-15-2), cuya eficacia de clasificación fue de un 91,83 %. No obstante, no se desecha el resto de las redes, pues pueden ser utilizadas en nuevos procesos de prueba y/o clasificación.

Este valor de eficacia también es bueno en comparación con los alcanzados en trabajos similares. En [21] la mejor eficacia lograda fue de un 84% y en [10] fue de un 85,8%. Dentro de la amplia bibliografía consultada, solo un trabajo precedente muestra mejores resultados en la detección de personas desnudas, se trata de Sibunruang, et al. [6], donde se alcanzó un 94,2% de eficacia con un método más complejo, no basado en redes MLP. En general, los resultados obtenidos son comparables con estudios muy recientes, en los que se aplican algoritmos más avanzados y complejos [22].

Por otra parte, el promedio de la velocidad de clasificación fue de 5,7 imágenes por segundo. Este es un índice muy bueno, si se toma en consideración que todo el proceso de entrenamiento y de prueba se desarrolló en computadoras personales.

4. CONCLUSIONES

En este artículo se presentaron los detalles del diseño, la implementación y la comprobación del funcionamiento de un clasificador de imágenes

sencillo, basado en redes neuronales del tipo MLP, que alcanza una eficacia del 91,83% en la detección de personas desnudas, considerada buena para la mayor parte de las aplicaciones.

En el diseño del clasificador se aplicó un grupo de etapas básicas, propias del procesamiento de imágenes. En particular debe destacarse la decisión de eliminar la etapa de Preprocesado (buscando mayor velocidad) y de incluir una etapa de Entrenamiento (para la red neuronal utilizada). Esta última permite agregar nuevas muestras, o quitar algunas que estén influyendo negativamente en los resultados de clasificación, y volver a entrenar la RNA.

Los resultados alcanzados permitieron incluir este clasificador en el MCD del sistema FILPACON, obteniéndose buenos resultados. Además, el mismo clasificador puede ser empleado en motores similares para la revisión de documentos, de mensajería electrónica y en muchas otras aplicaciones.

La eficacia de la clasificación y la velocidad de ejecución convierten este clasificador en una solución de bajo costo, técnica y económicamente, lo cual permite emplearlo en diversos entornos (como escuelas, ministerios y hogares) y en múltiples servicios (correo electrónico, navegación Web, transferencia y almacenamiento de documentos, entre otros).

REFERENCIAS

- [1] J. F. Gantz, C. Chute, A. Manfrediz, S. Minton, D. Reinsel, W. Schlichting y A. Toncheva, "The diverse and exploding digital universe", EMC-IDC, 2008. [En línea]. Disponible: <http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf>.
- [2] Netcraft, "May 2010 Web Server Survey", Netcraft, 2010. [En línea]. Disponible: http://news.netcraft.com/archives/2010/05/14/may_2010_web_server_survey.html.
- [3] J. Ropelato, "Internet pornography statistics", *Internet Filter Review*, 2006. [En línea]. Disponible: <http://internet-filter-review.toptenreviews.com/internet-pornography-statistics.html>.
- [4] A. Faizi, "Robust face detection using template matching algorithm", Tesis de maestría, Department of Electrical Engineering, University of Toronto, 2008. [En línea]. Disponible: https://tspace.library.utoronto.ca/bitstream/1807/17164/1/Faizi_Amir_200811_MASc_thesis.pdf.

- [5] D. Brown and I. Craw, "A SOM Based Approach to Skin Detection with Application in Real Time Systems", *12th British Machine Vision Conference*, Manchester, 2001. [En línea]. Disponible: http://www.comp.leeds.ac.uk/bmvc2008/proceedings/2001/papers/33/accepted_33.pdf.
- [6] C. Sibunruang, J. Polpinij, R. Camchong, A. Chotthanomet, and S. Puang-pronpitag, "A pornographic web patrol system based-on hierarchical image filtering techniques", *9th International Conference on Computer Science and Informatics*, Taiwán, 2006. [En línea]. Disponible: <http://www.atlantis-press.com/publications/aisr/jcis-06/index.html>.
- [7] N. Sprague and J. Luo, "Clothed people detection in still images", *16th International Conference on Pattern Recognition (ICPR)*, Canadá, 2002. [En línea]. Disponible: <http://max.cs.kzoo.edu/~nsprague/papers/pplfind.pdf>.
- [8] P. Yan and K. W. Bowyer, "Biometric recognition using 3D ear shape", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, n° 8, pp.1297-1308, 2007.
- [9] Z. Ming-Ji, W. Wei-Qiang, Z. Qing-Fang and G. Weng, "Skin-color detection based on adaptive thresholds", *Third International Conference on Image and Graphics. IEEE Computer Society Press*, Hong Kong, 2004.
- [10] M. Jones y J. Rehg, "Statistical color models with application to skin detection", *International Journal of Computer Vision*, vol. 46, n° 1, pp.81-96, 2002 [En línea]. Disponible: <http://www.cc.gatech.edu/~rehg/Papers/SkinDetect-IJCV.pdf>.
- [11] A. Guerrero, S. Rodríguez, J. Porven et al., "Sistema adaptativo de filtrado de contenidos", *VII Seminario Iberoamericano de Seguridad en Tecnologías de la Información, XII Convención Internacional Informática*, Cuba, 2007.
- [12] R. Ap-Apid, "An algorithm for nudity detection", *5th Philippine Computing Science Congress*, Filipinas, 2005. [En línea]. Disponible: <http://www.math.admu.edu.ph/~raf/pcsc05/proceedings/AI4.pdf>.
- [13] A. Florencia-Ysiquio, "Modelado de Sistemas de control de un robot manipulador basado en procesamiento digital de Imágenes", Tesis de Maestría, Departamento de Ingeniería en Sistemas Computacionales, Universidad de las Américas, Puebla, México, 2004. [En línea]. Disponible: http://catarina.udlap.mx/u_dl_a/tales/documentos/msp/florencia_y_an/.
- [14] T. Asano, D. Z. Chen, N. Katoh and T. Tokuyama, "Polynomial-time solutions to image segmentation", *7th Annual SIAM-ACM Conference on Discrete Algorithms*, Atlanta, 1996.
- [15] L. Shapiro and G. Stockman, *Computer Vision*, New Jersey: Prentice-Hall, 2001.
- [16] V. Vezhnevets, V. Sazonov and A. Andreeva, "A survey on pixel-based skin color detection techniques", *International Conference on Computer Graphics and Vision*, Rusia, 2003. [En línea]. Disponible: <http://graphics.cs.msu.ru/en/publications/text/gc2003vsa.pdf>.

- [17] C. Gutiérrez, M. Meneses, A. Gutiérrez, A. Vallarta, L. Andrade, A. Ávila, C. Montoya, R. Sánchez y P. Santana “Depilación con láser”, *Revista de la Asociación Mexicana Cirugía Plástica Estética y Reconstructiva*, vol. 15, n° 1, 2005. [En línea]. Disponible: <http://www.medigraphic.com/pdfs/cplast/cp-2005/cp051d.pdf>.
- [18] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, Pacific Grove: PWS Pub., 1999.
- [19] N. Otsu, “A threshold selection method from gray-level histograms”, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, n° 1, pp.62-66, 1979.
- [20] B. Martín del Brío y A. Sanz-Molina, *Redes Neuronales y Sistemas Difusos*, 2ª edición ampliada y revisada. Madrid: RA-MA, 2001. ISBN 84-7897-466-0.
- [21] J. C. Cortizo-Pérez y D. Expósito-Gil, “Detección de Imágenes Pornográficas mediante Redes Neuronales Artificiales”, AINetSolutions Technical Report, n° 02, 2004. [En línea]. Disponible: <http://www.esp.uem.es/jccortizo/RNAS.pdf>.
- [22] M. Rehenuma, A. Gias, M. Kamal, H. Muhammad, M. Ibrahim, A. Khan, A. Imran, S. Islamm, G. Rabbani, S. Mostafa, S. Islam, and Z. Begum, “Comparative study of statistical detection algorithms for sub-continental human images”. *Information Technology Journal*, vol. 4, n° 9, pp.811-817, 2010.