

تشخیص تصاویر دیجیتال حاوی برهنگی با استفاده از شبکه عصبی و ماشین بردار پشتیبان

سید مجید خراشادی زاده^۱، وحید آزادزاده^۲، علی محمد لطیف^{۳*}

۱- کارشناسی ارشد، ۳- استادیار دانشگاه یزد

(دریافت: ۹۴/۱۰/۱۷، پذیرش: ۹۵/۰۸/۱۰)

چکیده

در این مقاله رویکرد جدیدی برای تشخیص تصاویر با محتوای غیراخلاقی ارائه شده است. سیستم پیشنهادی شامل سه گام است، در گام اول تصویر به صورت پیکسل به پیکسل توسط یک شبکه عصبی به نواحی پوستی و غیرپوستی بخش بندی می شود. پس از این بخش مجموعه ای از ویژگی ها از جمله نسبت پیکسل های پوستی به کل پیکسل های تصویر استخراج می شوند. در پایان از ماشین بردار پشتیبان برای طبقه بندی تصویر به عنوان یک تصویر سالم یا غیراخلاقی بر حسب ویژگی های استخراج شده استفاده می شود. در این تحقیق از یک مجموعه شامل ۴۰۰ تصویر برای آموزش و تست سیستم استفاده شده است. نرخ بخش بندی تصویر به نواحی پوستی و غیر پوستی ۹۱/۸ درصد است و نرخ تشخیص تصاویر غیراخلاقی توسط روش پیشنهادی ۸۹/۹ درصد است. لازم به ذکر است نرخ مثبت صحیح و نرخ مثبت کاذب برای روش پیشنهادی به ترتیب ۹۲٪ و ۰/۱۲۵٪ به دست آمد.

واژه های کلیدی: طبقه بندی تصاویر، تشخیص برهنگی، شبکه عصبی پرسپترون چندلایه، ماشین بردار پشتیبان، تخمین چگالی

۱- مقدمه

یکی از راه حل های این مسأله، مسدود کردن کامل دسترسی به برخی از وب سایت ها بر مبنای یک لیست سیاه است که این کار می تواند با استفاده از تعدادی کلمات کلیدی انجام شود. اکثریت روش های رایج تشخیص محتوای غیراخلاقی، مبتنی بر متن هستند. به عبارت دیگر عمل فیلتر کردن با استفاده از محتوای متنی و عنوان سایت و هم چنین برچسب های تصاویر انجام می شود. این رویکرد ممکن است برای سایت های مختص به تصاویر غیراخلاقی کارآمد به نظر برسد؛ اما در یک شبکه اجتماعی که هم حاوی محتوای مفید و هم حاوی محتوای غیراخلاقی است، مسدود کردن کل سایت ناکارآمد است. اکثر روش های مبتنی بر متن برای پالایش محتوای وب دقت خوبی ندارند، چرا که این روش ها از تحلیل محتوایی تصاویر استفاده نمی کنند، در حالی که تصاویر یکی از مؤلفه های اصلی اینترنت هستند. بنابراین می توان با بررسی محتوای تصاویر، عملکرد روش های پالایش وب را بهبود داد.

این مقاله به صورت زیر سازماندهی شده است: در بخش دوم برخی از مقالات ارائه شده در این زمینه به اختصار معرفی شده اند. در بخش سوم سیستم پیشنهادی و اجزاء آن به طور مفصل تشریح می شود. بخش چهارم به ارزیابی و ارائه نتایج حاصل از پیاده سازی سیستم پیشنهادی اختصاص یافته است و در نهایت نتیجه گیری در بخش پنجم مطرح شده است.

بدون شک یکی از بزرگ ترین اختراعات بشر در قرن بیستم، اختراع رایانه بوده است. هم چنین ایجاد شبکه اینترنت در دهه ۱۹۹۰ میلادی یکی از تأثیرگذارترین اتفاقات در تاریخ استفاده از رایانه است. با وجود این که از عمر اینترنت و وب^۱ بیش از ۱۵ سال نمی گذرد، امروزه بسیاری از خدمات، به ویژه در کشورهای توسعه یافته، از طریق اینترنت ارائه می شوند. با وجود مزایای فراوانی که اینترنت به همراه دارد، دسترسی کنترل نشده به آن می تواند مشکل ساز باشد. در اینترنت مطالب و تصاویر مستهجن و غیراخلاقی فراوانی وجود دارد. از آنجایی که این نوع مطالب تأثیر منفی بر روی کودکان و نوجوانان دارد، در اکثر کشورهای جهان، قوانین و اقدامات خاصی برای محافظت از کودکان در برابر این نوع مطالب وجود دارد، که هدف آن ها جلوگیری از دسترسی کودکان و نوجوانان به محتوای پایگاه های غیراخلاقی است [۱]. بنابراین تکنیک ها و روش های هوشمندی که بتوانند تصاویر و محتوای غیراخلاقی را به صورت خودکار تشخیص دهند، به شدت در حوزه پالایش محتوای وب و نرم افزارهای کنترل والدین بر کودکان مورد نیاز هستند.

* رایانامه نویسنده مسئول: alatif@yazd.ac.ir

۲- پیشینه تحقیق

بررسی روش‌های پیشین در زمینه تشخیص تصاویر غیراخلاقی که برخی از آن‌ها در ادامه به‌طور جزئی‌تر بررسی خواهند شد، نشان می‌دهد که به‌طور کلی سیستم‌های بازشناسی تصاویر غیراخلاقی از سه گام عمده تشکیل شده‌اند: بخش‌بندی نواحی پوستی، استخراج ویژگی و طبقه‌بندی تصاویر.

از آن‌جا که تصاویر غیراخلاقی همواره شامل برهنگی هستند لذا همبستگی قوی‌ای بین تصاویر غیراخلاقی و درصد نواحی پوست موجود در تصویر وجود دارد. به همین دلیل اولین گام در سیستم‌های طبقه‌بندی تصاویر غیراخلاقی تشخیص نواحی پوستی در تصویر است. با این حال موانعی برای بخش‌بندی دقیق تصویر به نواحی پوستی و غیرپوستی وجود دارد که از آن جمله می‌توان به کیفیت پایین تصویر، وجود اشیاء غیر انسان با رنگ مشابه رنگ پوست و عدم اشباع نواحی پوستی به دلیل انعکاس پوست و شرایط نوری اشاره کرد.

روش‌های بخش‌بندی نواحی پوستی به دو دسته روش‌های مبتنی بر پیکسل و روش‌های مبتنی بر ناحیه تقسیم‌بندی می‌شوند. در مراجع [۲-۳] از جدول مراجعه‌نرمال شده^۱، برای تعیین احتمال پوست بودن یک پیکسل رنگی استفاده می‌شود. در این روش فضای رنگ به تعدادی بازه گسسته‌سازی می‌شود. مقدار فراوانی در هر یک از این بازه‌ها یک هیستوگرام دو بعدی یا سه بعدی را به نام جدول مراجعه تولید می‌کنند. فراوانی هر بازه تعداد دفعات رخداد بازه رنگی متناظر آن را در نمونه‌های آموزشی ذخیره می‌کند. در صورتی که احتمال پوست بودن یک پیکسل از یک حد آستانه بیش‌تر باشد، به عنوان پیکسل پوستی طبقه‌بندی می‌شود. مشکل این روش استفاده از آستانه و عدم تعمیم‌پذیری است.

هر چند که اطلاعات رنگی نقش اصلی در مدل‌سازی پوست را ایفا می‌کند، با این حال انتخاب فضای رنگ مناسب برای نمایش پوست نیز امری مهم است. مقایسه‌های متعددی بین فضاهای رنگی مختلف برای تشخیص پوست را می‌توان در مرجع [۴-۵] پیدا کرد. مرجع [۶] از ویژگی فضای رنگی YIQ و مدل آماری پوست و روش k-میانگین برای تشخیص پوست استفاده کرده‌اند.

در مرجع [۷] از مدل مخلوط گوسی برای مدل‌سازی رنگ پوست استفاده می‌شود. به‌طور معمول پیکسل‌های پوستی نمی‌توانند به درستی با یک توزیع گوسی مدل شوند به همین دلیل مدل مخلوط گوسی مناسب‌تر است. هنگام آموزش یک مدل مخلوط گوسی متداول‌ترین روش برای تعیین پارامترها استفاده از الگوریتم EM^۲ است.

در مرجع [۸] از تقریب خوشه‌های رنگی پوست با ۳ توزیع

گوسی سه بعدی در فضای رنگ YCbCr استفاده شده است. در این روش در صورتی که فاصله ماهالانوبیس بردار رنگ c تا نزدیک‌ترین خوشه مدل از یک حد آستانه از پیش تعریف شده کم‌تر باشد، به عنوان پیکسل پوستی در نظر گرفته می‌شود.

Lee و Yoo [۹] با مقایسه توزیع‌های رنگ پوست و غیر پوست در فضاهای رنگی مختلف به این نتیجه رسیدند که به دلیل عدم تقارن خوشه رنگ پوست حول پیک چگالی، استفاده از یک مدل گوسی متقارن منجر به بالا رفتن نرخ تشخیص اشتباه می‌شود. به همین دلیل از مدل مرز بیضی شکل برای آموزش و ارزیابی استفاده کردند که نسبت به مدل تک گوسی و مدل مخلوط گوسی نتایج بهتری را فراهم می‌کند.

Bosson و همکاران در [۱۰] از نرخ درست‌نمایی پوست در فضای رنگ HSV برای تشخیص پوست استفاده کردند. در این مقاله در گام طبقه‌بندی مدل خطی تعمیم یافته، k نزدیک‌ترین همسایه، شبکه عصبی پرسپترون چند لایه^۳ و ماشین بردار پشتیبان^۴ با هم مقایسه شدند. نتایج ارزیابی برتری MLP را نسبت به سایر روش‌های طبقه‌بندی نشان می‌دهد.

در مقاله [۱۱] از توصیفگر هندسی مقاوم در برابر تبدیلات افاین و مدل پوست در مدل تصادفی مارکوف^۵ برای شناسایی ناحیه پوستی حاوی محتوای غیر اخلاقی استفاده شده است.

در مقاله [۱۲] پس از انتقال مدل رنگ از فضای RGB به HSV و تقسیم‌بندی پیکسل‌ها به پوستی و غیر پوستی، با استفاده از درصد پیکسل‌هایی که به عنوان رنگ پوست شناسایی شده‌اند، شناسایی تصاویر غیر اخلاقی انجام می‌شود. مشکل این روش تعیین آستانه برای درصد پیکسل‌های پوستی است.

در مقاله [۱۳] یک پایگاه داده از تصاویر سالم و غیر اخلاقی ایجاد شده است. برای شناسایی نوع تصویر ورودی، ابتدا ناحیه‌ی پوستی از پس‌زمینه حذف می‌شود، سپس ۱۰۰ تصویر مشابه شامل ویژگی‌های بصری مانند توصیفگرهای رنگ، بافت و شکل از بین تصاویر سالم و غیر اخلاقی از پایگاه داده استخراج می‌شود. در صورتی تعداد تصاویر تطبیق یافته از پایگاه داده که مربوط به تصاویر غیر اخلاقی هستند بیش‌تر از یک حد آستانه باشد، تصویر ورودی نیز به عنوان محتوای غیر اخلاقی دسته‌بندی می‌شود.

مروری بر پیشینه تحقیقات نشان می‌دهد اکثر روش‌های موجود تنها بر روی استخراج ویژگی‌های مناسب برای بخش‌بندی پیکسل‌های پوستی تمرکز کرده‌اند و در جداسازی تصاویر غیراخلاقی تنها از نسبت پیکسل‌های پوستی به کل پیکسل‌های تصویر استفاده کرده‌اند. در این مقاله علاوه بر ارائه رویکردی جدید برای بخش‌بندی پیکسل‌های پوستی بدون نیاز به آستانه تعریف شده توسط کاربر، از ویژگی‌های بیشتری برای تمایز به

3 -MultiLayer Perceptron (MLP)

4 -Support Vector Machine (SVM)

5 -Markov Random Fields

1 -Normalized Lookup Table

2 -Expectation-Maximization

تصاویر سالم و غیراخلاقی استفاده شده است.

تصاویر گرفته شده در شرایط نوری زیاد رخ می‌دهد.

۳- سیستم پیشنهادی

شکل (۱) دیاگرام بلوکی کلی سیستم پیشنهادی را نشان می‌دهد. سیستم شامل چهار گام اصلی است. در گام اول مجموعه‌ای از پیش‌پردازش‌ها بر روی تصویر ورودی انجام می‌شود. در گام بعد تصویر به دو بخش نواحی پوستی و نواحی غیرپوستی بخش‌بندی می‌شود. در گام سوم مجموعه‌ای از ویژگی‌های مناسب برای طبقه‌بندی تصاویر از تصویر بخش‌بندی شده استخراج می‌شود و در نهایت در گام چهارم طبقه‌بندی تصاویر توسط طبقه‌بند و با کمک ویژگی‌های استخراج شده انجام می‌شود.

۳-۱- پیش‌پردازش

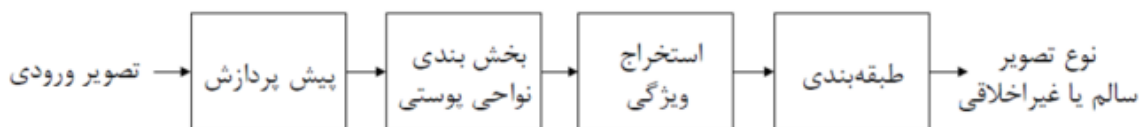
همه تصاویر دارای کیفیت یکسان نیستند. برخی از تصاویر ممکن است در شرایط نوری نامناسب تصویربرداری شده باشند و یا دارای اثرات مصنوعی ناشی از فشرده‌سازی و یا دارای اندازه کوچک باشند که از جزئیات خوبی برخوردار نیستند. به منظور استفاده از تکنیک تشخیص پوست مبتنی بر پیکسل، نیاز است تا تصویر ورودی تا حد امکان متعادل‌سازی شود. در این‌جا از روش اعمال کنتراست خودکار^۱ برای بهبود کیفیت تصویر ورودی استفاده شده است.

شکل (۲) نتیجه حاصل از اعمال کنتراست خودکار بر روی یک تصویر نمونه نشان می‌دهد. تابع کنتراست خودکار، محدوده رنگی تصویر را به بازه ماکزیمم بسط می‌دهد. به بیان دیگر، در تصاویر با کیفیت نوری پایین، که پوست تاریک‌تر از حالت شرایط نرمال خود دیده می‌شود، کنتراست خودکار محدوده رنگی را به محدوده قهوه‌ای قابل تشخیص تبدیل می‌کند. همین اتفاق برای

۳-۲- بخش‌بندی نواحی پوستی

به دلیل همبستگی بالا بین تصاویر غیراخلاقی و میزان نواحی پوستی موجود در تصویر در اکثر روش‌ها از بخش‌بندی تصویر به دو ناحیه پوستی و غیر پوستی استفاده شده است. در واقع می‌توان گفت مهم‌ترین عامل در تشخیص برهنگی قابلیت تشخیص پوست است. در حالت ایده‌آل، سیستم باید قادر به تشخیص هرگونه پوست با رنگ‌های متفاوت (سیاه پوست، سفیدپوست، سبزه و موارد دیگر) باشد. هم‌چنین سیستم باید نسبت به تغییرات نوری مقاوم باشد، چرا که محیط تصویربرداری ممکن است شرایط نورپردازی مناسبی نداشته باشد. به منظور تشخیص پیکسل‌های با رنگ پوست بهتر است بهترین فضای رنگ انتخاب شود. هدف از انتقال فضای رنگ کاهش هم‌پوشانی پیکسل‌های پوستی و غیر پوستی در فضای رنگ جدید است تا در نهایت بتوان به طبقه‌بندی صحیح‌تری از پیکسل‌های پوستی و غیرپوستی دست یافت.

به طور معمول تأثیر تغییرات نوری در مؤلفه شدت روشنایی^۲ بیشتر از مؤلفه‌های رنگی^۳ است و لذا می‌توان یک تشخیص‌گر مقاوم پوست را با حذف مؤلفه روشنایی و استفاده از مؤلفه رنگی به‌دست آورد. L در فضای رنگ JHS، G در RGB، nG در YnRGB در YCbCr مؤلفه شدت روشنایی هستند. Shin در [۱۴] نشان داد در بین ۹ فضای رنگ، فضای رنگ دوبعدی YCbCr گرفته شده از فضای رنگ سه بعدی YCbCr بهترین تفکیک بین پیکسل‌های پوستی و غیر پوستی را فراهم می‌کند. هم‌چنین نتایج بدست آمده توسط [۱۵] نیز فضای رنگ YCbCr را بهترین فضای رنگ برای تفکیک پیکسل‌های پوستی و غیرپوستی نشان می‌دهند.



شکل (۱). ساختار سیستم پیشنهادی



(ب) تصویر بعد از اعمال کنتراست خودکار



(الف) تصویر اصلی

شکل (۲). پیش‌پردازش

بر خلاف فضای رنگ RGB که تغییرات شدت روشنایی محیط بر روی هر سه مؤلفه اثر می‌گذارد، در فضای رنگ YCbCr به راحتی می‌توان تغییرات شدت روشنایی را از محتوای رنگی جدا کرد. از همین رو در این‌جا نیز از همین فضای رنگ برای جداسازی پیکسل‌های پوستی استفاده شده است.

فضای رنگ YCbCr به طور گسترده در پردازش تصاویر دیجیتال استفاده می‌شود. در این فضای رنگ، اطلاعات درخشندگی^۱ با یک مولفه‌ی Y نمایش داده می‌شود و اطلاعات رنگ به صورت دو مولفه‌ی مختلف Cb و Cr ذخیره می‌گردد. مولفه‌ی Cb تفاوت بین مولفه‌ی آبی و یک مقدار مرجع، و مولفه‌ی Cr تفاوت بین مولفه‌ی قرمز و یک مقدار مرجع است. تبدیل استفاده شده برای نگاشت فضای رنگ RGB به فضای رنگ YCbCr توسط رابطه (۱) ارائه شده است.

برای تخمین تابع چگالی توزیع پیکسل‌های پوستی از هیستوگرام دو بعدی ساخته شده از مقادیر لایه Cb و Cr استفاده شده است. محتوای هر ستون هیستوگرام، تعداد پیکسل‌های پوستی با رنگ متناظر آن ستون در فضای CbCr را نشان می‌دهد. شکل (۶) هیستوگرام دو بعدی ساخته شده را نشان می‌دهد.

به منظور نرمال‌سازی تمامی مقادیر بر کل تعداد پیکسل‌های پوستی تقسیم شده‌اند تا یک هیستوگرام نرمال شده حاصل شود. هیستوگرام نرمال شده، احتمال پوست بودن هر پیکسل با رنگ مشخص را نشان می‌دهد. هیستوگرام نرمال شده را می‌توان به صورت رابطه (۲) بیان کرد.

$$p_{skin}(c) = \frac{\text{اندازه ستون هیستوگرام مربوط به رنگ } c}{\text{تعداد کل پیکسل‌های پوستی}} \quad (2)$$

مقدار $p_{skin}(c)$ با یک حد آستانه Th مقایسه می‌شود. در صورتی که احتمال $p_{skin}(c)$ از حد آستانه بیش‌تر باشد پیکسل به عنوان پیکسل پوستی در نظر گرفته می‌شود. به منظور اجتناب از آستانه، از یک شبکه عصبی مصنوعی^۲ برای تصمیم‌گیری در مورد پیکسل استفاده می‌شود. در تئوری، ANN نتیجه بهتری نسبت به استفاده از آستانه در هر کانال را می‌دهد زیرا شبکه عصبی روابط بین کانال‌ها را نیز در نظر می‌گیرد. این شبکه عصبی مصنوعی با استفاده از همان مجموعه داده قبلی که برای ساخت هیستوگرام دو بعدی استفاده شد، آموزش داده می‌شود.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112.000 \\ 112.000 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

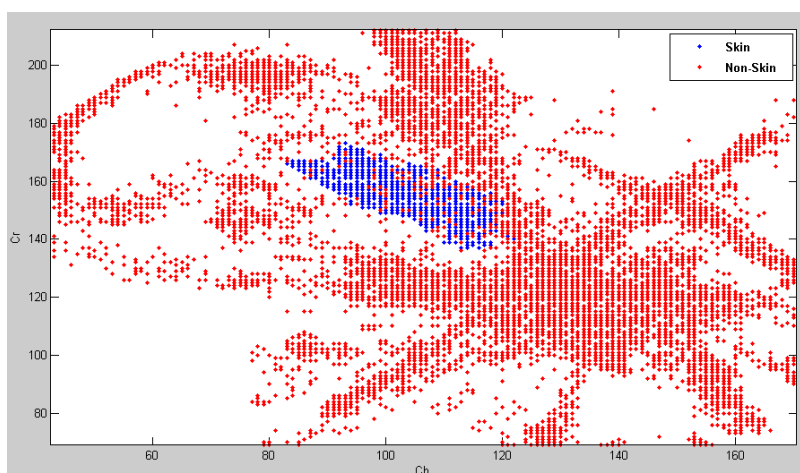
شکل‌های (۳-۴) دو تصویر شامل پوست‌های با نژاد رنگی مختلف را به همراه سه مؤلفه فضای رنگ YCbCr نشان می‌دهد. برای تشخیص پیکسل‌های پوستی از هیستوگرام نرمال شده پیکسل‌های پوستی استفاده شده است. برای این منظور ابتدا از یک مجموعه از پیکسل‌های برجسب‌دار در فضای RGB استفاده شده است که شامل ۲۴۵۰۵۸ پیکسل است که از این میان ۵۰۸۵۹ پیکسل مربوط به پوست و باقی آن مربوط به پیکسل‌های غیر پوست است. مقادیر این پیکسل‌ها به فضای رنگ



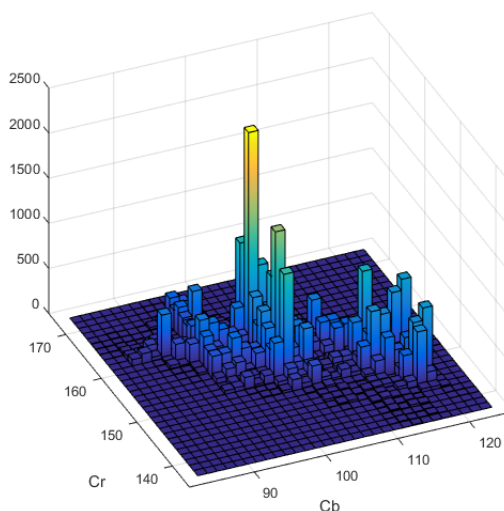
شکل (۳). تصویر یک شخص با نژاد سیاه پوست در فضای YCbCr



شکل (۴). تصویر تصویر یک شخص با نژاد سفید پوست در فضای YCbCr



شکل (۵). توزیع رنگ‌ها در فضای رنگ CbCr، محور افقی نشانگر مؤلفه Cb و محور عمودی نشانگر مؤلفه Cr.



شکل (۶). هیستوگرام دو بعدی در فضای CbCr مربوط به پیکسل‌ها پوستی

می‌شود. المان ساختاری مورد استفاده برای عملیات مورفولوژی، مربعی با اندازه 5×5 در نظر گرفته شده است.

۳-۳- استخراج ویژگی از تصویر

پس از بخش‌بندی تصویر به نواحی پوستی و غیر پوستی، مجموعه‌ای از ویژگی‌های متمایز کننده برای طبقه‌بندی تصاویر غیراخلاقی استخراج می‌شوند. هر چه ویژگی‌های استخراج شده دارای قابلیت متمایزکنندگی بیش‌تری باشند، فرایند طبقه‌بندی با دقت بالاتری انجام می‌شود. در ادامه ویژگی‌های استفاده شده در سیستم پیشنهادی بیان خواهند شد.

۳-۳-۱- اندازه بزرگ‌ترین ناحیه پوست

اندازه بزرگ‌ترین ناحیه پوست می‌تواند با اهمیت باشد. این ناحیه به صورت درصدی از کل تصویر محاسبه می‌گردد. در صورتی که ناحیه‌های پوستی کوچک بسیاری در تصویر شناسایی شده باشد این ویژگی، کمک شایانی به طبقه‌بندی می‌کند.

شبکه عصبی استفاده شده برای تشخیص پیکسل پوستی شامل چهار نرون ورودی است. سه نرون ورودی برای سه مؤلفه فضای رنگ RGB و یک نرون ورودی برای میزان احتمال $P_{skin}(c)$. هم‌چنین از یک لایه پنهان با ۱۰ نرون پنهان و دو نرون به عنوان خروجی شبکه عصبی استفاده شده است. به منظور نرمال‌سازی ورودی‌ها، مقادیر RGB قبل از ورود به شبکه عصبی مصنوعی، بر ۲۵۵ تقسیم شده‌اند. و لذا هر سه ورودی مقدار بین صفر و یک دارند. برای آموزش و تست شبکه عصبی از پایگاه داده UCI^۱ شامل ۲۴۵۰۵۸ پیکسل استفاده شده است که ۷۰ درصد داده‌ها برای آموزش و بقیه برای تست استفاده شده‌اند.

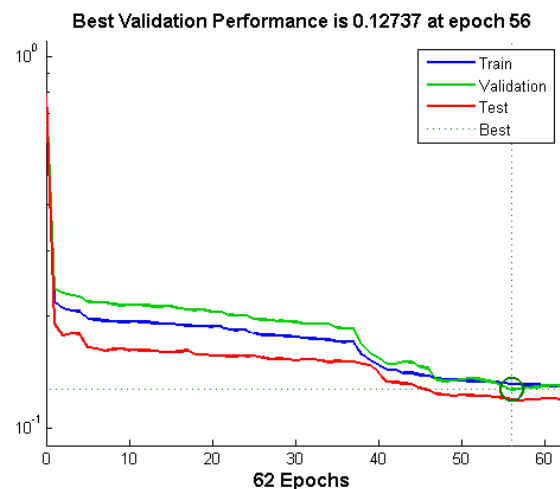
پس از تشخیص ناحیه‌های پوستی معمولاً تصویر باپیری خروجی شامل حفره‌ها یا نویزهایی است که به منظور کاهش این اثرات از فیلترهای مورفولوژی ریاضی استفاده شده است. برای این منظور برای حذف حفره‌ها و هم‌چنین برای حذف نویزهای موجود به ترتیب از عملیات بستن و بازکردن تصویر استفاده

دو کلاس از یکدیگر در راستای عمود بر مرز تصمیم‌گیری حداکثر شود.

۴- نتایج و پیاده‌سازی

سیستم پیشنهادی در نرم‌افزار متلب نسخه ۸ (۲۰۱۴a) و بر روی سیستم با ۸ مگابایت حافظه اصلی و پردازنده ۲،۳ گیگاهرتزی پیاده‌سازی و آزمایش شد. همان‌طور که در بخش ۲-۳، گفته شد، از یک مجموعه داده آموزشی برچسب‌دار برای محاسبه هیستوگرام نرمال شده استفاده شد. برای این منظور از مجموعه داده بخش‌بندی پوست [16] استفاده شده است. این دیتاست به صورت تصادفی مقادیر R، G و B را از تصاویر چهره مربوط به گروه‌های سنی مختلف (جوان، میان‌سال و پیر) و نژادهای مختلف (سفید، سیاه و آسیایی) نمونه‌برداری کرده است. این مجموعه داده شامل ۵۰۸۵۹ پیکسل پوستی و ۱۹۴۱۹۸ پیکسل غیر پوستی است. در حقیقت مجموعه داده مذکور شامل یک فایل متنی است که هر سطر آن مقادیر لایه‌های رنگی R، G و B و برچسب متناظر با رنگ (صفر به معنای غیرپوست و یک به معنای پوست) را در بردارد.

همان‌طور که گفته شد پس از تخمین چگالی پیکسل‌های پوستی با استفاده از روش هیستوگرام نرمال شده از یک شبکه عصبی برای طبقه‌بندی پیکسل‌های پوستی و غیر پوستی استفاده می‌شود. شکل (۷) نمودار خطای شبکه عصبی در طول فرایند آموزش را نشان می‌دهد. همان‌طور که نشان داده شده است شبکه بعد از ۵۶ تکرار و با کارایی ۰/۱۲ متوقف شده است. همچنین دلیل توقف فرایند آموزش، افزایش متوالی خطای مربوط به داده‌های اعتبارسنجی به تعداد از پیش تعیین شده (شش مرتبه) بوده است.



شکل (۷). خطای شبکه عصبی مربوط به طبقه‌بندی پیکسل‌ها در طول فرایند آموزش

۳-۲- تعداد چهره‌ها

تعداد چهره‌های موجود در یک تصویر اطلاعات مفیدی در رابطه با سالم یا غیراخلاقی بودن تصویر فراهم می‌کند. به عنوان مثال اگر ۷۰ درصد پیکسل‌های درون تصویر را پوست تشکیل دهد. با این اطلاعات به احتمال زیاد تصویر مورد نظر از نوع غیر اخلاقی خواهد بود، اما در صورتی که بدانیم تعداد افراد موجود در تصویر بیش‌تر از یک نفر است احتمال غیر اخلاقی بودن تصویر کاهش می‌یابد. به منظور این که این ویژگی همواره عددی بین صفر و یک باشد از معکوس این معیار استفاده شده است. در این مقاله از الگوریتم Viola - Jones [5] برای استخراج چهره از تصویر استفاده می‌شود.

۳-۳- درصد چهره

تصاویر پرتله^۱ چهره نیز حاوی درصد بالایی از پوست هستند. در این حالت تصویر با برچسب غیراخلاقی طبقه‌بندی می‌گردد. تصاویری که دارای تعدادی زیادی چهره هستند و سایر نواحی پوست بدن را به صورت اندک دارند، با استفاده از این تکنیک می‌توانند به صورت مطلوب طبقه‌بندی شوند.

۳-۴- تعداد نواحی پوستی

تعداد نواحی تشخیص داده شده به عنوان پوست می‌تواند بر روی سالم یا غیراخلاقی بودن تصویر اثرگذار باشد. به منظور حذف نواحی کوچک (نویز در تصویر) الگوریتم تنها نواحی که اندازه آن‌ها از یک آستانه مشخص بزرگ‌تر هستند را می‌شمارد.

۳-۵- درصد ناحیه پوست در تصویر

نسبت پیکسل‌های پوستی به کل پیکسل‌های تصویر معیار مناسبی در تشخیص تصاویر غیر اخلاقی است چرا که در تصاویر غیر اخلاقی اغلب بخش زیادی از تصویر به پیکسل‌های پوستی تعلق دارد. نحوه محاسبه این نسبت در رابطه (۳) نشان داده شده است.

$$Skin_Ratio = \frac{\text{تعداد کل پیکسل‌های تصویر}}{\text{پیکسل‌های پوستی تصویر}} \quad (3)$$

۴-۳- طبقه‌بندی تصاویر

پس از استخراج ویژگی‌های مرتبط، سیستم پیشنهادی، از طبقه‌بند ماشین بردار پشتیبان برای طبقه‌بندی تصاویر در دو دسته سالم و غیر اخلاقی استفاده می‌کند. ماشین بردار پشتیبان روشی برای به دست آوردن مرز بهینه جدا کننده دو کلاس است که این فرآیند مستقل از توزیع احتمالاتی نمونه‌های آموزشی هر یک از کلاس‌ها انجام می‌گیرد. در این روش مرز تصمیم‌گیری در جایی قرار می‌گیرد که فاصله نزدیک‌ترین نمونه‌های آموزشی هر

شده است.

جدول (۱). مقایسه روش پیشنهادی روی پایگاه داده UCI

روش	Recall	Precision
مرجع [۶]	۶۹/۰۰	۸۱/۶۵
مرجع [۱۷]	۸۰/۷۹	۵۰/۸۲
روش پیشنهادی	۸۰/۴	۸۰/۲



(ب) تصویر حاصل از بخش‌بندی مربوط به تصویر (الف)



(الف) تصویر اصلی



(د) تصویر حاصل از بخش‌بندی مربوط به تصویر (ج)



(ج) تصویر اصلی

شکل (۹). چند تصویر نمونه از بخش‌بندی تصویر به نواحی پوستی و غیر پوستی

به منظور حذف حفره‌ها و هم‌چنین نویزهای موجود در تصویر باینری حاصل از عملیات مورفولوژی ریاضی استفاده شده است. شکل (۱۰) تصویر حاصل از شناسایی پیکسل‌های پوستی توسط شبکه عصبی را قبل و بعد از اعمال فیلترهای مورفولوژیکی نشان می‌دهد. همان‌طور که دیده می‌شود اعمال این فیلترها می‌تواند نتیجه حاصل از بخش‌بندی را برای پردازش‌های بعدی بهبود بخشد چرا که مؤلفه‌های همبندی کوچک حذف شده و حفره‌های ناشی از نویز نیز پر شده‌اند.



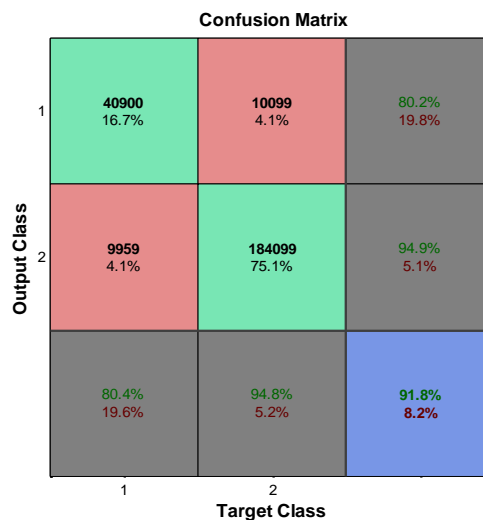
(ب) تصویر پس از اعمال عمل بستن



(الف) تصویر حاصل از بخش‌بندی

شکل (۱۰). نتیجه حاصل از اعمال عملیات ریخت شناسی

شکل (۸)، ماتریس درهم‌ریختگی مربوط به طبقه‌بندی پیکسل‌های پوستی و غیرپوستی را نشان می‌دهد که حاکی از دقت ۹۱/۸ درصد برای بخش‌بندی صحیح پیکسل‌های پوستی است. همان‌طور که مشاهده می‌شود تعداد ۴۰۹۰۰ پیکسل معادل ۱۶/۷ درصد از کل داده‌ها، از کلاس یک (کلاس پوست) به طور درست به عنوان پوست طبقه‌بندی شده‌اند. همچنین ۱۸۴۰۹۹ پیکسل معادل ۷۵/۱ درصد از کل داده‌ها مربوط به کلاس دو (کلاس غیرپوست) بوده است که به درستی تشخیص داده شده است. در مقابل ۹۹۵۹ پیکسل پوستی معادل حدوداً ۴/۱ درصد از کل داده‌ها مربوط به کلاس یک (کلاس پوست) بوده و به نادرست به عنوان غیر پوست تشخیص داده شده است. همچنین ۱۰۰۹۹ معادل حدوداً ۴/۱ درصد از کل داده‌ها مربوط به کلاس دو (کلاس غیرپوست) بوده است که به طور نادرست در کلاس یک یعنی کلاس پوست طبقه‌بندی شده است. جدول (۱) مقایسه بین صحت^۱ و فراخوانی^۲ را برای طبقه‌بندی پیشنهادی و دو روش دیگر نشان می‌دهد.



شکل (۸). ماتریس درهم‌ریختگی حاصل از بخش‌بندی پوستی

در جدول (۱) مقایسه بین چند طبقه‌بندی پوستی با روش پیشنهادی بر روی پایگاه داده UCI نشان داده شده است. نحوه محاسبه این مقادیر در روابط (۴) و (۵) آورده شده است.

$$P = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \quad (۴)$$

$$R = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \quad (۵)$$

پس از آموزش شبکه عصبی، از تصاویر موجود در پایگاه داده YACVI به منظور بخش‌بندی با استفاده از طبقه‌بندی حاصل به دو ناحیه پوستی و غیر پوستی استفاده شد. چندین تصویر از این پایگاه داده و تصویر دودویی متناظر آن‌ها در شکل (۹) نشان داده

۷۰ درصد داده‌ها برای آموزش و از ۳۰ درصد برای تست و اعتبار سنجی استفاده شد. در نهایت دقت ۸۹/۸ درصد با طبقه‌بند ماشین بردار پشتیبان حاصل شد. در آزمایشی دیگر تمامی نمونه‌های آموزش و تست ترکیب شده و از رویکرد اعتبار سنجی متقاطع^۲ 5-fold برای به دست آوردن دقت استفاده شد. در رویکرد اعتبار سنجی متقاطع 5-fold، تمامی نمونه‌ها به صورت تصادفی به ۵ بخش مساوی تقسیم می‌شوند. یکی از پنج بخش به عنوان داده اعتبار سنجی برای تست دسته‌بند و باقی داده‌ها برای آموزش استفاده می‌شود. این فرایند اعتبار سنجی متقاطع پنج مرتبه تکرار می‌شود، در نهایت میانگین ارزیابی‌ها محاسبه می‌شود. مزیت استفاده از رویکرد اعتبار سنجی متقاطع این است که تمامی نمونه‌ها برای ارزیابی و آموزش استفاده می‌شوند. مقادیر به دست آمده از فرایند اعتبار سنجی متقاطع 5-fold به صورت جدول (۲) می‌باشد که میانگین نرخ دسته‌بندی ۹۰/۹۱ است.

نمونه‌ای از تصاویر سالم استفاده شده در شکل (۱۱) نشان داده شده است. هم‌چنین شکل (۱۲) چند نمونه از تصاویر غیراخلاقی را پس از بخش‌بندی نشان می‌دهد.

جدول (۲). نتایج حاصل از اعتبار سنجی متقاطع 5-fold در ماشین بردار پشتیبان

مرحله اول	مرحله دوم	مرحله سوم	مرحله چهارم	مرحله پنجم	میانگین 5-fold
۸۹/۲۷	۹۲/۱۱	۹۰/۹۵	۹۱/۴۱	۹۰/۸۴	۹۰/۹۱



شکل (۱۲). چند نمونه از تصاویر غیراخلاقی پس از بخش‌بندی

خروجی هر طبقه‌بند دودویی چهار معیار پی‌آمد دارد. نرخ مثبت درست^۱ تعداد نمونه‌هایی را نشان می‌دهد که طبقه‌بند به درستی یک نمونه مثبت را مثبت طبقه‌بندی کرده است. هر گاه نمونه‌ای مثبت باشد و توسط طبقه‌بند منفی طبقه‌بندی شود این نمونه به عنوان

پس از استخراج ویژگی‌های مربوطه از نواحی پوستی، برای هر تصویر یک بردار ویژگی در فضای پنج بعدی به دست می‌آید که از این بردار ویژگی برای طبقه‌بندی تصویر با استفاده از یک ماشین بردار پشتیبان استفاده می‌شود.

دقت ماشین بردار پشتیبان، تا حد زیادی به انتخاب پارامترهای مدل بستگی دارد. در این مقاله از الگوریتم جستجوی شبکه‌ای [۱۸] برای انتخاب مدل مناسب استفاده می‌شود. در این روش ماشین بردار پشتیبان با تمامی جفت‌های ممکن از پارامترها آموزش داده می‌شود، در نهایت پارامترهایی که بهترین نتیجه را فراهم کنند برای آموزش مدل استفاده می‌شوند. دامنه جستجو برای پارامترها به صورت $\gamma = [2^{-11}, 2^{-9}, \dots, 2^1, 2^3]$ و $C = [2^{-5}, 2^{-3}, \dots, 2^3, 2^5]$ می‌باشد، بنابراین ماشین بردار پشتیبان با $6 \times 8 = 48$ ترکیب ممکن از جفت پارامترها آموزش داده می‌شود که بهترین نرخ اعتبار سنجی با مقادیر $\gamma = 2^{-7}$ و $C = 2^3$ به دست آمد. همچنین به منظور غلبه بر بیش برآزش^۱ از رویکرد اعتبار سنجی 5-fold استفاده می‌شود.

با توجه به عدم وجود مجموعه داده استاندارد در این زمینه، به منظور آموزش ماشین بردار پشتیبان از یک مجموعه شامل



شکل (۱۱). چند نمونه تصویر سالم در ارزیابی طبقه‌بند



شکل (۱۴). نمونه تصویری که روش پیشنهادی قادر به شناسایی آن نیست

۵- نتیجه گیری

در این مقاله روشی جدید برای شناسایی تصاویر غیر اخلاقی با استفاده از ویژگی‌های رنگی پوست ارائه شده است. فیلتر پیشنهادی از سه قسمت آشکارسازی نواحی پوستی، استخراج ویژگی و طبقه‌بندی تصاویر تشکیل شده است.

در ابتدا از ویژگی‌های شدت روشنایی در فضای رنگی RGB و احتمال پوست بودن هر پیکسل در فضای رنگی YCBCR به عنوان ورودی شبکه عصبی استفاده می‌شود تا بخش‌بندی نواحی پوستی و غیر پوستی انجام شود. سپس از نواحی پوستی آشکار شده پنج ویژگی شامل اندازه بزرگ‌ترین ناحیه پوستی، تعداد چهره‌ها، درصد چهره، تعداد نواحی پوستی و درصد ناحیه پوست در تصویر استخراج می‌شود در نهایت از این ویژگی‌ها به منظور دسته‌بندی تصاویر به سالم و غیر اخلاقی توسط ماشین بردار پشتیبان استفاده می‌شود. هر دو طبقه‌بند (طبقه‌بند پیکسل‌های پوستی و طبقه‌بند تصاویر سالم و غیر اخلاقی) نتایج قابل قبولی را از نظر نرخ طبقه‌بندی درست ارائه دادند. نوآوری مقاله را از دو منظر می‌توان بیان کرد. در بخش آشکارسازی پیکسل‌های پوستی، روشی برای تشخیص پیکسل‌های پوستی از غیرپوستی بدون استفاده از آستانه تعیین شده توسط کاربر ارائه شده است که قابلیت تفکیک بالایی دارد. در قسمت طبقه‌بندی تصاویر غیر اخلاقی نیز از جمله مزایای روش پیشنهادی می‌توان به قدرت تفکیک بیشتر بین دو کلاس سالم و غیر اخلاقی اشاره کرد که علت آن استفاده از ویژگی‌های بیش‌تر از جمله تعداد چهره‌ها برای طبقه‌بندی می‌باشد. با این حال از معایب روش پیشنهادی عدم تشخیص درست تصاویر پوستی مربوط به چهره بزرگ‌نمایی شده یا سایر نواحی مجاز بدن است. به عنوان کارهای آینده، پیشنهاد می‌شود در بخش‌بندی پیکسل‌ها به نواحی پوستی و غیرپوستی از اطلاعات مبتنی بر شکل ناحیه و بافت ناحیه نیز استفاده شود تا دقت تشخیص نواحی پوستی را بهبود داد. همچنین استخراج و شناسایی اعضای مختلف بدن بی شک در بهبود دقت موفق خواهد بود.

منفی نادرست^۱ شمارش می‌شود. هرگاه یک نمونه منفی به کلاس منفی طبقه‌بندی شود به عنوان منفی درست^۲ شمارش می‌شود و در پایان منفی نادرست تعداد نمونه‌های منفی را نشان می‌دهد که به اشتباه در کلاس مثبت طبقه‌بندی شده‌اند. شکل (۱۳) ماتریس درهم‌ریختگی مربوط به طبقه‌بندی ۴۰۰ تصویر پایگاه داده جمع‌آوری شده را نشان می‌دهد.

دقت طبقه‌بندی پیشنهادی روی مجموعه داده‌های جمع‌آوری شده برابر ۸۹/۸٪ است. از ماتریس درهم‌ریختگی معیارهای متفاوتی قابل محاسبه است. در این میان دو مورد از مهم‌ترین آن‌ها عبارتند از نرخ مثبت صحیح و نرخ مثبت کاذب که به ترتیب توسط روابط (۶) و (۷) محاسبه شده‌اند. بیش‌تر بودن مقدار کمی TPR و کم‌تر بودن مقدار کمی FPR به معنای بهتر بودن طبقه‌بند ارائه شده است.

		Target Class		
		1	2	
Output Class	1	184 46.0%	25 6.3%	88.0% 12.0%
	2	16 4.0%	175 43.8%	91.6% 8.4%
		1	2	
		92.0% 8.0%	87.5% 12.5%	89.8% 10.3%

شکل (۱۳). ماتریس درهم‌ریختگی مربوط به طبقه‌بندی تصاویر

$$TPR = \frac{\text{True Positive}}{\text{All Positive}} = \frac{184}{200} = 0.92 \quad (۶)$$

$$FNR = \frac{\text{False Negative}}{\text{All Negative}} = \frac{25}{200} = 0.125 \quad (۷)$$

نتایج حاصل از شبیه‌سازی نشان می‌دهد زمانی که یک تصویر حاوی مقدار زیادی نواحی پوستی است و الگوریتم ویولاجونز قادر به یافتن چهره در آن نیست، تصویر بدون هرگونه معیار دیگری در گروه تصاویر غیر اخلاقی تشخیص داده می‌شود. این مشکل به دلیل عدم وجود ویژگی‌هایی برای تشخیص عضو بدن است و نمی‌توان آن را به ناکارآمدی طبقه‌بند نسبت داد. شکل (۱۴) نمونه‌ای از این تصاویر سالم که به صورت غیر اخلاقی تشخیص داده شده‌اند را نشان می‌دهد.

۶- مراجع

- [9] J. Lee and S. Yoo, "An elliptical boundary model for skin color detection," In International Conference on Imaging Science, Systems, and Technology, 2002 .
- [10] A. Bosson, G. Cawley, Y. Chan, and R. Harvey, "Non-retrieval: blocking pornographic images," Image and Video Retrieval, pp. 50-60, 2002 .
- [11] M. Islam, P. Watters, and J. Yearwood "Real-time detection of children's skin on social networking sites using Markov random field modelling," Information Security Technical Report , pp. 51-58, 2011 .
- [12] J. Marcial-Basilio, G. Aguilar-Torres, G. Sanchez-Perez, L. Toscano-Medina, and H. Perez-Meana, "Detection of pornographic digital images," International journal of computers, pp. 298-305, 2011 .
- [13] J. Shih, C. Lee, and C. Yang, "An adult image identification system employing image retrieval technique," Pattern Recognition Letters, pp. 2367-2374., 2007 .
- [14] M. Shin, K. Chang, and L. Tsap, "Does colorspace transformation make any difference on skin detection," in Proceedings In Applications of Computer Vision, 2002 .
- [15] K. Shaik, P. Ganesan, V. Kalist, B. Sathish, and J. Jenitha, "Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color Space," in Procedia Computer Science, 2015 .
- [16] R. Bhatt and A. Dhall, "Skin segmentation dataset," UCI Machine Learning Repository, 2010 .
- [17] S. Fekri-Ershad, M. Saberi, and F. Tajeripour, "An Innovative Skin Detection Approach Using Color Based Image Retrieval Technique," in arXiv preprint arXiv, 2012 .
- [18] C. Hsu, C. Chang, and C. Lin, "A practical guide to support vector classification," 2003.
- [1] H. Zhu, S. Zhou, J. Wang, and Z. Yin, "An algorithm of pornographic image detection," In Fourth International Conference on Image and Graphics, 2007 .
- [2] J. Basilio, G. Torres, G. Perez, L. Medina, and H. Meana, "Explicit image detection using YCbCr space color model as skin detection," Applications of Mathematics and Computer Engineering, pp. 123-128, 2011 .
- [3] Y. Chen, K. Hu, and S. Ruan, "Statistical skin color detection method without color transformation for real-time surveillance systems," Engineering Applications of Artificial Intelligence , pp. 1331-1337, 2012 .
- [4] C. Prema and D. Manimegalai, "Survey on skin tone detection using color spaces," International Journal of Applied Information Systems, vol. 2, no. 2, pp. 18-26, 2012 .
- [5] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001 .
- [6] H. Al-Mohair, J. Saleh, and S. Suandi, "Hybrid Human Skin Detection Using Neural Network and K-Means Clustering Technique," Applied Soft Computing, pp. 337-347, 2015 .
- [7] T. Jebara and A. Pentland, "Parametrized structure from motion for 3D adaptive feedback tracking of faces," In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997 .
- [8] R. Hsu, M. Abdel-Mottaleb and A. Jain, "Face detection in color images," IEEE transactions on pattern analysis and machine intelligence, pp. 696-706, 2002 .

Detection of Pornographic Digital Images Using Support Vector Machine and Neural Network

S. M. Khorashadizadeh, V. Azadzadeh, A. M. latif*

*Yazd University

(Received: 07/01/2016, Accepted: 31/10/2016)

ABSTRACT

In this paper a new approach for detecting explicit content or pornographic images is proposed. The suggested system contains three phases. In the first phase, the image is segmented into skin region and non-skin region by an ANN in a pixel-wise manner. After image segmentation, a set of discriminative features including the ratio of skin pixels to total image pixels, the number of faces in the image, the area of largest skin area are extracted. Finally, support vector machines are used to classify the image in two normal image and explicit image using the extracted feature vector. A set of 400 images is used to learn and test the system. The detection rate for classifying pixels in skin and non-skin is 91.8 and as the result, the proposed system could obtain 89.9 accuracy in classifying images based on their content. Furthermore, the true positive rate and false positive rate for the proposed system are 92% and 0.125%, respectively.

Key words: Images Classification, Explicit Detection, Multi-Layer Perceptron, Support Vector Machine, Dense Estimation