

آشکارسازی چهره با شبکه‌های عصبی در تصاویر رنگی

محمود محلوجی

محمد باقر منهج

گروه برق و کامپیوتر دانشگاه آزاد اسلامی - واحد کاشان

دانشکده مهندسی برق دانشگاه صنعتی امیر کبیر

E-mail: mmahlouji@yahoo.com

E-mail: menhaj@aut.ac.ir

چکیده - در این مقاله، آشکارسازی چهره با شبکه‌های عصبی در تصاویر رنگی بررسی می‌شود. با توجه به اینکه رنگ پوست در فضای *ycbcr* تشکیل یک خوشه را می‌دهد، لذا یک شبکه عصبی پرسپترون چندلایه یا *MLP* برای جداسازی نواحی پوستی استفاده شد. برای تست شبکه از نمونه‌های پوست صورت موجود در تصاویر تهیه شده از یک دوربین دیجیتال و یک پایگاه عکس دانشجویی استفاده گردید و نتایج زیر حاصل شد: در فریم‌های ویدئویی به دلیل سادگی زمینه تصویر، با یک شبکه (1-4-6-2) نتایج بسیار خوبی حاصل شد. در دومین تست شبکه به دلیل پیچیدگی زمینه عکس‌ها، بهترین شبکه با ساختار (1-60-90-2) نتایج قابل قبولی را ارائه داد.

کلید واژه - آشکارسازی چهره، رنگ پوست، شبکه عصبی *MLP*، فضای رنگی *ycbcr*

1- مقدمه

آشکارسازی چهره در تصاویر، کاربردهای وسیعی در نظارت ویدئویی، تعامل انسان کامپیوتر و ردیابی چهره دارد. تحقیقات زیادی در این مورد انجام شده است [1-9]. بعضی از کارهای انجام شده در این مورد به آشکارسازی چهره در تصاویر سطح خاکستری محدود شده است [5]، که در آن آشکارسازی چهره در تصاویر از روبرو با تاکید بر ویژگی‌های صورت نظیر چشم‌ها و گوشه‌های لب انجام گرفته است. کارهای دیگری نظیر [9] به شناسایی چهره یک فرد در تصویر محدود شده است، درحالیکه آشکارسازی چهره در تصویری که دارای چندین چهره انسان هستند، مورد نیاز است. بر عکس در این مقاله، هدف قابلیت آشکارسازی چهره در محدوده وسیعی از جهات مورد دید یا از دیدگاه‌های مختلف با تاکید بر ویژگی‌های عمومی چهره است. ویژگی استفاده شده برای آشکارسازی چهره، رنگ پوست صورت انسان است. اولین قدم برای رسیدن به این منظور آشکارسازی ناحیه پوست در تصویر است. یکی از روش‌هایی که در آشکارسازی رنگ پوست مورد استفاده قرار گرفته

است مبتنی بر ارائه مدلی برای رنگ پوست است [6]. با توجه به اینکه رنگ پوست در فضای رنگی *ycbcr* تشکیل یک کلاستر یا خوشه را می‌دهد، به همین جهت یک رویکرد مستقیم به شبکه‌های عصبی برای آشکارسازی رنگ پوست مناسب به نظر می‌رسد. در این مقاله از شبکه‌های عصبی پرسپترون چندلایه با الگوریتم یادگیری پس‌انتشار خطا یا BP برای تفکیک نواحی پوستی و غیرپوستی در تصویر استفاده شده است. همچنین برای آموزش شبکه عصبی از نمونه‌های پوست صورت موجود در تصاویر تهیه شده از یک دوربین دیجیتال و یک پایگاه عکس دانشجویی استفاده شد.

2- مشخصات رنگ پوست انسان

در بسیاری از کاربردها نظیر ذخیره‌سازی و انتقال ویدئویی معمولاً رنگ‌ها به مولفه‌های روشنایی و رنگی تفکیک می‌شوند. یکی از دلایل این تقسیم بندی، حساسیت کمتر چشم انسان به تغییرات مولفه‌های رنگی است. تجربیات روانشناسی نشان می‌دهد که ادراک رنگ‌ها دارای سه مشخصه است: شدت نور، اشباع و رنگ. در فضای رنگی

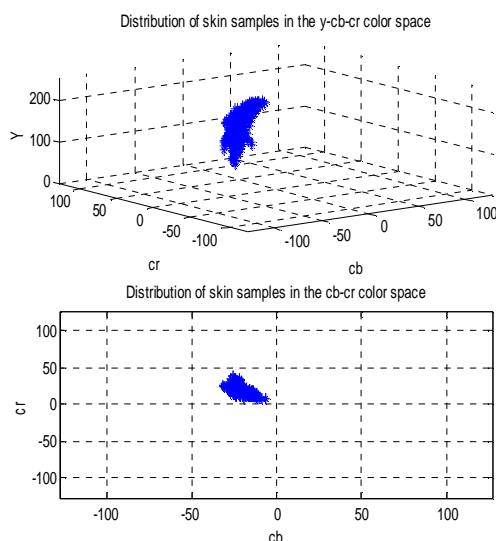
RGB را در اختیار قرار می‌دهند. برای تبدیل فضای رنگی RGB به فضای رنگی ycbcr می‌توان از رابطه زیر استفاده نمود [8]:

$$\begin{bmatrix} y \\ cb \\ cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

برای تولید رنگ پوست در فضای رنگ cbcR نمونه‌های پوستی از تصاویر صورت انتخاب شده و از رابطه (1) استفاده می‌شود. شکل 2 توزیع رنگ پوست در فضای رنگی cbcR را برای یک نمونه تصویر داده شده در شکل 1 نشان می‌دهد. ناحیه آبی رنگ در شکل 2 متناظر با مقادیر رنگی می‌باشد که دارای احتمال غیرصفر در نمونه‌های انتخاب شده می‌باشند. چنانچه در این شکل دیده می‌شود واضح است که رنگ پوست در یک ناحیه خیلی کوچک از فضای رنگی قرار دارد.



شکل 1: نمونه پوستی



شکل 2: توزیع نمونه پوستی شکل 1 در فضاهای رنگی

ycbcr شدت نور متناظر با مقدار روشنایی y است، در حالیکه اشباع و رنگ در مولفه‌های رنگی cb ، cr نگهداری می‌شوند. رنگ پوست انسان تشکیل طبقه خاصی از رنگ‌ها را می‌دهد که از اکثر رنگ‌های اشیای طبیعی دیگر تفکیک‌پذیر است. اگر چه رنگ‌های پوست از شخصی به شخصی و از نژادی به نژادی دیگر فرق می‌کند، لیکن دارای توزیع خیلی کوچک در صفحه رنگی است. این بدان معناست که رنگ‌های پوست نسبتاً در اشباع و رنگ ثابت اند [7]. رنگ‌هایی به‌عنوان رنگ پوست طبقه‌بندی می‌شوند که اشباع و رنگ آن‌ها در محدوده خاصی قرار داشته باشند. با حذف مولفه روشنایی از رنگ‌ها، اختلاف بین رنگ‌های پوست افراد و نژادهای مختلف و تاثیر شرایط نورانی کاهش داده می‌شود. در این مقاله از مشخصات رنگ پوست انسان با به‌کارگیری مولفه‌های cb و cr استفاده شده و با استفاده از توزیع رنگ پوست در فضای رنگی cbcR پروسه دسته‌بندی بین رنگ پوست و غیرپوست (چهره و غیرچهره) انجام گرفته است.

2-1- استخراج رنگ پوست

علی‌رغم اینکه رنگ پوست انسان از یک شخص به شخص دیگر و از یک نژاد به نژاد دیگر بسیار فرق می‌کند، لیکن توزیع آن در یک ناحیه خیلی کوچک در فضای رنگی cbcR می‌باشد [8]. در این مقاله، از فضای رنگی ycbcr برای آشکارسازی رنگ پوست استفاده شده است. در فضای رنگ ycbcr اطلاعات روشنایی از اطلاعات رنگ جدا می‌شود. شکل 2 توزیع رنگ پوست در فضای رنگ ycbcr را برای یک نمونه رنگ پوست که در شکل 1 نشان داده شده است، نمایش می‌دهد. چنانچه در این شکل دیده می‌شود، مقدار شدت روشنایی y تاثیر چندانی در یافتن رنگ پوست ندارد، در نتیجه از مقدار روشنایی y صرف‌نظر می‌شود. همچنین در شکل 2 توزیع رنگ پوست در فضای رنگی cbcR نشان داده شده است. چنانچه در این شکل دیده می‌شود، رنگ پوست تشکیل یک کلاستر کوچک در فضای cbcR را می‌دهد، بنابراین استفاده از مقادیر cb ، cr برای آموزش شبکه عصبی MLP مناسب به‌نظر می‌رسد.

2-2- تولید رنگ پوست در فضای رنگی cbcR

خروجی اکثر دوربین‌های رنگی، تصاویر رنگی در فضای رنگ

3- شبکه‌های عصبی پیشنهادی

چنانچه بیان شد، کلاس پوست انسان تشکیل یک ناحیه کوچک در فضای رنگ cbcR را می‌دهد، در نتیجه استفاده از شبکه‌های عصبی برای آشکارسازی رنگ پوست مناسب به نظر می‌رسد. با توجه به توانایی شبکه‌های عصبی MLP در طبقه‌بندی الگوها بخصوص در مواردی که مرز تصمیم‌گیری بین کلاس‌های مختلف الگوها دارای پیچیدگی زیاد و کاملاً غیرخطی است، از این نوع شبکه‌ها برای ایجاد تمایز بین کلاس‌های پوست و غیرپوست استفاده گردید. برای آموزش شبکه از الگوریتم BP و مقادیر cb و cr در فضای رنگ cbcR استفاده شد، بدین ترتیب تعداد ورودی‌های شبکه برابر دو انتخاب گردید. هر چند وجود یک لایه مخفی در شبکه MLP کافی به نظر می‌رسد لیکن قرار دادن دو لایه مخفی، همگرایی سریعتر و هموارتر شبکه را باعث می‌شود [3]. تعداد نرون‌های موجود در لایه‌های مخفی در ابتدا برابر مقادیر مشخصی لحاظ شده سپس در صورتیکه شبکه همگرا نمی‌شد، تعداد نرون‌های لایه‌های مخفی افزایش و در غیر این صورت تعداد آن‌ها کاهش داده شدند تا یک مقدار بهینه حاصل شود. لایه خروجی دارای یک نرون بوده و تنها خروجی شبکه دارای مقداری در محدوده 1- و 1 می‌باشد. برای آموزش شبکه عصبی، انتخاب داده‌های یادگیری در دو حالت زیر انجام گرفته و شبکه مورد تست قرار گرفت:

الف) یک فریم از یک برش ویدئویی که با استفاده از یک دوربین دیجیتال از یک شخص در حال حرکت تهیه شده بود، انتخاب شده سپس نمونه‌های آموزشی از نمونه‌های پوست صورت شخص حاضر در تصویر و زمینه آن فراهم گردید. پس از آموزش شبکه با این نمونه‌ها، تست نهایی شبکه با انتخاب فریم‌های دیگری از برش ویدئویی انجام گرفت.

ب) تعدادی عکس از یک پایگاه عکس دانشجویی انتخاب شده سپس تعدادی از آن‌ها برای آموزش شبکه و تعدادی دیگر برای تست شبکه به کار برده شدند. مشابه نمونه‌های آموزشی از نمونه‌های پوست صورت افراد حاضر در عکس‌های انتخاب شده برای آموزش شبکه و زمینه آن‌ها فراهم شد. آموزش شبکه با این نمونه‌ها انجام شده و نهایتاً تست شبکه با نمونه عکس‌های انتخابی برای تست شبکه انجام گردید.

4- نتایج آزمایشات

4-1- نتایج آزمایشات روی فریم‌های ویدئویی

اولین تصاویر آزمایش شده فریم‌های برگرفته از یک برش ویدئویی بودند که از طریق یک دوربین دیجیتال از یک شخص در حال حرکت گرفته شده بود. شکل 3 یک فریم نمونه برای آموزش شبکه است که از ناحیه صورت شخص حاضر در تصویر، نمونه پوستی و از زمینه آن، نمونه‌های غیرپوستی نشان داده شده در شکل 4 انتخاب شده‌اند. در شکل 5 توزیع نمونه‌های پوستی و غیرپوستی انتخاب شده، نشان داده شده است. برای آموزش شبکه عصبی از تصویر نمونه پوستی تعداد 375 نقطه و از تصاویر نمونه‌های غیرپوستی تعداد 403 نقطه با توزیع یکنواخت انتخاب شدند. سپس با توجه به اینکه بعضی از نمونه‌های غیرپوستی در زمینه تصویر اصلی دارای رنگ پوست بوده و در فضای رنگی cbcR با خوشه تشکیل شده از نمونه‌های پوستی همپوشانی داشتند، با انتخاب یک معیار فاصله با آستانه یک واحد در فضای رنگی cbcR، نمونه‌های غیرپوستی که دارای فاصله اقلیدسی کمتر از یک واحد در فضای رنگی cbcR از نمونه‌های پوستی بودند، حذف گردیدند. بدین ترتیب بین دو کلاس پوست و غیرپوست حداقل یک واحد فاصله در فضای رنگی انتخاب شده وجود دارد و در نتیجه شبکه عصبی امکان تفکیک دو کلاس را خواهد داشت. در این مثال خاص تعداد نقاطی که به این ترتیب با اعمال معیار فاصله فوق حذف گردیدند، 7 نقطه بود که در شکل 5 با رنگ آبی نمایش داده شده است. این روش در نتیجه نهایی تأثیر منفی محسوسی ایجاد نکرد. خلاصه اینکه مجموعه داده‌های یادگیری ورودی و خروجی در این مثال خاص شامل 771 بردار ورودی انتخاب گردید که از این تعداد 375 بردار مربوط به نمونه‌های پوستی با مقدار خروجی یک



شکل 3: تصویر اصلی

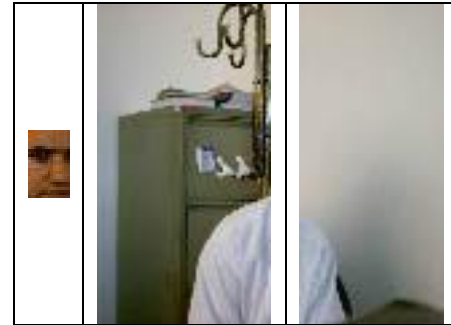
3667 سیکل موفق به یادگیری داده‌ها شده است. ساختارهای مختلف دیگری برای شبکه عصبی مورد آزمایش قرار گرفتند، از بین این ساختارها بهترین ساختار شبکه (2-6-4-1) بود که معرفی شد.

4-1-1- آشکارسازی چهره

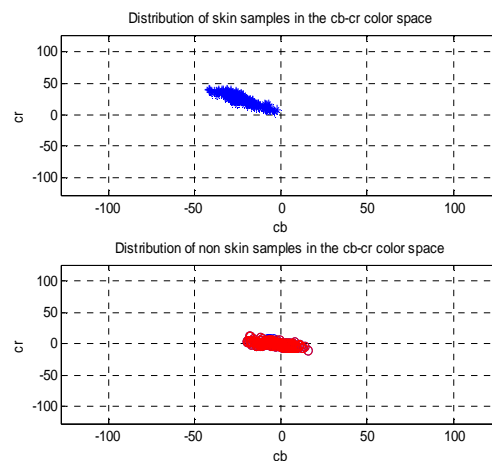
خروجی شبکه عصبی مقداری در محدوده 1- و 1 به هر نقطه تصویر نسبت می‌دهد، بدین ترتیب هر نقطه تصویر دارای مقداری است که نشان می‌دهد این نقطه به پوست شباهت دارد یا خیر. با در نظر گرفتن آستانه‌ای برابر صفر می‌توان تصویر ورودی را به یک تصویر باینری تبدیل نمود که در آن نقاط پوستی و غیرپوستی مشخص می‌شوند. شکل 7 نتایج حاصل از به‌کاربردن عملیات مذکور را روی تصویر اصلی که در شکل 3 نشان داده شده است و به‌عنوان تصویر نمونه برای آموزش شبکه‌های عصبی مطرح شده به‌کار رفتند، نشان می‌دهد. در این شکل نقاط سفید نمایانگر نقاط پوستی و نقاط سیاه نمایانگر نقاط غیرپوستی می‌باشند. چنانکه در تصویر موجود در شکل 7 دیده می‌شود بعضی از نقاط زمینه تصویر به‌عنوان نقاط پوستی معرفی شده‌اند. دلیل این مطلب وجود بعضی از رنگ‌های پوستی در زمینه تصویر می‌باشد. در شکل 8 و 9 نتایج آشکارسازی چهره روی دو فریم دیگر از برش ویدئویی با شبکه (2-6-4-1) نشان داده شده است. چنانکه در این تصاویر دیده می‌شود علاوه بر چهره، نقاط پوستی دست شخص حاضر در تصویر نیز آشکار شده است. همچنین نقاطی از زمینه تصویر نیز دارای رنگ پوست بوده و آشکارسازی شده است.



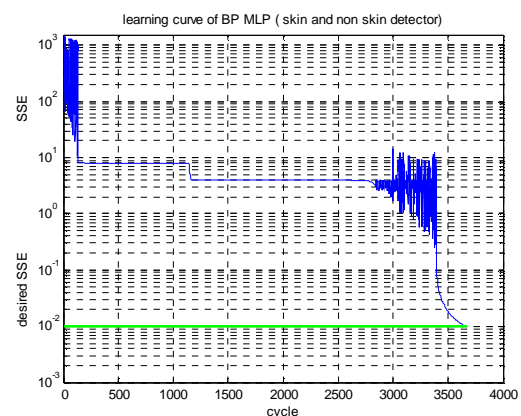
شکل 7: نتایج آشکارسازی چهره روی تصویر اصلی



شکل 4: نمونه‌های غیرپوستی و نمونه پوستی



شکل 5: توزیع نمونه‌های پوستی و غیرپوستی در فضای رنگی



شکل 6: منحنی یادگیری شبکه (2-6-4-1)

و مابقی مربوط به نمونه‌های غیرپوستی با مقدار خروجی منفی یک می‌باشند.

در ابتدا شبکه عصبی MLP با ساختار (2-6-4-1) در نظر گرفته شد. مقادیر اولیه پارامترهای شبکه بصورت تصادفی انتخاب شدند. منحنی یادگیری شبکه در شکل 6 آورده شده است. چنانچه در این شکل دیده می‌شود، شبکه بعد از

4-2- نتایج آزمایشات روی عکس‌های

دانشجویی

نمونه دیگری از عکس‌ها که مورد تست قرار گرفتند، یک مجموعه عکس دانشجویی بود که تعداد 100 عکس به عنوان نمونه انتخاب شدند. از بین این عکس‌ها تعداد 50 عکس برای آموزش انتخاب شدند. نمونه‌های رنگ پوست از رنگ پوست صورت افراد حاضر در عکس‌ها انتخاب شده و نمونه‌های غیر پوست از زمینه عکس‌های مزبور انتخاب گردیدند. با توجه به وجود و مشاهده بعضی از نمونه‌های رنگ پوست در اشیاء غیرپوستی در فضای رنگ cbr، رنگ‌های پوست و غیرپوست با یکدیگر همپوشانی دارند. مخصوصاً این همپوشانی در ناحیه مرزی دو کلاس پوست و غیرپوست در فضای رنگ فوق باعث می‌شود تا تفکیک این دو کلاس با مشکل روبرو گردد زیرا فاصله بین دو کلاس در نواحی مرزی اطراف کلاس پوست در نقاط مختلف به اندازه کافی نبود، یا عبارت دیگر در ناحیه مرزی دو کلاس نوعی ابهام در تشخیص و جداسازی این دو کلاس وجود دارد. برای تمایز کامل دو کلاس پوست و غیرپوست و رفع ابهام فوق در مرزبندی این دو ناحیه، نیاز به الگوریتم خاصی وجود داشت. الگوریتمی بشرح زیر به کار گرفته شد:

ابتدا نقاطی از کلاس پوست با فاصله آستانه‌ای مشخص از نقاط کلاس غیرپوست را حذف نموده سپس همین عمل برای کلاس غیرپوست با آستانه دیگری مشابه انجام گرفت، بدین ترتیب که نقاطی از کلاس غیر پوست با فاصله آستانه‌ای مشخص از نقاط باقیمانده کلاس پوست حذف گردیدند. نمونه‌هایی از عکس‌های انتخاب شده که آشکارسازی چهره در آن‌ها از طریق شبکه آموزش یافته (1-60-90-2) انجام شده است، در شکل 10 آورده شده‌اند. چنانکه در نمونه‌های این شکل دیده می‌شود، مشکل زمینه تصویر برای مواردی که دارای رنگ پوست نسبتاً زیاد و بهم پیوسته می‌باشد، وجود دارد. این مشکل در بعضی موارد حتی باعث تغییر شکل هندسی و ساختار صورت نیز گردیده است. برای حل این مشکل نیاز به در نظر گرفتن ویژگی‌های دیگر ضروری به نظر می‌رسد.



شکل 8: نتایج آشکارسازی چهره با شبکه (1-4-6-2)

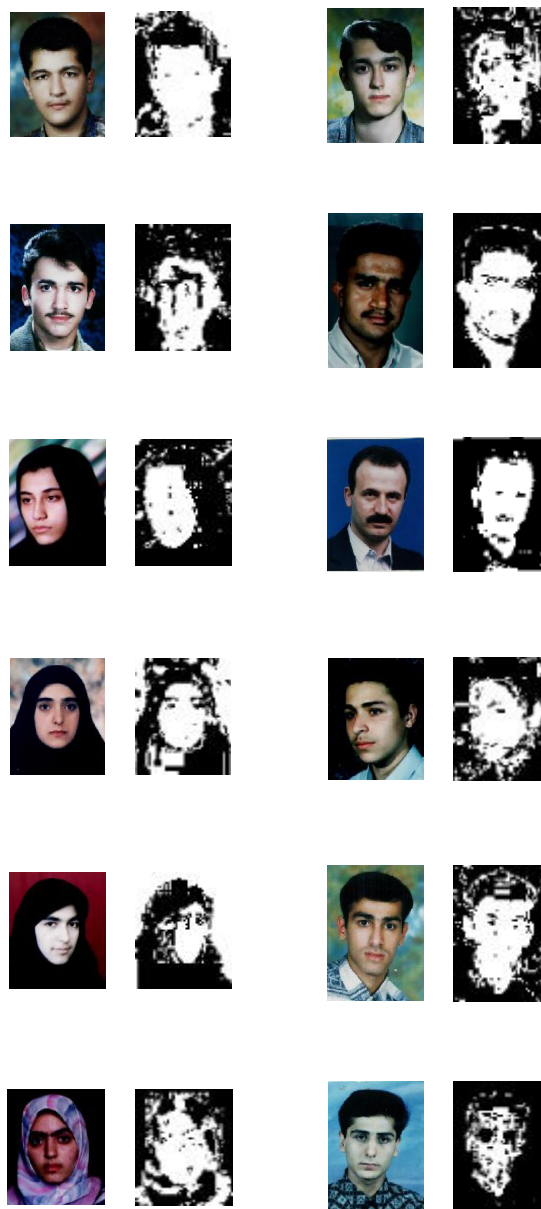


شکل 9: نتایج آشکارسازی چهره با شبکه (1-4-6-2)

منبع مختلف تهیه گردیده و نتایج زیر حاصل شد: در فریم‌هایی که از دوربین دیجیتال گرفته شدند، به دلیل سادگی زمینه تصویر، با یک شبکه (1-4-6-2) نتایج خوبی حاصل شد که با توجه به کم بودن تعداد پارامترهای شبکه، براحتی می‌توان از این شبکه برای آشکارسازی چهره در سکانس‌های ویدئویی استفاده نمود، بدین ترتیب این روش در کاربردهای ردیابی چهره قابل استفاده است. دومین تست شبکه با عکس‌های یک پایگاه عکس دانشجویی انجام شد. به دلیل پیچیدگی و تنوع زمینه عکس‌های مذکور، شبکه بدست آمده دارای نرون‌های زیادتری در لایه‌های مخفی بود که بهترین شبکه دارای ساختار (1-60-90-2) می‌باشد. نتایج بدست آمده برای شبکه بهینه در حد قابل قبولی بوده لیکن در تصاویری که زمینه آن‌ها دارای رنگ‌های پوستی نسبتاً زیاد و بهم پیوسته می‌باشند، این روش ممکن است دچار مشکل گردد، بخصوص در مواردی که این بخش‌های زمینه با قسمت پوستی آشکار شده صورت انسان تشکیل یک ناحیه بهم پیوسته را می‌دهد و در نتیجه باعث تغییر شکل هندسی و ساختار بخش آشکار شده صورت انسان می‌گردد.

مراجع

- [1] L. L. Huang, A. Shimizu, Y. Hagihara, H. Kobatake, "Face detection from cluttered images using a polynomial neural network," *Elsevier Neurocomputing* 51, pp. 197-211, 2003.
- [2] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face Detection in Color Images," *CSE Department, Michigan State University, MI 48824*, 2000.
- [3] M. Seshadrinathan, J. Ben-Arie, "Face Detection by Integration of Evidence," *ECE Department, University of Illinois at Chicago*, 2003.
- [4] G. A. Ramirez, O. Fuentes, "Face Detection Using Combinations of Classifiers," *Proc. IEEE Second Canadian Conf. on computer and Robot Vision (CRV'05)*, 2005.
- [5] H.A.Rowley, S.Baluja and T.Kanade, "Neural Network-Based Face detection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, Jan. 1998, pp. 23-38.
- [6] Haiyuan Wu, Qian Chen and Masahiko Yachida, "Face Detection from Color Image Using a Fuzzy Pattern Matching Method," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 6, pp. 557-563, June 1999.
- [7] J. C. Terrilon, M. N. Shirazi, H. Fukamachi and S. Akamatsu, "Comparative Performance of Different Chrominance Models and Chrominance Spaces for the Automatic Detection of Human Faces in Color Images," *Proc. IEEE Int'l Conf. on Face and gesture recognition*, pp. 54-61, 2000.
- [8] H. Wang and S. F. Chang, "A Highly Efficient System for Automatic Face Region Detection in MPEG Video," *IEEE Trans. on Circuit Systems for Video Technology*, Vol. 7, No. 4, pp. 615-628, 1997.
- [9] رضا آقایی، حسین نظام‌آبادی‌پور و احسان‌اله کبیر، "آشکارسازی چهره با شبکه عصبی"، دهمین کنفرانس مهندسی برق ایران، اردیبهشت 81.



شکل 10: نتایج آشکارسازی چهره روی تصاویر مختلف با شبکه (1-90-60-2) با آستانه‌های به ترتیب برابر 1 و 2

5- نتیجه‌گیری

در این مقاله، آشکارسازی چهره با شبکه‌های عصبی در تصاویر رنگی مورد بررسی قرار گرفت. از ویژگی‌های مختلفی که می‌توان برای آشکارسازی چهره استفاده نمود، ویژگی رنگ به‌عنوان یک ویژگی عمومی به‌کار گرفته شد. شبکه استفاده شده یک شبکه MLP با دو لایه مخفی و الگوریتم یادگیری پس‌انتشار خطا بود. فضای رنگی به‌کار برده شده فضای cbrc می‌باشد. نمونه‌های آموزشی و تست شبکه از دو