



دانشگاه صنعتی شاهرود

دانشکده برق و رباتیک

پایان نامه دوره کارشناسی ارشد مهندسی برق-الکترونیک

تشخیص احساس شادی و غم از طریق بررسی

تصاویر دو بعدی چهره

نوشین نبی زاده

استاد راهنما

دکتر علی سلیمانی ایوری

استاد مشاور

دکتر علیرضا احمدی فرد

تیر ماه ۱۳۸۸

چکیده

یکی از مهمترین راه های ارتباطی انسان از طریق چهره و حالات آن است. حالات چهره در انتقال منظور و ارتقاء کیفیت ارتباطات انسانی نقشی اساسی بازی می کند. در دنیای صنعتی امروز تشخیص خودکار احساسات چهره در طیف وسیعی از زمینه ها از قبیل مطالعات روان شناسی و حقوقی، انیمیشن سازی، روباتیک، لب خوانی، درک تصاویر و ویدئو کنفرانس ها، مخابرات و ارتباطات تلفنی، تشخیص موارد مشکوک در محافظت های امنیتی و ضد تروریستی، شناسایی افراد و هم چنین تقابل انسان و ماشین کاربرد دارد. دانشمندان از سی سال گذشته تا کنون فعالیت های زیادی در این زمینه انجام داده اند و موفقیت های چشمگیری در رشد و توسعه این سیستم ها به دست آورده اند، به طوری که چه در زمینه ی تشخیص حالات هفت گانه در تصاویر ویدئویی و زنده و چه در زمینه ی بررسی تصاویر استاتیک و ضبط شده به سیستم هایی با بازده ی بالای ۸۰٪ رسیده اند.

تمامی این سیستم ها مبتنی بر سه بخش انتخاب ویژگی، استخراج ویژگی و کلاسه بندی آن ها هستند و از یک نقطه نظر، روش های استخراج ویژگی به دو دسته ی روش های هالستیک و روش های آنالیتیک تقسیم بندی می شوند. از جنبه ی دیگر این روش ها در سه گروه روش های تصویر پایه، مدل پایه و ویژگی پایه قابل تفکیک اند.

در این پایان نامه پس از بررسی های بسیار این نکته آشکار شد که عموم روشهای موجود یا دارای پیچیدگی های محاسباتی بالا و یا راندمان پایین هستند و بزرگترین مشکل در این گونه مسائل ایجاد مصالحه ای بین این دو عامل است. لذا به دنبال الگوریتم های موجود چندین پیشنهاد برای تکمیل و بهینه کردن آنها ارائه شد و نهایتاً الگوریتمی با نرخ تشخیص ۹۹٪ در تشخیص احساسات شادی و غم به دست آمد. در ابتدا برای استخراج نقاط اصلی چهره از الگوریتم بهبود یافته ی لوپپاس استفاده شد. استخراج ویژگی به کمک آنالیز مولفه های اساسی و تلفیق آن با فیلترهای گابریل صورت گرفت و برای کلاسه بندی روش های معیار فاصله، شبکه عصبی و تلفیق آن با شبکه آماری به کار رفت. بازده ۹۹٪ حاصل پیاده سازی فیلترهای گابریل و آنالیز مولفه های اساسی روی چشم و دهان استخراج شده از طریق الگوریتم نقاط برجسته ی لوپپاس و کلاسه بندی به کمک شبکه عصبی پرسپترون است. روش پیشنهادی شبکه آماری زمینه ی بازی را برای بررسی بیشتر در کارهای آینده ارائه می دهد.

Abstract

One of the important ways of human's communication is established by face and its expressions. Human expression plays an important role in transmission of inner emotions and improvement of quality of the human's communications. Nowadays, in industrialized world, automatic recognition of face expression covers a broad spectrum ranging from psychological and legal studies, animation synthesis, robotic, images comprehension and video conference, communication and broadcasting networks, recognition of suspicious treat for taking security and antiterrorist measures, to human-machine interface. From thirty years ago till now, Scientifics have done many researches and activities in this field and they have achieved salient successes in advancement and improvements of these systems. As we can see they could achieve the recognition rate up to 80 percent in recognition of seven expressions of human both in real time-dynamic image and static and record images.

All of these systems are based upon three sections as features selection, features extraction, and features classification. From one point of view, facial feature extraction methods are categorized in holistic and analytic methods. From other point of view, these methods are categorized in feature-based, model-based, and image based methods.

In this thesis after deep studies, it founds that a lot of current techniques either have mathematical complexity or low recognition rate. So, finding a tradeoff between these two is the biggest question, therefore by following current algorithms, we proposed several supplementary proposal to optimize former methods and finally we approached to an algorithm with a recognition rate up to 99 percent accuracy in sadness or happiness expression recognition. At first step, we used modified Loupias salient point method for separating the basic element of face. Feature extraction was done with using principal component analysis and Gabor filters and principal component analysis. Distance criterion and artificial neural network and statistical network were used for feature classification. Recognition rate of 99% was achieved of using Gabor filters and principal component analysis on eyes and mouth that were achieved of Loupias salient point method, and perceptron neural network for classification. Proposed statistical network method announce broad field for more scrutiny in future researches.

فهرست مطالب

فصل اول.....مقدمه

- ۱-۱-مقدمه.....(۱)
- ۱-۲-کاربرد سیستم تشخیص احساس.....(۲)
- ۱-۳-مروری بر تاریخچه و روش ها.....(2)
- ۱-۴-پایگاه داده ها.....(8)
- ۱-۵-ساختار پایان نامه.....(۱۰)

فصل دوم.....مروری بر روش های پر کاربرد تشخیص احساس چهره

- ۱-۲-۱-مقدمه.....(۱۱)
- ۲-۲-تشخیص احساس چهره از طریق روش ویژگی-پایه ی بررسی واحد های حرکتی.....(۱۱)
- ۱-۲-۲-تشخیص واحد های حرکتی پایین چهره برای آنالیز احساس چهره.....(۱۲)
- ۱-۲-۲-۱-معرفی سیستم.....(۱۲)
- ۱-۲-۲-۲-مدل های چند حالتی برای اجزا چهره.....(۱۳)
- ۱-۲-۲-۳-مدل های چند حالتی لب.....(۱۴)
- ۱-۲-۲-۴-استخراج ویژگی های پایین چهره.....(۱۴)
- ۱-۲-۲-۵-نمایش ویژگی های پایین چهره.....(۱۶)
- ۲-۲-۲-تشخیص واحد های حرکتی بالای چهره برای آنالیز احساس چهره(۱۸)
- ۱-۲-۲-۲-معرفی سیستم.....(۱۸)
- ۱-۲-۲-۲-مدل چشم دو حالتی.....(۱۸)
- ۱-۲-۲-۳-مدل ابرو و گونه و چین و چروک.....(۱۹)
- ۱-۲-۲-۴-ویژگی های چشم.....(۲۰)
- ۱-۲-۲-۵-ویژگی های ابرو و گونه.....(۲۱)
- ۱-۲-۲-۶-ویژگی های گذرا.....(۲۱)
- ۱-۲-۲-۷-نمایش ویژگی بالای چهره.....(۲۱)
- ۳-۲-تشخیص احساس چهره از طریق روش ویژگی-پایه ی مبتنی بر MPEG4 و کلاسیفایر فازی.....(۲۲)
- ۱-۳-۲-معرفی سیستم.....(۲۲)
- ۲-۳-۲-چارچوب MPEG4 برای حالات چهره.....(۲۳)
- ۳-۳-۲-سیستم فازی برای کلاس بندی احساسات چهره.....(۲۴)
- ۱-۳-۳-۲-نمایش فازی ها.....(۲۴)
- ۲-۳-۳-۲-نمایش فازی احساسات.....(۲۵)
- ۴-۳-۲-توضیح الگوریتم سیستم فازی.....(۲۵)
- ۵-۳-۲-نتایج سیستم مبتنی بر MPEG4 و کلاسیفایر فازی.....(۲۶)
- ۴-۲-تشخیص احساس چهره از طریق روش مدل-پایه ی گراف های کلی و محلی.....(۲۷)
- ۱-۴-۲-معرفی سیستم.....(۲۷)
- ۲-۴-۲-گراف ناحیه ای ویژگی های چهره(FRG).....(۲۷)
- ۳-۴-۲-تطبیق ویژگی های صورت با گراف محلی.....(۲۸)
- ۵-۴-۲-توسعه گراف برای تشخیص احساسات چهره.....(۲۹)
- ۵-۲-تشخیص احساس چهره از طریق روش تصویر-پایه ی مدل ظهور فعال.....(۳۰)
- ۱-۵-۲-معرفی سیستم.....(۳۰)
- ۲-۵-۲-استخراج ویژگی های چهره.....(۳۰)

- ۲-۵-۳- شبکه عصبی برای کلاسه بندی احساس صورت..... (۳۱)
- ۲-۶- تشخیص احساس چهره از طریق روش تصویر-پایه ی مدل ظهور فعال و ویژگیهای مکانی..... (۳۳)
- ۲-۶-۱- معرفی سیستم..... (۳۳)
- ۲-۶-۲- استخراج نقاط ویژگی..... (۳۳)
- ۲-۶-۳- کلاسه بندی و تشخیص حالات چهره..... (۳۴)
- ۲-۷- تشخیص احساس چهره از طریق روش تصویر-پایه ی ماتریس انتقال حالت..... (۳۶)
- ۲-۷-۱- معرفی سیستم..... (۳۶)
- ۲-۷-۲- ماتریس انتقال حالت..... (۳۷)
- ۲-۷-۳- تشخیص احساس چهره..... (۳۷)
- ۲-۸- تشخیص احساس چهره از طریق روش تصویر-پایه ی دینامیکی بر اساس توابع ویولت هار..... (۳۸)
- ۲-۸-۱- معرفی سیستم..... (۳۸)
- ۲-۸-۲- نمایش ویژگی های دینامیکی شبه هار..... (۳۹)
- ۲-۸-۳- کد کردن ویژگی های دینامیکی..... (۴۰)
- ۲-۸-۴- آموزش آدابوست..... (۴۱)
- ۲-۹- تشخیص احساس چهره از طریق روش تصویر-پایه ی استفاده از فیلترهای گابر..... (۴۲)
- ۲-۹-۱- معرفی سیستم..... (۴۲)
- ۲-۹-۲- استخراج ویژگی..... (۴۲)
- ۲-۹-۳- آنالیز اجزا اصلی..... (۴۴)
- ۲-۹-۴- کلاس بندی به کمک شبکه عصبی LVQ..... (۴۴)

فصل سوم..... مروری بر تبدیل موجک و گابر

- ۳-۱- مقدمه..... (۴۶)
- ۳-۲- تاریخچه ی موجک ها..... (۴۷)
- ۳-۳- کاربرد موجک در پردازش تصویر..... (۴۸)
- ۳-۴- توضیح مختصری درباره تبدیل گابر..... (۵۲)

فصل چهارم..... مروری بر شبکه های عصبی و آنالیز مولفه های اساسی

- ۴-۱- مقدمه..... (۵۵)
- ۴-۲- نگاهی به انواع شبکه های مصنوعی عصبی..... (۵۶)

- ۱-۲-۴- شبکه های عصبی پرسپترون.....(۵۶)
- ۲-۲-۴- فراگیری در مدل نرون.....(۵۷)
- ۳-۲-۴- پرسپترون چند لایه ای.....(۵۷)
- ۴-۲-۴- شبکه های عصبی RBF.....(۶۰)
- ۵-۲-۴- شبکه های بازگشتی.....(۶۲)
- ۱-۵-۲-۴- شبکه های المان.....(۶۲)
- ۲-۵-۲-۴- شبکه هاپفیلد.....(۶۲)
- ۶-۲-۴- شبکه های با فراگیری بدون سرپرست.....(۶۳)
- ۱-۶-۲-۴- شبکه های خود سازمان ده کوهنن.....(۶۳)
- ۳-۴- تحلیل مولفه های اساسی^۱ (PCA).....(۶۴)
- ۱-۳-۴- مقدمه.....(۶۴)
- ۲-۳-۴- مختصری درباره الگوریتم تحلیل مولفه های اساسی.....(۶۵)
- ۳-۳-۴- توضیح الگوریتم.....(۶۵)
- ۴-۳-۴- کاربرد تحلیل مولفه های اساسی در پردازش تصویر.....(۶۶)

فصل پنجم..... روش های پیشنهادی و پیاده سازی آن ها

- ۱-۵- مقدمه.....(۶۹)
- ۲-۵- مشخصه های احساسات شش گانه.....(۷۱)
- ۱-۲-۵- استخراج ویژگی های چهره.....(۷۱)
- ۳-۵- روش های پیشنهادی جهت تشخیص احساسات.....(۷۵)
- ۱-۳-۵- روش تشخیص توسط آنالیز مولفه های اساسی و معیار فاصله.....(۷۵)
- ۲-۳-۵- روش تشخیص احساسات توسط آنالیز مولفه های اساسی و شبکه عصبی.....(۸۱)
- ۱-۲-۳-۵- روش آنالیز مولفه های اساسی با یک گروه پایه-شبکه عصبی.....(۸۱)
- ۲-۲-۳-۵- الگوریتم آنالیز مولفه های اساسی با دو گروه پایه-شبکه عصبی.....(۸۳)
- ۳-۳-۵- الگوریتم تشخیصی آنالیز مولفه های اساسی- شبکه عصبی با ورودی اصلاح شده توسط گابریل.....(۸۸)
- ۴-۳-۵- الگوریتم تشخیص احساسات آماری-شبکه عصبی.....(۹۷)

فصل ششم..... نتیجه گیری و پیشنهادات

¹ Principal Component Analysis

۶-۱- نتیجه گیری.....(۱۰۴)

۶-۲- پیشنهادات.....(۱۰۶)

مراجع

پیوست

فهرست اشکال

فصل اول:

شکل (۱-۱) نمایش سیستمی بر اساس الکترومیوگرافی.....(۴)

فصل دوم:

شکل (۱-۲) سیستم feature-based تشخیص AUها.....(۱۲)

شکل (۲-۲) حالت های چرخش سر.....(۱۳)

شکل (۳-۲) اجزاء مختلف چهره دید از روبرو.....(۱۴)

شکل (۴-۲) ویژگی های گذرا برای حالات مختلف (a) چشم ها کشیده و لبخند (b) چشم ها و دهان باز (c) چشم ها بسته یا کشیده (d) دهان به شدت بسته.....(۱۶)

شکل (۵-۲) چروکهای ناسلابال.....(۱۷)

شکل (۶-۲) نمایش ویژگی های پایین چهره.....(۱۸)

شکل (۷-۲) شکل مدل چشم دو حالت. (a) چشم باز. (b) چشم بسته. (c) دیاگرام حالت گذرا. (d) مدل پارامتری چشم باز. (e) مدل پارامتری چشم بسته.....(۱۹)

شکل (۸-۲) ماسک نیم دایره عنیبه.....(۲۰)

شکل (۹-۲) نتایج دنبال کردن در (a) چشم نیمه باز و (b) چشم کاملا باز به همراه پلک زدن.....(۲۰)

شکل (۱۰-۲) ویژگی های بالای چهره.....(۲۱)

شکل (۱۱-۲) توابع عضویت احساسات مختلف.....(۲۵)

شکل (۱۲-۲) تابع عضویت حالت تعجب.....(۲۵)

شکل (۱۳-۲) شمای کلی سیستم تشخیص احساس فازی.....(۲۶)

شکل (۱۴-۲) چند حالت آزمایش شده.....(۲۶)

شکل (۱۵-۲) نمونه ی پیچیده ای از تطابق ناحیه ای.....(۲۸)

شکل (۱۶-۲) گرافهای LG حالات مختلف.....(۲۹)

شکل (۱۷-۲) حداقل سازی درجه دوم در AAM.....(۳۱)

شکل (۱۸-۲) نمونه هایی از دنبال کردن چهره با AAM.....(۳۲)

- شکل (۱۹-۲) نمونه ای از شکل و بافت برای پنج حس مختلف.....(۳۲)
- شکل (۲۰-۲) اولین ویژگی آموزش داده شده توسط آدابوست.....(۳۴)
- شکل (۲۱-۲) مکان نقاط ویژگی بر چهره.....(۳۵)
- شکل (۲۲-۲) پارامترهای دهان.....(۳۵)
- شکل (۲۳-۲) پارامترهای چشم.....(۳۶)
- شکل (۲۴-۲) ویژگی های دهان و بینی.....(۳۶)
- شکل (۲۵-۲) ماتریس انتقال حالت. (a) مپینگ مستقیم SVD (b).....(۳۷)
- شکل (۲۶-۲) انتقال تصاویر (a) حالت خنثی و (b) حالات تبدیل شده.....(۳۸)
- شکل (۲۷-۲) ساختار سیستم مبتنی بر هار.....(۳۹)
- شکل (۲۸-۲) نمونه هایی از ویژگی های شبه هار پیاده شده روی چهره.....(۳۹)
- شکل (۲۹-۲) نمونه ای از یک واحد ویژگی دینامیکی در سکانس تصویر.....(۴۰)
- شکل (۳۰-۲) نمونه ای از کدینگ یک واحد ویژگی دینامیکی.....(۴۱)
- شکل (۳۱-۲) پروسه ی تشخیص احساس.....(۴۱)
- شکل (۳۲-۲) کرنل های فیلتر گابری استفاده شده برای حصول بردار ویژگی.....(۴۳)
- شکل (۳۳-۲) مکان نقاط ثابت روی چهره.....(۴۴)
- شکل (۳۴-۲) ساختاری از شبکه LVQ.....(۴۵)

فصل سوم:

- شکل (۱-۳) نمایش دو روش بازسازی در موجک.....(۴۹)
- شکل (۲-۳-الف) تصویر اصلی لنا.....(۵۰)
- شکل (۲-۳-ب) تصویر بازسازی شده از ۱۶ سطح کوآنتیزه ی موجک.....(۵۰)
- شکل (۲-۳-ج) تصویر بازسازی شده از ۶۴ سطح کوآنتیزه ی موجک.....(۵۱)
- شکل (۲-۳-د) تصویر بازسازی شده از اعمال معکوس موجک در کل تصویر.....(۵۱)
- شکل (۳-۳) بانک فیلتری برای تبدیل موجک پیوسته تصویر.....(۵۱)
- شکل (۴-۳) اثر تغییر طول موج را در شرایط یکسان، λ های ۵ و ۱۰ و ۱۵ از چپ به راست.....(۵۳)
- شکل (۵-۳) اثر تغییر θ را در باقی شرایط یکسان، زاویه چرخش ۰ و ۴۵ و ۹۰ از چپ به راست.....(۵۳)
- شکل (۶-۳) اثر تغییر φ را در باقی شرایط یکسان، φ های ۰ و ۱۸۰ و ۹۰- و ۹۰ از چپ به راست.....(۵۳)
- شکل (۷-۳) تجزیه تصویر را با استفاده از الگوریتم هرمی گابری. (a) تصویر چهره ورودی، (b) نمایش اندازه کرنل های گابری روی تصویر چهره با سه مقیاس و هشت چرخش.....(۵۴)

فصل چهارم:

- شکل (۱-۴) مدلی برای نرون.....(۵۶)

شکل (۲-۴) مدل نرون با تابع انتقال $\phi(\dots)$ (۵۸)

شکل (۳-۴) تابع انتقال Log-Sigmoid..... (۵۸)

شکل (۴-۴) تابع انتقال Tan-Sigmoid..... (۵۹)

شکل (۵-۴) تابع انتقال خطی..... (۵۹)

شکل (۶-۴) یک شبکه MLP با دو لایه پنهان..... (۶۰)

شکل (۷-۴) ساختار شبکه عصبی RBF..... (۶۱)

شکل (۴-۸-الف) تصویر اصلی و اجزای سه رنگش..... (۶۷)

شکل (۴-۸-ب) تصاویر حاصل از PCA روی تصویر مورد نظر..... (۶۷)

شکل (۴-۸-ج) مقایسه دو روش مهم در فشرده سازی تصاویر. سمت چپ تصویر حاصل از مجموع وزنی رنگ های R,G,B و سمت راست تصویر حاصل از پیاده سازی PCA..... (۶۸)

فصل پنجم:

شکل (۱-۵) نمونه هایی از تصاویر پایگاه داده FG_Net و چشم و دهان استخراج شده از طریق تبدیل موجک..... (۷۴)

شکل (۲-۵) نمونه ای از آن تصاویر پایگاه داده JAFEE..... (۷۵)

شکل (۳-۵) چند نمونه چشم، بینی و دهان های استخراج شده توسط موجک از پایگاه داده JAFEE در حالت شاد..... (۷۵)

شکل (۴-۵) نمونه ای از تصاویر مورد بررسی از پایگاه داده FG_Net..... (۷۶)

شکل (۵-۵-الف) تصویر پایه ی اول در PCA از پایگاه داده FG_Net..... (۷۷)

شکل (۵-۵-ب) تصویر پایه ی دوم در PCA از پایگاه داده FG_Net..... (۷۷)

شکل (۵-۵-ج) تصویر پایه ی سوم در PCA از پایگاه داده FG_Net..... (۷۷)

شکل (۵-۵-د) تصویر پایه ی چهارم در PCA از پایگاه داده FG_Net..... (۷۸)

شکل (۶-۵) بلوک دیاگرام سیستم تشخیص احساس با الگوریتم آنالیز مولفه های اساسی روی کل چهره..... (۷۸)

شکل (۷-۵) نمودار بازده کلاس بندی بر حسب تعداد پایه های PCA برای الگوریتم تشخیص توسط PCA روی کل چهره..... (۷۹)

شکل (۸-۵) نمودار بازده کلاس بندی بر حسب تعداد پایه های PCA برای الگوریتم تشخیص توسط PCA روی لب..... (۸۰)

شکل (۹-۵) بلوک دیاگرام سیستم تشخیص احساس با الگوریتم آنالیز مولفه های اساسی روی اجزای چهره..... (۸۰)

شکل (۱۰-۵) بلوک دیاگرام سیستم تشخیص احساس با الگوریتم آنالیز مولفه های اساسی و شبکه عصبی روی کل چهره..... (۸۱)

شکل (۱۱-۵) بلوک دیاگرام سیستم تشخیص احساس با الگوریتم آنالیز مولفه های اساسی و شبکه عصبی روی اجزای چهره..... (۸۳)

- شکل (۱۲-۵) نمودار بازده کلاسه بندی بر حسب تعداد پایه های PCA برای الگوریتم تشخیص توسط PCA دو پایه روی کل چهره.....(۸۴)
- شکل (۱۳-۵) تصاویر استفاده شده برای ساختن پایه ها برای کلاس غمگین.....(۸۴)
- شکل (۱۴-۵) تصاویر استفاده شده برای ساختن پایه ها برای کلاس شاد.....(۸۵)
- شکل (۱۵-۵) بلوک دیاگرام روش تشخیص احساس توسط آنالیز مولفه های اساسی با دو گروه پایه-معیار فاصله.....(۸۵)
- شکل (۱۶-۵) بلوک دیاگرام روش تشخیص احساس توسط آنالیز مولفه های اساسی با دو گروه پایه-شبکه عصبی.....(۸۷)
- شکل (۱۷-۵) مقایسه چشم افراد در حالت شاد و غمگین.....(۸۸)
- شکل (۱۸-۵) دو نمونه از تصاویر چهره و خروجی فیلتر گابر آن در دو حالت شاد و غمگین در به ترتیب چرخش ۳۶۰ و ۱۸۰ درجه و گام اول.....(۸۹)
- شکل (۱۹-۵) خروجی ۱۲ فیلتر گابر برای لب شاد.....(۸۹)
- شکل (۲۰-۵) خروجی ۱۲ فیلتر گابر برای لب غمگین.....(۹۰)
- شکل (۲۱-۵) خروجی ۱۲ فیلتر گابر برای چشم چپ شاد.....(۹۰)
- شکل (۲۲-۵) خروجی ۱۲ فیلتر گابر برای چشم چپ غمگین.....(۹۰)
- شکل (۲۳-۵) خروجی PCA برای یک لب شاد (نگاشت روی ۴۰ پایه) برای مقایسه.....(۹۲)
- شکل (۲۴-۵) خروجی PCA برای یک لب غمگین (نگاشت روی ۴۰ پایه) برای مقایسه.....(۹۲)
- شکل (۲۵-۵) خروجی PCA برای یک لب شاد و یک لب غمگین (نگاشت روی ۴۰ پایه) برای مقایسه.....(۹۳)
- شکل (۲۶-۵) خروجی PCA برای میانگین لب های شاد و لب های غمگین (نگاشت روی ۴۰ پایه) برای مقایسه.....(۹۳)
- شکل (۲۷-۵) نمایش همگرا شدن PCA با ۳۰ پایه برای لب.....(۹۴)
- شکل (۲۸-۵) نمایش همگرا شدن PCA با ۳۰ پایه برای چشم.....(۹۴)
- شکل (۲۹-۵) بلوک دیاگرام سیستم تشخیص احساس با الگوریتم فیلترهای گابر و آنالیز مولفه های اساسی و شبکه عصبی روی اجزای چهره.....(۹۵)
- شکل (۳۰-۵) اجزای برجسته استخراج شده یک تصویر غمگین با استفاده از روش نقاط برجسته موجک.....(۱۰۲)
- شکل (۳۱-۵) اجزای برجسته استخراج شده یک تصویر شاد با استفاده از روش نقاط برجسته موجک.....(۱۰۲)
- شکل (۳۲-۵) تبدیل ماتریسهای لب و چشم به بردار برای یک فرد غمگین.....(۱۰۳)
- شکل (۳۳-۵) تبدیل ماتریسهای لب و چشم به بردار برای یک فرد شاد.....(۱۰۳)

فهرست جداول

فصل دوم

- جدول (۱-۲) مدل اجزا چند حالتی چهره از دید روبرو.....(۱۵)
- جدول (۲-۲) تشخیص AU های پایین چهره.....(۱۷)
- جدول (۳-۲) مدل ابرو و گونه و چین و چروک.....(۱۹)
- جدول (۴-۲) تعریف پارامترهای بالای چهره.....(۲۲)
- جدول (۵-۲) استفاده از MPEG FAPها برای استفاده در سیستم فازی.....(۲۳)
- جدول (۶-۲) نتایج تشخیص سیستم تشخیص احساس فازی.....(۲۶)

فصل پنجم

- جدول (۱-۵): مقایسه نتایج بدست آمده از روشهای ۱-۳-۵ تا ۳-۳-۵.....(۹۶)
- جدول (۲-۵) مقایسه چشمها و لبها در دو فرد غمگین-غمگین.....(۹۹)
- جدول (۳-۵) مقایسه چشمها و لبها در دو فرد شاد-شاد.....(۱۰۰)
- جدول (۴-۵) مقایسه چشمها و لبهای دو فرد شاد-غمگین.....(۱۰۰)
- جدول (۵-۵) مقایسه چشمها و لبها در دو فرد شاد-عادی.....(۱۰۰)
- جدول (۶-۵) مقایسه چشمها و لبها در دو فرد غمگین-عادی.....(۱۰۰)
- جدول (۷-۵) نتایج نهایی پیاده سازی این الگوریتم بر روی پایگاه تصاویر.....(۱۰۳)

فصل اول

مقدمه

۱-۱- مقدمه

چهره انسان دارای ظاهر بسیار پیچیده و فوق العاده گویایی است، به طوری که یکی از راه های اصلی ارتباطی انسان ها از طریق چهره و حالات آن است. حالات چهره در ارتباطات انسانی نقش مهمی دارد و هم زمان با تعاملات اجتماعی رخ می دهد و به انتقال منظور و فهم مطلب کمک می کند. بر خلاف باور عمومی، تحقیقات روان شناسی و اجتماعی نشان داده است آنچه در مکالمات و ارتباطات انسانی مفهوم و منظور را می رساند بیشتر حالات چهره است نه کلمات گفته شده. محرابیان مشخص کرد بخشی زبانی از یک پیام، که از کلمات تشکیل شده است تنها ۷٪ پیام و بخش غیرزبانی و پارالینگوئیستیک^۱ که مربوط به لحن ادای جمله است، ۳۸٪ پیام را می رساند، در حالیکه حالات چهره ی گوینده ۵۵٪ پیام را منتقل می کند [1,3]. این درصد درجه اهمیت حالات چهره را در ارتباطات انسانی نشان می دهد و لازم است در سیستم های HCI^۲ در نظر گرفته شود.

به دلیل پیچیدگی آناتومی چهره انسان ساخت سیستمی که حالات و احساسات ظریف یک چهره را ثبت و ارائه دهد دشوار است. این مسئله با در نظر گرفتن حساسیت طبیعی انسان ها نسبت به حالات چهره بغرنج تر می شود. کوچکترین تغییر شکل غیر طبیعی در چهره ی فرد توسط مخاطب انسانی به سادگی قابل تشخیص است. با تحقیقات زیادی که در این زمینه انجام شده است تشخیص احساسات چهره همچنان یک زمینه ی تحقیقاتی باز به شمار می رود. هدف نهایی تحقیقات در این زمینه ساخت سیستمی است که با دقت بالایی احساسات شخص را تشخیص دهد، به صورت بلادرنگ عمل کند، تا حد ممکن خودکار باشد و نسبت به تغییرات فردی و شرایط محیطی مقاوم باشد.

۱-۲- کاربرد سیستم تشخیص احساس

¹ paralinguistic

² Human Computer Interface

تشخیص خودکار احساسات چهره در رنج وسیعی از کاربردها در دنیای صنعتی امروز حائز اهمیت است. کاربردهایی از قبیل لب خوانی، ویدئو کنفرانس، درک و شاخص تصاویر، مطالعات روانشناسی و مراقبت از بیماران افسردگی و اسکیزوفرنی، سنتز انیمیشن چهره، رباتیک، تشخیص موارد مشکوک در محافظت های امنیتی و ضد تروریستی، شناسایی افراد و تعداد زیادی کاربردهای HCI از موارد کاربرد سیستم تشخیص خودکار احساسات چهره اند [2,7]. در زمینه ی HCI توسعه ی سیستمی که بتواند حضور یک چهره را تشخیص دهد و شخص را با استفاده از چهره، صدا یا هردو شناسایی کند و در ادامه رفتار فرد را با استفاده از تشخیص احساسات صوت یا تشخیص احساس چهره یا هردو درک کند، مورد نظر است و تقابل انسان و کامپیوتر را در سیستم تعاملات اجتماعی بهبود و توسعه می دهد و ارتباط ماشین و انسان را نزدیک تر و مؤثرتر می سازد. سیستم های کامپیوتری که توانایی درک احساسات انسان را دارند رنج وسیعی از کاربردها را در زمینه های مختلف تحقیقاتی از قبیل کاربردهای امنیتی، اجرای قوانین، موارد پزشکی، تحصیلاتی، روان شناسی و ارتباطات تلفنی به خود اختصاص داده اند. این زمینه تحقیقاتی برای دانشمندان علوم کامپیوتر، مهندسين، روان شناسان و پزشکان اعصاب و روان و ماموران امنیتی و پلیس مطرح و مورد علاقه بوده است. تشخیص احساسات چهره نیازی اساسی در توسعه تقابلات انسان و کامپیوتر محسوب می شود و توضیح احساسات چهره توسط یک مدل ریاضی کار دشواری است.

۱-۳- مروری بر تاریخچه و روش ها

اگر به تاریخچه ی این مطالعات نظری بیفکنیم خواهیم دید از سی سال گذشته تحقیقات زیادی روی تشخیص احساسات انسان از طریق بررسی حالت چهره انجام شده است، که توسط پاول اکمن^۱ آغاز گردیده است. اکمن و گروهش روی تشخیص حالت انسان مطالعات جالبی انجام دادند [4,100,101]. آن ها حالت های مختلف انسان را به شش دسته ی کلی تقسیم کردند که شامل شادمانی، ناراحتی، عصبانیت، ترس، تعجب و تنفر است و حالات جهانی چهره نامیده می شود. اکمن و فریسن^۲ برای اولین بار در گسترش سیستم کدینگ چهره (FACS)^۳ کوشیدند که حالت های چهره را بر اساس یک سری واحد حرکت^۴ که مربوط به وضعیت برخی ماهیچه های چهره است، تشخیص می داد. احساسات چهره ناشی از تغییر ماهیچه های چهره به طور طبیعی است. FACS سیستمی برای کد کردن تغییرات کوچک و ظریف در صورت، شامل ۴۴ واحد حرکت است که مربوط به فعالیت ماهیچه های چهره اند و تمام تغییرات

¹ Paul Ekman

² Friesen

³ Facial Action Coding System

⁴ Action Unit

کوچک قابل تشخیص در چهره را به صورت زبانی توضیح می دهند. اکمن هم روی تصاویر ویدئویی و هم روی تک تصویر کار کرد که بعدها توسط دانشمندان دیگر ادامه یافت.

دانشمندان متوجه شدند که نوع و شدت حالات چهره در فرهنگ های مختلف برای بیان هر احساسی تا حدودی متفاوت است. مثلاً هنگام نمایش یک فیلم بیننده های ژاپنی در مقایسه با آمریکایی ها و اروپایی ها کمتر احساسات خود را از طریق چهره بروز می دهند. مطالعات روان شناسی ژانگ¹ در آسیا نشان می دهد که آسیایی ها در ابراز احساسات تنفر و ترس مشکلاتی دارند [5]. همچنین بروز احساسات در طیف های سنی مختلف، متفاوت است. بچه ها به طور گسترده تر و بدون هیچ قید و ملاحظه ای احساسات خود را نشان می دهند. هم چنین بحث های زیادی در روان شناسی و علم رفتارشناسی موجود است که آیا احساسات چهره جهانی است یا نه و اینکه آیا این حالات به صورت غیر ارادی رخ می دهد و یا ارادی. تحقیقات درباره ی جهانی بودن و ویژگی های فرهنگی، مدارکی ارائه می کند که با آنکه نمایش احساسات از طریق حالات چهره در فرهنگ های مختلف متفاوت است اما احساسات از طریق حالات چهره به طور صحیحی توسط افراد سراسر دنیا با نرخی بسیار بالاتر از آنچه فقط شانس به ما اجازه می دهد شناسایی می شوند و این نشان می دهد که احساسات چهره جهانی هستند و می توان آنها را به هفت گروه پایه ی شادمانی، ناراحتی، عصبانیت، نفرت، ترس، تعجب و حالت خنثی تقسیم بندی کرد [8,47,18]. اما در حالت عملی چهره ها حاوی ترکیب این احساسات هستند و مطمئن نیستیم بتوان حالات چهره را به صورت یکتا در یکی از این هفت کلاس مشخص قرار داد.

تا کنون تنها دو سیستم برای تشخیص خودکار حالات چهره به کار رفته است: سیستم کامپیوتر ویزن و سیستمی بر اساس الکترومیوگرافی. سیستم اول برای آموزش و استفاده وقت گیر است و برای استانداردسازی به خصوص در آزمایشگاه مشکلاتی دارد. سیستم دوم همان طور که در شکل 1-1 نشان داده شده است به جاگذاری سنسورهای روی صورت نیاز دارد که ممکن است فعالیت های خاص چهره را نشان دهد [6,8]. ما در اینجا سیستم نخست را مورد بررسی قرار می دهیم.



شکل (1-1) نمایش سیستمی بر اساس الکترومیوگرافی [6]

¹ Zhang

به طور کلی دو روش مختلف و رایج در تشخیص احساس بر اساس کامپیوتریون وجود دارد: روش های ویدئو-پایه¹ با استفاده از سکناس تصاویر و روش های استاتیک تصویر-پایه² با استفاده از تصاویر دو بعدی مجزا [9,10]. روش های مبتنی بر سکناس تصاویر معمولاً از آنالیز جریان نوری³ و ابزارهای تشخیص پترن برای تشخیص پترن های جریان نوری برای حالات خاص چهره استفاده می کنند [18]. این روش به چندین فریم از تصاویر برای تشخیص حالت احتیاج دارد و در کاربردهای زنده⁴ و غیر مقاوم در مقابل تغییرات، محدودیت هایی را شامل می شود. اما در کل می توان گفت روشهای ویدئو-پایه که از ویژگی های دینامیک استفاده می کنند نسبت به دسته دوم نتایج بهتری دارند چرا که ویژگی های دینامیک، ویژگی های بهینه ای هستند [11]. در روش های تصویر-پایه تشخیص احساس از طریق بررسی تصاویر دو بعدی انجام می شود و معمولاً از روش های ویژگی-پایه⁵ استفاده می کنند که نسبتاً عملکرد سریعی دارند اما درگیر مشکلاتی در زمینه توسعه ی روش های استخراج ویژگی اند تا بتوانند بدون توجه به تغییرات فردی و شرایط محیطی به درستی کار کنند. این روش ها از ویژگی های استاتیک سود می جویند و فرض می کنند که حالات چهره استاتیک اند و تشخیص به صورت فریم به فریم بدون استفاده از اطلاعات زمانی انجام می شود. اما این روش ها عملکرد خوبی در سیستم های عملی ندارد [9,12,38].

تمام روشهای به کار رفته برای تشخیص و آنالیز احساس چهره پایه و اصول مشابهی دارند. آنها در ابتدا چهره را تشخیص داده، از باقی تصویر جدا می کنند. سپس بعضی ویژگی ها را از تصویر یا ویدئو استخراج کرده و بعد این ویژگی ها را به عنوان ورودی های سیستم کلاس بندی به کار می برند. تفاوت روشها در انتخاب ویژگی ها، روش استخراج آنها و نوع کلاس بندی به کار رفته است. برای تشخیص چهره روش های بسیار زیادی به کار رفته اند و سیستم های موفق همان شبکه های عصبی بوده اند. بعد از تشخیص چهره، تصویر برای داشتن اندازه یکسان یا مقدار سیاه سفیدی⁶ یکسان نرمالایز می شود. سپس به یکی از مهمترین بخش ها در تشخیص احساس می رسیم که همانا انتخاب ویژگی و نحوه ی استخراج آنهاست که تنوع شدیدی در کارهای موجود می بینیم. روش های استفاده شده در این بخش را می توان به دو دسته ی اصلی روش های آنالیتیک⁷ و روش های هالستیک⁸ تقسیم کرد [40]. روش های هالستیک شامل روش های مدل-پایه⁹ [13] و تصویر-پایه¹⁰ [14] یا بر پایه تطابق الگوی کلی¹ اند که تصویر چهره را بدون

¹ Video based

² Image based

³ Optical Flow

⁴ Real time

⁵ Feature-based

⁶ Grey-scale

⁷ Analytic

⁸ Holistic

⁹ Model-based

¹⁰ Image-based

آنکه به بخش هایی تقسیم کنند به طور کامل آنالیز می کنند و یک الگو که می تواند یک پیکسل تصویر یا بردار ویژگی باشد که از بررسی کل تصویر چهره حاصل می شود، به دست می آید. در روش های تصویر پایه یک الگو از بررسی کل تصویر چهره حاصل می شود و در روش مدل پایه یک مدل آماری از مجموعه ی آموزش ساخته می شود و برای تشخیص احساس چهره به کار می رود. دسته ی دوم روش های آنالیتیک یا روش هندسی ویژگی-پایه [15] است که در آن مکان و شکل اجزا اصلی چهره مثل ابرو و چشم و بینی و دهان به عنوان بردار ویژگی حالت استخراج می شود. تعداد اجزا و نحوه ی استخراج آنها مستقیماً بر سرعت تشخیص اثر می گذارد. بنابراین روش مناسبی برای استخراج اجزا که ساده، دقیق و سریع باشد و نسبت به نویز، روشنایی، و فرد خاص تغییر ناپذیر بماند، مورد نیاز است.

در سیستم هندسی ویژگی-پایه اجزا اصلی چهره یا نقاط ویژگی استخراج می شوند و مسافت بین نقاط ویژگی برای تشکیل بردار ویژگی محاسبه می شود. نقاط ویژگی می توانند گراف هندسی ای برای نمایش چهره ایجاد کنند. در روش های تصویر-پایه برخی آنالیز های مکانی هالستیک و آنالیز های مکانی محلی استفاده می شود. در روش مدل-پایه مدل آماری از مجموعه ی آموزش ساخته می شود و برای تشخیص احساس چهره به کار می رود [16].

در یک مقایسه، روش های ویژگی-پایه از نظر محاسباتی نسبت به روش های الگو-پایه^۲ یا مدل-پایه گران تر و پیچیده ترند اما نسبت به تغییرات مقیاس، سایز، روشنایی، چرخش سر و مکان چهره مقاوم ترند. در ادامه ی این مقایسه روش های ظاهر-پایه^۳ یا تصویر-پایه نسبت به ویژگی-پایه^۴ بهترند زیرا گروه اول به تنظیم غیر دقیق تصویر و ناپیوستگی حرکت تصاویر حساس نیستند.

برای شناسایی احساس چندین روش تشخیص پترن در سیستم تشخیص احساس چهره جمع آوری شده است. روشهای آموزش ماشین از قبیل آدابوست^۵ و LDA^۶ و SVM^۷ برای شناسایی و آنالیز احساس چهره به کار رفته است. این روشها چندین نقص دارند: اکثر روشهای استخراج ویژگی ممکن است ویژگی ها را نتوانند به طور مقاوم و کامل، به خاطر آشفتگی ناشی از مو، شدت روشنایی و چین و چروک استخراج کنند. به طوری که تمام روشهای گفته شده تنها برای وضوح بالای تصویر قابل استفاده اند.

دسته ای از دانشمندان کوشیدند احساسات را از طریق بررسی تصاویر دینامیک کلاس بندی کنند. سوا^۸ تشخیص احساس را با دنبال کردن حرکت ۲۰ نقطه خاص آنالیز کرد [17]. میس^۱ از میانگین و واریانس

¹ Holistic template matching

² template-based

³ appearance-based

⁴ feature-based

⁵ Adaboost

⁶ Linear Discriminate Analysis

⁷ Support Vector Machine

⁸ Suwa

جریان نور^۱ در بلوک های کوچک تقسیم شده استفاده کرد[18]. یاکوب و دیویس^۲ از حرکت درون فریمی لبه های استخراج شده در ناحیه دهان، بینی، چشم و ابرو استفاده کردند[19]. بارتلت^۳ ترکیب جریان نوری و اجزاء اصلی به دست آمده از تفریق تصاویر را به کار برد[20]. اسا و پنتلند^۴ یک مدل پارامتری دینامیک توسط دنبال کردن حرکات چهره ساختند[21]. آنها برای تشخیص حالات چهره سیستمی براساس الگوهای مختلف جریان نوری ارائه کردند. آنها از الگوهای زمانی-مکانی برای تشخیص ۲ واحد حرکت چهره و ۳ وضعیت احساسی استفاده کردند. اما الگوها برای تشخیص واحدهای حرکت نامناسب بودند چرا که برای هر فرد و هر ترکیبی از واحدهای حرکت، نیاز به تعریف الگوی جدایی بود. یاکوب از جریان نوری برای محاسبه ی حرکت ویژگی های چهره برای دریافت احساس استفاده کرد[22]. ژانگ استفاده از ویژگی های هندسی را با مجموعه ی چند مقیاسه، چند چرخشه ی ویولت گابر در ۳۴ نقطه ی ثابت مقایسه کرد[5]. لیستی و رومل هارت^۱ یک شبکه عصبی برای جدا سازی حالت خنثی و خنده به کار بردند[23]. پاجت و کنترل^۶ نیز از شبکه ی عصبی سود جستند و روزنبلوم^۸ از یک شبکه ی عصبی RBF^۹ برای کلاس بندی ۶ حالت پایه و فلنز^{۱۰} از یک مدل پرسپترون چند لایه برای کلاس بندی حالت های خنثی و ناراحتی و شادی و عصبانیت استفاده کردند[24,25,26].

دسته ای از دانشمندان از طریق روشهای تصویر-پایه احساسات چهره را تشخیص دادند. سبه^{۱۱} مدل سه بعدی شبکه سیمی را برای دنبال کردن ویژگی های مختلف چهره به کار برد و مقایسه ای بین شبکه ی بیزین، درخت تصمیم، پرسپترون، SVM و KNN^{۱۲} برای کلاس بندی ۴ حالت روی پایگاه داده کوهن انجام داد و هم چنین تکنیک بگینگ و بوستینگ^{۱۳} را برای بهبود انتخاب ویژگی ها پیاده کرد[27,28]. بارتلت^۴ روشهای مختلف را مقایسه کرد و آدابوست، SVM و LDA را روی نمایش ویولت - گابر پیاده نمود[29]. اسا و پنتلند از پارامترهای دینامیک سه بعدی برای تشخیص ۵ حالت استفاده کردند[21]. ابود و کوییلنبرگ^{۱۵} از مدل های AAM^{۱۶} استفاده کردند[31,32,30]. گودتوک^۱ از ترکر با مدل سه بعدی برای

¹ Mase

² optical flow

³ Yacoob & davis

⁴ Bartlett

⁵ Essa & Pentland

⁶ Lisetti & Rumelhart

⁷ Padget & Cottrell

⁸ Rosenblum

⁹ Radial Basis Function

¹⁰ Fellenz

¹¹ Sebe

¹² K Nearest Neighborhood

¹³ Bagging & Boosting

¹⁴ Bartlett

¹⁵ Abboud & Kuilenburg

¹⁶ Active Appearance Model

برای دنبال کردن حرکت سر و چرخش استفاده کرد و سپس توسط SVM کلاسه بندی نمود[33].دوناتو^۱ برای تشخیص احساس نمایش ویولت گابر و آنالیز اجزاء مستقل و هانگ^۲ مدل توزیع نقطه ای را به کار برد[14].

برخی سیستم ها برای تشخیص حالات چهره توسط واحدهای حرکت از طریق FACS تلاش می کنند. تیان^۳ از کانتور لب، کانتور چشم، ویژگی های گونه و چروک ها و شبکه عصبی برای کلاس بندی گروهی از واحدهای حرکت استفاده کرد[34]. والاستار^۴ از الگوهای زمانی و کلاس بندی ترکیبی KNN و قانون- پایه^۵ سود جست[35]. اخیرا پنتیک و پاتراس^۶ از یک ترکر با فیلتر جزئی و قوانین زمانی برای کلاس بندی واحدهای حرکت از تصاویر غیر روبرو استفاده کرده اند[36]. مکنزی و کیمی^۷ سیستم خود تطبیقی گزارش کردند که اطلاعات چهره را به یک مجموعه واحدهای حرکت و بعد به یک مجموعه از برچسب های احساس تبدیل می کند.

دسته ی دیگری از دانشمندان کوشیدند احساسات را در تصاویر استاتیک کلاس بندی کنند. ترک و پنتلنت^۸ تصویر چهره را با مقادیر ویژه ی چهره ی به دست آمده از آنالیز خطی اجزاء اصلی نشان دادند[37]. پاژت و پاترل^۹ از روش مشابه ی مقادیر ویژه، اما با تفاوت بلوک های ۷ پیکسلی از نواحی ویژگی چشم ها و دهان استفاده کردند[38]. کنترل و متکالی^{۱۰} نمایش هالستیک را بر اساس اجزاء اصلی استخراج شده از شبکه های تغذیه ی مستقیم به کار بردند[39]. رهارجا^{۱۱} نیز از نمایش هالستیک با شبکه ی عصبی اما با نمایش تصاویر در ساختار هرمی استفاده کرد و لانیتیس^{۱۲} الگوهای پارامتری و منعطفی برای چهره به کار برد[40,41].

لیونز^{۱۳} در ۱۹۹۸ سیستم کدینگ احساسات چهره انسان بر اساس گابر ویولت را پیشنهاد کرد و نشان داد این روش نمایش درجه بالایی از همبستگی با اجزا اصلی چهره دارد[42,43]. ژانگ در ۱۹۹۸ نشان داد کدینگ حالات چهره بر اساس فیلتر بانک های گابر برای استخراج ویژگی ها و کلاس بندی آنها بر اساس

¹ Goktutk

² Donato

³ Haung

⁴ Tian

⁵ Valstar

⁶ Rule-based

⁷ Pantic & Patras

⁸ Keamy و Mckenzi

⁹ Turk & Pentlant

¹⁰ Padgett & Cottrell

¹¹ Cottrell & Metcalfe

¹² Rahardja

¹³ Lanitis

¹⁴ Lyons

پرسپترون چند لایه نتایج بهتری نسبت به تشخیص حالات چهره بر اساس روشهای تشخیص ویژگی هندسی دارد [5].

همان طور که دیدیم روش های بسیاری برای شناسایی احساسات چهره وجود دارد. از قبیل روشهای جریان نوری که نیاز به محاسبات پیچیده دارند و برای کاربردهای زنده مناسب نیستند و همچنین از فقدان مدل مناسب چهره برای نمایش احساسات رنج می برند. روشهای قانون-پایه که نسبت به خطاهای کوچک در اندازه گیری حساس اند. روش هایی که بر اساس تغییر شکل ماهیچه های صورت و واحدهای حرکت بیان می شوند که محاسبات پیچیده و وقت گیر و زیادی می طلبند. مشکل سیستم های مبتنی بر FACS ماهیت غیر کمی و وابسته به فردی آنهاست. چرا که یک حالت می تواند ناشی از فعالیت یک ماهیچه یا گروهی از ماهیچه ها باشد. توجه کنید تشخیص خودکار واحدهای حرکت کار دشواری است چرا که واحدهای حرکت تعریف کمی و مشخصی ندارند و می توانند در بیش از ۷۰۰۰ ترکیب پیچیده ظاهر شوند. برخی سیستم های دیگر از پارامترهای انیمیشن چهره^۱ برای کلاس بندی و سنتز استفاده می کنند که درگیر مشکلاتی مشابه سیستم های مبتنی بر FACS هستند.

۱-۴- پایگاه داده ها

مرحله ی بعد تهیه ی پایگاه داده یا همان دیتا بیس^۲ است. تهیه ی یک پایگاه داده ی خوب ملزم به داشتن تخصص، زمان و آموزش سیستم است. یکی از مهمترین کارهای ما این است که مجموعه ای بسازیم که در آن احساسات چهره پاسخی به حالت های احساسی باشد که شخص در لحظه در آن قرار می گیرد، نه حالت های مصنوعی صورت که از وقایع احساسی به دست نیامده و کاملاً غیر واقعی و ساختگی در مقابل دوربین است. مشکل آشکار در تهیه پایگاه داده گرفتن تصاویری با حالت های طبیعی است. مثلاً برای یک لبخند، معمولاً شخص هنگام لبخند زدن جلوی دوربین واقعا احساس شادمانی نمی کند و تصویر به دست آمده با یک لبخند حقیقی که از فیلم برش داده می شود، بسیار متفاوت است. در کل احساسات تعریف شده ی کمی وجود دارد که در این میان تنها تعداد کمی از آنها را می توان در محیط آزمایشگاه شبیه سازی کرد. تشخیص پز که در مقابل تشخیص حالت قرار دارد به بررسی حالت های غیر واقعی و مصنوعی می پردازد.

پایگاه داده های معروف در زمینه تشخیص احساس پایگاه داده احساس Cohn-Canade [28]، پایگاه داده غیر رایگان POFA از اکمان^۳ شامل ۱۱۰ تصویر سیاه سفید، پایگاه داده احساس چهره CMU AMP،

^۱ Facial Animation Parameter

^۲ Data base

^۳ Ekman

پایگاه داده تشخیص احساس و ژست^۱ FG_Net و پایگاه داده^۲ JAFEE است. پایگاه داده رایگان احساس چهره CMU AMP شامل ۷۵ تصویر از ۱۳ فرد است که شش احساس مختلف را نشان داده اند. پایگاه داده احساس غیر رایگان Cohn-Canade شامل ۳۴۰ سکانس تصویر از هر یک از ۴۸ نفری است که در تهیه پایگاه داده کمک کرده اند. پایگاه داده تشخیص احساس و ژست FG_Net ابتدا در سال ۲۰۰۲ توسط پروفیسور فرانک والهوف^۳ از دانشگاه مونیخ تهیه شده است و شش حالت مختلف تعریف شده توسط اکمن و فریسن^۴ را نشان می دهد و در تحقیقات انجمن تحقیقات اروپا روی تشخیص چهره و احساس گسترش یافته است. در این پایگاه داده تلاش شده است که تصاویری که از افراد تهیه می شود در زمانی باشد که احساسات واقعی و نه ساختگی نشان می دهند. به این منظور تصاویری به صورت متوالی و سکانس تصاویر از فرد هنگامی که با تماشای کلیپ های ویدئویی در موقعیت احساس مورد نظر قرار گرفته است تهیه شده است و شامل چرخش محدود سر در تمام جهات است و ۷ احساس جهانی را نشان می دهد. پایگاه داده معروف دیگر پایگاه داده JAFEE است که شامل ۲۱۳ تصویر ۲۵۶ در ۲۵۶ پیکسل از احساسات چهره زنان ژاپنی است که توسط کاماچی و گیوبا^۵ از دانشگاه کیوشو^۶ تهیه شده است. برای تهیه این پایگاه داده از ده نفر خواسته شده است که در مقابل دوربین احساسات مختلفی از خود نشان دهند و وقتی آماده بودند از خود عکس بگیرند. تصاویر اصلی جوری تغییر مقیاس داده شده و برش خورده که چشم افراد تقریباً در یک مکان و با فاصله ۶۰ پیکسل از کنار عکس قرار بگیرد و از هر فرد حدود ۲۰ تصویر از هفت احساس مختلف تهیه شد. ما در این رساله از پایگاه داده FG_Net و JAFEE استفاده کرده ایم.

۱-۵- ساختار پایان نامه

در فصل اول به معرفی مختصر سیستم تشخیص احساس، بیان اهمیت و کاربرد و شرح خلاصه ی روش های موجود پرداختیم. در فصل دوم به توضیح مختصر چندین روش پایه و پرکاربرد در این زمینه می پردازیم و نمونه هایی از هر سه روش ویژگی-پایه، مدل-پایه، و تصویر-پایه ارائه می کنیم. در فصل سوم به معرفی اجمالی تبدیل موجک و فیلترهای گابر پرداخته، در فصل چهارم شبکه های عصبی و تحلیل مولفه های اساسی^۷ (PCA) را بسیار خلاصه توضیح می دهیم. در فصل پنجم به توضیح و بررسی نتایج حاصل از روش خود پرداخته که در ابتدا از طریق به دست آوردن نقاط برجسته به کمک الگوی

^۱ Face and Gesture Recognition Research Network

^۲ Japanese Female Expression

^۳ Frank Wallhof

^۴ Friesenkp

^۵ Kamachi و Gyoba

^۶ Kyushu

^۷ Principle Component Analysis

پیشنهادی از طرف سبه و لوپاس¹ بر اساس تبدیل موجک نواحی اصلی چهره یعنی چشم و دهان را جدا می‌کنیم [44,45]، سپس برای استخراج ویژگی یکبار از تحلیل مولفه‌های اساسی و بار دیگر از تلفیق تحلیل مولفه‌های اساسی و فیلترهای گابر استفاده می‌کنیم که روی نواحی استخراج شده و کل چهره اعمال می‌شود تا نتایج را بتوانیم در مراحل بعد با هم مقایسه کنیم. برای کلاس بندی کردن ویژگی‌های استخراج شده نیز از معیار فاصله و شبکه‌های عصبی پرسپترون چند لایه سود می‌جوییم و بعد از آن نیز یک روش پیشنهادی مبتنی بر اطلاعات آماری و شبکه عصبی را پیشنهاد می‌کنیم. و در نهایت در فصل ششم به نتیجه گیری و بیان راه کارها و پیشنهاداتی برای ادامه ی این پروژه خواهیم پرداخت.

¹ Loupias , Sebe

فصل دوم

مروری بر روش های پر کاربرد تشخیص احساس چهره

۲-۱-۱-مقدمه

در این فصل به توضیح مختصر چندین روش پایه و پرکاربرد در زمینه تشخیص احساس چهره از طریق تحلیل تصاویر چهره می پردازیم و نمونه هایی از هر سه روش الگو-پایه و مدل-پایه و تصویر-پایه ارائه می کنیم.

۲-۲-تشخیص احساس چهره از طریق روش ویژگی-پایه ی بررسی واحد های حرکتی

از اولین و مهمترین روش های تشخیص احساس افراد از طریق بررسی حالات چهره استفاده از سیستم¹ FACS و واحد های حرکتی² است که هنوز هم با وجود پیچیدگی های آن، به علت دقت بالا و انعطاف و تعمیم پذیری سیستم به عنوان روشی مطرح در میان دانشمندان و محققان محبوبیت دارد. این سیستم را اکمن و گروهش به عنوان اولین افرادی که در این زمینه به فعالیت علمی و مدون پرداختند ابداع کردند و گسترش دادند. در بخش ۲-۲-۱ و ۲-۲-۲ به شرح مختصری از این سیستم می پردازیم [46,4].

۲-۲-۱-تشخیص واحد های حرکتی پایین چهره برای آنالیز احساس چهره

۲-۲-۱-۱-معرفی سیستم

در این روش بر اساس مدل های چند حالته، ویژگی های چهره از قبیل چشم، ابرو، دهان، چانه و چین و چروک های مربوط به آنها، تشخیص داده و دنبال می شود. برای کلاس بندی معمولاً توسط الگوریتم شبکه های عصبی یازده AU³ با نرخ تشخیص ۹۶٪ قابل تشخیص است. نتایج نشان می دهد این سیستم قادر است AUها را چه به صورت تکی چه به صورت ترکیبی شناسایی کند.

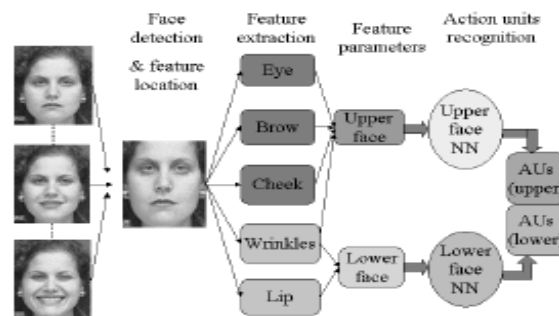
¹ Face Action Units Coding System

² Action Unit

³ Action Unit

در این روش از FACS که شامل ۴۹ واحد یا AU که مربوط به مناطق چشم ها ، دهان و پیشانی است، استفاده می شود که سی AU مربوط به آناتومیک ماهیچه های چهره اند. گرچه تعداد واحدها محدود هستند اما بیشتر از هفت هزار ترکیب از آنها مشاهده شده است. تشخیص خودکار AUها مسئله ی دشواری است چرا که تعریف مشخصی از آنها و ترکیباتشان وجود ندارد.

روش هایی از جریان نوری، سیستم هیبریدی و سیستم جریان چگالی استفاده می کنند و روشی نیز بر اساس آنالیز لبه های استخراج شده از چهره وجود دارد [48,49] ولی بازدهی سیستم تشخیص احساس از طریق سیستم FACS بیشتر است. شکل ۲- ۱ یک نمای کلی از این سیستم را نشان می دهد. ابتدا چرخش سر و مکان چهره مشخص می شود. سپس تغییرات ظریف در اجزاء چهره اندازه گیری می گردد. مطابق با FACS ، این تغییرات به عنوان جمعی از پارامترهای ویژگی نمایش داده می شود. سپس این AUها توسط شبکه ی عصبی کلاس بندی می گردند [83,84].



شکل (۱-۲) سیستم feature-based تشخیص AUها

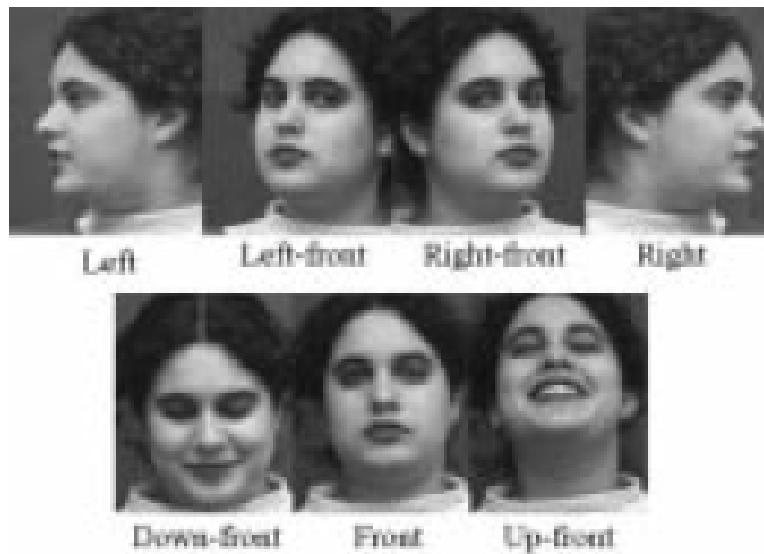
برای تشخیص تغییرات در احساسات چهره مدل چند حالت را گسترش می دهند که ویژگی های چهره را دنبال می کند. چرخش های مختلف سر و اجزاء مختلف چهره که حالت های مختلف را نشان می دهند تعریف می شوند. برای هر وضعیت سر و هر جزء چهره حالت های مختلف وجود دارد. برای هر حالت مختلف روش استخراج خاصی مورد نیاز است. براساس توضیح AUها ویژگی های چهره به دو گروه ویژگی های بالا و پایین چهره تقسیم می شوند که در اینجا تشخیص گروه پایین را ابتدا توضیح می دهیم، چرا که لب ها از چشم ها بیشتر تغییر وضعیت می دهند و در تشخیص مهم و پرکاربردترند. ۹ پارامتر برای توضیح شکل لب ها ، حرکت لب ها و حالت لب ها و چین و چروک هایش به کار می رود. بعد از استخراج صحیح ویژگیها، سیزده AU پایین چهره و ترکیب آنها توسط شبکه عصبی با سرعت تشخیص ۹۶٪ قابل تشخیص است. برخی دانشمندان پس از معرفی و تشخیص AUهای پایین چهره، از HMM^۱ برای کلاس بندی احساسات استفاده کردند [49]. در مقایسه با سیستم قبلی، سیستم فعلی دقت

^۱ Hidden Markov Model

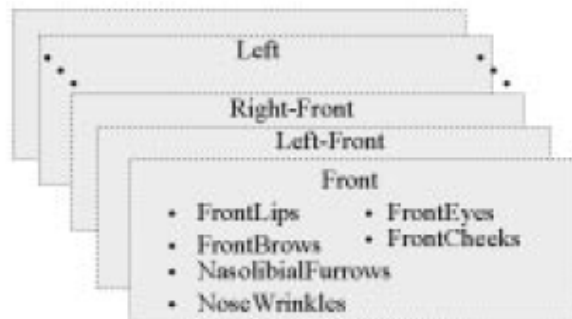
بیشتر و سرعت تشخیص بالاتری دارد و هم چنین می تواند بدون توجه به AUها که مجزا یا ترکیبی اند کار کند.

۲-۲-۱-۲-۲-مدل های چند حالتی برای اجزا چهره

چرخش سر روی زاویه ی نمای چهره اثر می گذارد. ۷ حالت چرخش سر در شکل ۲-۲ تعریف شده است. در شکل ۳-۲ اجزاء مختلف چهره دید از روبرو را نشان می دهد.



شکل (۲-۲) حالت های چرخش سر



شکل (۳-۲) اجزاء مختلف چهره دید از روبرو

۲-۲-۱-۳-۲-مدل های چند حالتی لب

مدل های مختلف اجزاء چهره برای حالت های مختلف سر استفاده می شوند. مثلاً مدل لب از جلو با مدل لب از نیمرخ فرق می کند. بر اساس نمای مختلف اجزاء چهره، مدل هندسی خاصی برای مدل کردن مکان و شکل هر جز داریم که در جدول ۱-۲ جزئیات مدل های اجزاء پایین چهره شامل اجزاء دائمی

مثل لب ها و اجزاء گذرا مثل چین و چروک ها در دید از روبرو را آورده ایم. هر کدام از اجزاء شامل چندین مدل حالت اند. مثلا لب سه مدل حالت باز، بسته و شدیداً بسته دارد و برای چین و چروک ها، مدل های حاضر و غایب داریم. الگوهای مختلفی برای کانتور و طرح لب به کار می رود. دو کمان برای مدل کردن مکان و چرخش و شکل لب ها استفاده می شود. الگوی لب های باز و بسته ۶ پارامتر $x_c, y_c, \omega, h_1, h_2, \theta$ دارد.

۲-۲-۱-۴- استخراج ویژگی های پایین چهره

جهت و بزرگی حرکت سطح پوست باعث ایجاد ویژگی های دائمی و گذرای چهره می شود. ویژگی های دائمی عبارت اند از چشم ها، لب و تمام چروک های پوست که با افزایش سن ایجاد شده اند. ویژگی های گذرا چروک هایی است که در حالت خنثی در چهره وجود ندارد. بعد از مقداردهی اولیه ی الگوهای ویژگی های دائمی چهره در فریم اول با حالت خنثی، ویژگی های دائمی و گذرا می توانند دنبال شوند. نتایج نشان می دهد این روش برای دنبال کردن ویژگی های چهره، حتی وقتی چرخش سر زیادی داشته باشیم روشی مقاوم است.

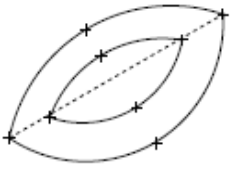
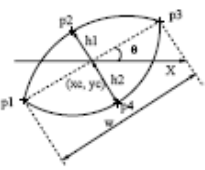

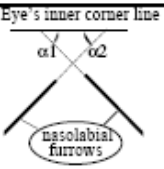
احساسات در چهره از ویژگی های گذرا ناشی می شود. چین و چروک ها و خط ها به خاطر وضعیت ماهیچه ها ایجاد می شوند. ویژگی های گذرا اطلاعات قاطعی برای تشخیص AUها فراهم می کنند. مثلا خط عمودی بین ابروها AU4 و خط افقی روی پیشانی AU1 را می سازد. بعضی از خطوط گذرا می تواند با بالا رفتن سن به خطوط دائمی تبدیل شود. حضور و غیاب چروک ها در تصویر می تواند با آنالیز هندسی ویژگی ها و آنالیز مقادیر ویژه مشخص شود. لوبو و ون^۱ چروک ها را توسط تابع مار^۲ تشخیص دادند تا افراد را به گروه های سنی مختلف تقسیم بندی کنند [50]. لین^۳ تمام لبه های افقی و عمودی و قطری را برای تشخیص احساس مشخص کرد [49].

جدول (۱-۲) مدل اجزا چند حالت چهره از دید روبرو

¹ Kwon و Lobo

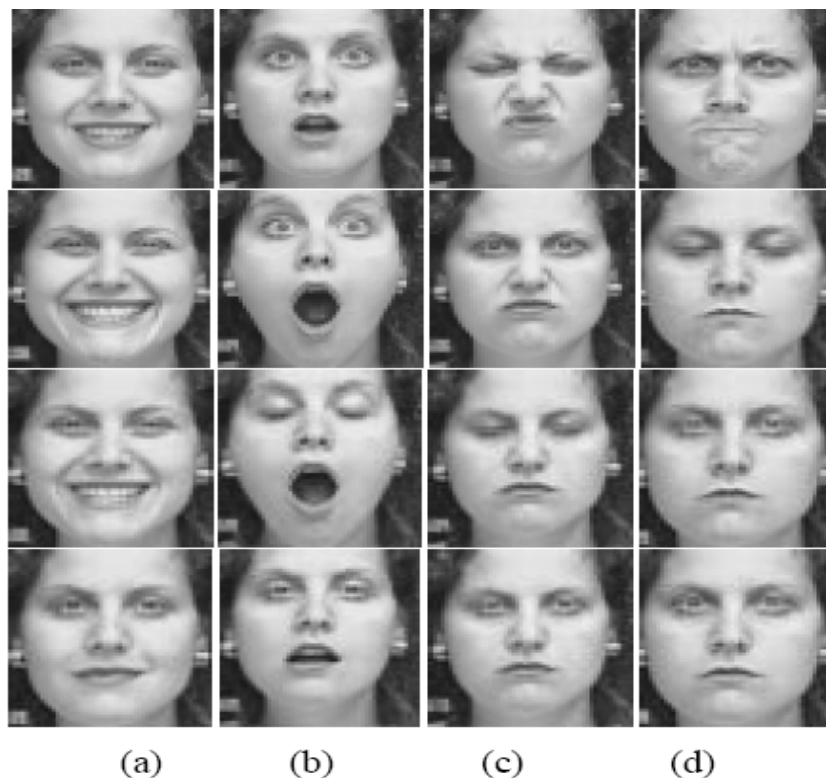
² Snake

³ Lien

Component	State	Description/Feature
Lip	Open	
	Closed	
	Tightly closed	
Furrow	Present	
	Absent	

سه گروه چروک مشخص می شوند: چروک های ناسلابیال^۱، چین های بینی، چین های کراس فیت^۲ که دو حالت حاضر و غایب دارند. در مقایسه با حالت خنثی چروک ها عمیق تر، طولانی تر یا با حضور هستند. حالت حاضر چروک بینی یک مربع بین دو گوشه ی داخلی چشم است. ناحیه چین های کراس فیت کنار گوشه های خارجی چشم و چروک ناسلابیال منطقه ی بین خط گوشه های داخلی چشم و خط گوشه های لب است. تغییر ویژگی های گذرا در حالات مختلف را در شکل ۲-۴ می بینید.

¹ nasolabial
² crows feet



شکل (۲-۴) ویژگی های گذرا برای حالات مختلف (a) چشم ها کشیده و لبخند (b) چشم ها و دهان باز (c) چشم ها بسته یا کشیده (d) دهان به شدت بسته

برای تشخیص اطلاعات لبه معمولاً از تشخیص لبه ی کنی^۱ استفاده می شود. برای چین بینی و چین کراس فیت، پیسکل لبه ی شماره ی E از فریم جاری را با پیسکل لبه ی E0 از فریم اول با حالت خنثی مقایسه می کنند. اگر E/E_0 بزرگتر از ترشلد مشخصی بود چین ها حضور دارند، در غیر این صورت آن ها غایب اند. برای چروک های ناسلابیال، لبه های قطری مثل شکل ۲-۵ مشخص می شود.

۲-۱-۲-۵-نمایش ویژگی های پایین چهره

از ۹ پارامتر که برای نمایش ویژگی های پایین چهره تعریف می شود ۶ پارامتر ویژگی های ثابت شکل لب، حالت لب و حرکت لب را نشان می دهد و ۳ پارامتر ویژگی های گذرا از چین و چروک بینی و ناسلابیال را مشخص می سازند. به خاطر آنکه گوشه های داخلی چشم ویژگی های بسیار پایداری در چهره اند و نسب به تغییر حالت چهره غیر حساس اند محور X خط بین دو گوشه ی داخلی چشم و محور Y عمود بر محور X تعریف می شود. تمام پارامتر های حرکت لب و چین های ناسلابیال در این دستگاه مختصات سنجیده می شوند و با حالت خنثی مقایسه می گردند.

¹ Canny

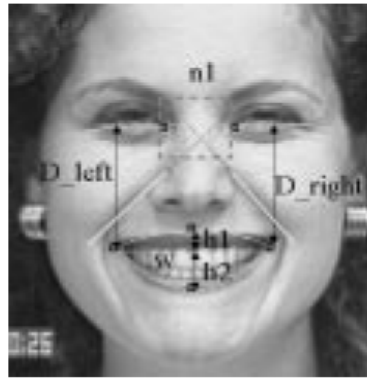


شکل (۲-۵) چروکهای ناسلابیال

برای حل مشکل تغییر سایز تصاویر نشان دهنده ی چهره، از نسبت پارامترها به جای استفاده ی مستقیم از آن ها استفاده می شود. تعریف پارامترهای پایین چهره در لیست جدول ۲-۲ آورده شده است. این اطلاعات به کمک خط واصل گوشه های داخلی دو چشم و نرمالایز کردن اطلاعات هر فرد حاصل می شود. معنای فیزیکی این پارامترها در شکل ۲-۶ مشخص شده است.

جدول (۲-۲) تشخیص AU های پایین چهره

Permanent features		
Lip height (r_{height})	Lip width (r_{width})	Left lip corner motion (r_{left})
$r_{height} = \frac{(h1+h2) - (h1_0+h2_0)}{(h1_0+h2_0)}$ If $r_{height} > 0$, lip height increases.	$r_{width} = \frac{w - w_0}{w_0}$ If $r_{width} > 0$, lip width increases.	$r_{left} = -\frac{D_{left} - D_{left0}}{D_{left0}}$ If $r_{left} > 0$, left lip corner move up.
Right lip corner (r_{right})	Top lip motion (r_{top})	Bottom lip motion (r_{btm})
$r_{right} = -\frac{D_{right} - D_{right0}}{D_{right0}}$ If $r_{right} > 0$, right lip corner move up.	$r_{top} = -\frac{D_{top} - D_{top0}}{D_{top0}}$ If $r_{top} > 0$, top lip move up.	$r_{btm} = -\frac{D_{btm} - D_{btm0}}{D_{btm0}}$ If $r_{btm} > 0$, bottom lip move up.
Transient features		
Left nasolabial furrow angle ($Angle_{left}$)	Right nasolabial furrow angle ($Angle_{right}$)	State of nose wrinkles (S_{nosew})
Left nasolabial furrow present with angle $Angle_{left}$.	Left nasolabial furrow present with angle $Angle_{right}$.	If $S_{nosew} = 1$, nose wrinkles present.



شکل (۶-۲) نمایش ویژگی های پایین چهره

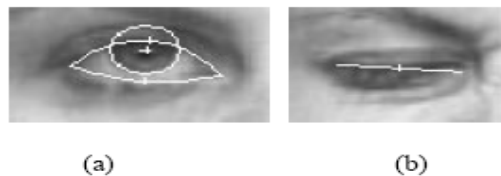
۲-۲-۲-۲- تشخیص واحد های حرکتی بالای چهره برای آنالیز احساس چهره

۲-۲-۲-۲-۱- معرفی سیستم

در اینجا به معرفی سیستم اتوماتیکی برای آنالیز دقیق تغییرات در حالات بالای چهره بر اساس دو دسته پارامتر ثابت و گذرا در دنباله تصاویر دید از روبرو می پردازیم. سیستم بر اساس FACS و AU های بالای چهره بوده و از مدل های اجزا چند حالتی برای دنبال کردن و مدل کردن ویژگی های مختلف چهره شامل چشم و ابرو و گونه و چروک ها استفاده می کند.

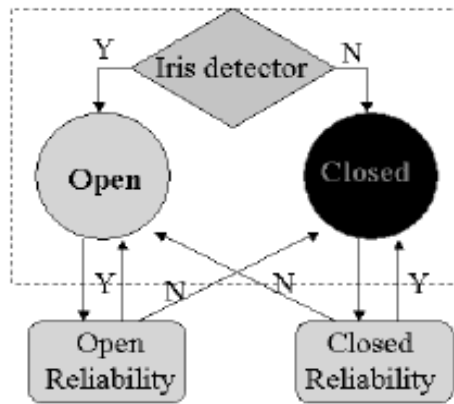
۲-۲-۲-۲-۲- مدل چشم دو حالتی

شکل ۷-۲ مدل چشم دو حالتی را نشان می دهد. با اطلاعات عنبیه دو حالت باز و بسته برای چشم قابل تشخیص است. برای حالت های مختلف، الگوهای خاص چشم و الگوریتم های گوناگون برای به دست آوردن ویژگی های چشم به کار می رود. برای چشم باز کانتور چشم را به صورت متقارن نسبت به خط عمود بر خط واصل دو گوشه ی چشم در نظر می گیرند. شکل ۷-۲ (d) از یک دایره با ۳ پارامتر (x_0, y_0, r) و ۲ کمان با ۶ پارامتر $(x_c, y_c, \omega, h_1, h_2, \theta)$ تشکیل شده است. برای چشم بسته مدل به ۴ پارامتر برای هر گوشه چشم کاهش می یابد.

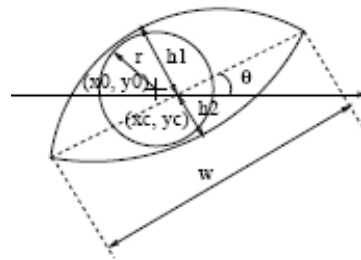


(a)

(b)



(c)



(d)



(e)

شکل (۷-۲) شکل مدل چشم دو حالت. (a) چشم باز. (b) چشم بسته. (c) دیاگرام حالت گذرا. (d) مدل پارامتری چشم باز. (e) مدل پارامتری چشم بسته

۲-۲-۳-مدل ابرو و چین و چروک

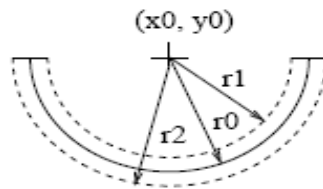
مدل ابرو و گونه و چین و چروک در جدول ۲-۳ آورده شده است. برای ابرو و گونه از مدل های تک حالتی مجزا استفاده شده است. یک مدل مثلث با ۶ پارامتر $P_1(x_1, y_1)$, $P_2(x_2, y_2)$ و $P_3(x_3, y_3)$ و چین و چروک ها با دو حالت حاضر و غایب وجود دارد.

جدول (۲-۳) مدل ابرو و گونه و چین و چروک

Component	State	Description/Feature
Brow	Present	
Cheek	Present	
Furrow	Present	
	Absent	

۲-۲-۲-۴-ویژگی های چشم

مدل دوگانه برای تشخیص چشم به کار می رود اما حالت پایه را چشم باز در نظر می گیرند. با توجه به الگوی چشم باز در فریم اول گوشه ی داخلی چشم به دقت با دنبال کردن نقطه ی ویژگی پیدا می شود. مشخص می شود که گوشه خارجی نسبت به گوشه داخلی ناپایدارتر و در ترک سخت تر است. پس فرض می شود گوشه های خارجی روی خط واصل گوشه های داخلی باشند و می توانند به کمک پهنای چشم که از فریم اول محاسبه می شود به دست آیند. عنبیه اطلاعات مهمی درباره چشم به دست می دهد. اطلاعات چگالی و لبه در تشخیص عنبیه به کار می رود. یک ماسک عنبیه نیم دایره برای به دست آوردن لبه های صحیح آن استفاده می شود. بعد از تشخیص عنبیه مرکز آن، مطابق شکل ۸-۲ مرکز ماسک را می سازد.



شکل (۸-۲) ماسک نیم دایره عنبیه

در یک دنباله تصاویر کانتور های پلک برای چشم باز با دنبال کردن نقاط ویژگی حاصل می شود. برای چشم بسته این کار شدنی نیست. خط واصل بین گوشه های داخلی و خارجی چشم به عنوان مرز شناخته می شود. برخی نتایج دنبال کردن چشم در شکل ۹-۲ نشان داده می شود.



(a)

(b)

شکل (۹-۲) نتایج دنبال کردن در (a) چشم نیمه باز و (b) چشم کاملا باز به همراه پلک زدن

۲-۲-۵-ویژگی های ابرو و گونه

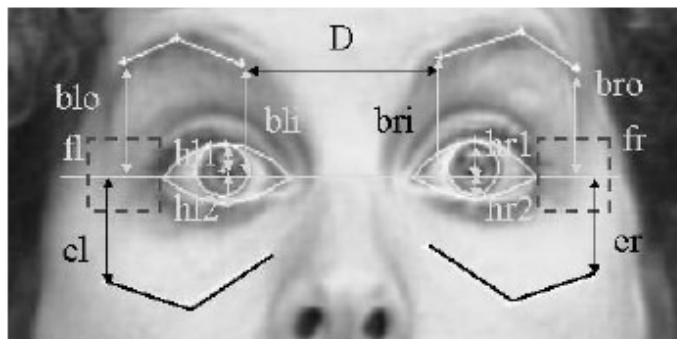
برای مدل کردن ابرو و گونه الگوهای مثلثی جدا گانه ای با ۶ پارامتر (x_1, y_1) , (x_2, y_2) و (x_3, y_3) استفاده شده است. ابرو و گونه توسط الگوریتم گرادین پیدای می شوند. شکل ۲-۹ شامل برخی نتایج ابرو برای حالت‌های مختلف نیز است.

۲-۲-۶-ویژگی های گذرا

احساسات به کمک ویژگی‌های گذرا قابل شناسایی است. این ویژگی‌ها اطلاعات قاطعی برای تشخیص AUها ایجاد می کند. بعد از یافتن گوشه های خارجی چشم، تشخیص گر لبه برای چروکهای کراس فیت شروع به جستجو می کند. تعداد پیکسل‌های لبه از فریم فعلی با تعداد پیکسل‌های لبه از فریم خنثی مقایسه می شود، اگر این نسبت از مقدار ترشلد بزرگتر بود چروکها حاضرند.

۲-۲-۷-نمایش ویژگی بالای چهره

برای تشخیص تغییرات دقیق احساسات، ویژگی بالای چهره را در ۱۵ پارامتر نمایش می دهند. ۱۲ پارامتر حرکت و شکل چشم، ابرو و گونه ها را توضیح می دهد. ۲ پارامتر حالت چروکهای کراس فیت و یک پارامتر فاصله ی بین ابروها را نشان می دهد.



شکل (۲-۱۰) ویژگی های بالای چهره

شکل ۲-۱۰ سیستم مختصات و تعریف پارامترها را نشان می دهد. به منظور خنثی سازی اثر اندازه های مختلف تصاویر چهره، تمام پارامترها به جز پارامترهای چروک فورو^۱، با تقسیم بر فاصله بین هر ویژگی و خط واصل بین دو گوشه داخلی چشم در فریم خنثی نرمالایز می شود. در جدول ۲-۴ تعریف پارامترهای بالای چهره را می بینیم.

جدول (۲-۴) تعریف پارامترهای بالای چهره

¹ furrow

Permanent features (Left and right)		
Inner brow motion (r_{binner})	Outer brow motion (r_{bouter})	Eye height ($r_{cheight}$)
$r_{binner} = \frac{bi - bi_0}{bi_0}$ If $r_{binner} > 0$, Inner brow move up.	$r_{bouter} = \frac{bo - bo_0}{bo_0}$ If $r_{bouter} > 0$, Outer brow move up.	$r_{cheight} = \frac{(h1+h2) - (h1_0+h2_0)}{(h1_0+h2_0)}$ If $r_{cheight} > 0$, Eye height increases.
Eye top lid motion (r_{top})	Eye bottom lid motion (r_{btm})	Cheek motion (r_{cheek})
$r_{top} = \frac{h1 - h1_0}{h1_0}$ If $r_{top} > 0$, Eye top lid move up.	$r_{btm} = \frac{h2 - h2_0}{h2_0}$ If $r_{btm} > 0$, Eye bottom lid move up.	$r_{cheek} = \frac{c - c_0}{c_0}$ If $r_{cheek} > 0$, Cheek move up.

معمولا از شبکه عصبی سه لایه با یک لایه مخفی برای تشخیص AUها استفاده می شود. هر خروجی یک تخمین از احتمال حضور AU در ورودی را ارائه می دهد.

۲-۳- تشخیص احساس چهره از طریق روش ویژگی-پایه ی مبتنی بر MPEG4 و کلاسیفایر

فازی

۲-۳-۱- معرفی سیستم

در اینجا روشی براساس اصول منطق فازی برای تشخیص احساسات چهره بررسی می شود که از تکنیک پردازش تصویر برای تشخیص «پارامترهای انیمیشن چهره MPEG4» از ورودی ویدئو استفاده می کند [51].

مشخص شد که تلاش برای تشخیص احساسات چهره در HCI¹ از دو مشکل اساسی رنج می برد. اول اینکه دشوار است یک مدل دقیق و مشخص برای نمایش حالات چهره ابداع شود. دوم اینکه تکنیک های پردازش تصویر برای استخراج تغییرات چهره همیشه اطلاعات دقیقی به ما نمی دهند. با توجه به این مشکلات، سیستمی براساس منطق فازی برای پردازش اطلاعات نادقیق به خوبی جواب می دهد و در سال های اخیر گسترش روش های فازی برای بررسی حالات چهره به وضوح به چشم می خورد. این سیستم ها براساس قوانین فازی مربوط به پارامترهای تعریف کننده ی حالات مختلف کار می کند [85]. سیستم کنونی پارامترها را در بخش هایی از MPEG4³ FAPs تعریف می کند. ویدئوی ورودی برای

¹ Moving Pictures Experts Group

² Human Computer Interface

³ Facial Animation Parameter

استخراج یک بردار FAP برای هر فریم پردازش می شود. قانون فازی بردار FAP را چک می کند تا شبیه ترین حالت نشان داده شده در فریم را پیدا کند.

۲-۳-۲- چارچوب MPEG4 برای حالات چهره

MPEG4 مجموعه ای از نقاط ویژگی و FAP پاسخ را فراهم می کند. برای توضیح حرکات چهره، MPEG4 از FAPها استفاده می کند که FAPها حرکات نقاط ویژگی یا FPهای مربوط به چهره اند. در واقع FAPها حرکت نرمال نقاط ویژگی را در جهت تعریف شده تعریف می کند. تمام FAPها شامل حرکات انتقالی اند که در بخش هایی به نام «واحدهای FAP» بیان می شود، که این واحدها به بخشهایی از مسافت بین ویژگی های کلیدی چهره پاسخ می دهد. در اینجا هر احساس به عنوان مجموعه ای از MPEG4 FAPs در نظر گرفته می شود. سیستم با استخراج AUهای چهره در فرم MPEG4 FAPs کار می کند. سیستم FAPها را با تعریف مجموعه های فازی برای هر FAP محدود می کند و تنها زیرمجموعه ای از نقاط ویژگی را که برای دنبال کردن هر حس کافی هستند در نظر می گیرد. جدول ۲-۵ FAPهای استفاده شده در این سیستم را نشان می دهد.

جدول (۲-۵) استفاده از MPEG FAPها برای استفاده در سیستم فازی

FAP Name	MPEG-4 No.	FAP Description
lower-t-midlip	4	Vertical displacement of upper midlip
raise-b-midlip	5	Vertical displacement of lower midlip
stretch-l-cornerlip	6	Horizontal displacement of left inner lip corner
stretch-r-cornerlip	7	Horizontal displacement of right inner lip corner
raise-l-cornerlip	12	Vertical displacement of left lip corner
raise-r-cornerlip	13	Vertical displacement of right lip corner
close-t-l-eyelid	19	Closing of top left eyelid
close-t-r-eyelid	20	Closing of top right eyelid
raise-l-i-eyebrow	31	Vertical displacement of left inner eyebrow
raise-r-i-eyebrow	32	Vertical displacement of right inner eyebrow
raise-l-m-eyebrow	33	Vertical displacement of left middle eyebrow

۲-۳-۳-سیستم فازی برای کلاس بندی احساسات چهره

سیستم های فازی به نحو موفقیت آمیزی در مدل کردن تعداد زیادی پدیده های زنده و آنی^۱ بدون قطعیت به کار می رود [52]. از یک سیستم قوانین فازی برای تشخیص احساسات چهره استفاده می شود که شامل مجموعه ای از قوانین فازی برای یافتن نگاشت بین احساسات و توزیع FAPهاست. توضیح اجزاء مختلف سیستم قوانین فازی در زیر آورده شده است.

۲-۳-۳-۱-نمایش فازی ها

FAPها در MPEG4 تنها وجوهی از فعالیت چهره را مشخص می کنند. به منظور اندازه گیری و پردازش ریاضی FAPها، آنها را با مجموعه های فازی با توابع عضویت خاص مدل می کنند. برای هر FAP دو نقطه P_1 و P_2 تعریف می شود و فاصله ی اقلیدسی بین این دو نقطه را به عنوان اندازه ی FAP می گیرند. رنج تغییر فاصله ی اقلیدسی با آنالیز مجموعه ی دیتای موجود به دست می آید. با استفاده از این رنج سه بخش برای هر FAP تعریف می شود: بالا، میانه و پایین^۲. تمام این بخش ها به عنوان مجموعه های فازی با تابع عضویت سیگموئید تعریف می شوند. توابع عضویت بر اساس آزمایش به دست می آیند. پایین یک مجموعه ی لبه چپ با کاهش سیگموئید است. بالا یک مجموعه ی لبه راست با افزایش سیگموئید و میانه ترکیب سیگموئید چپ و راست است. تابع عضویت برای هر یک از این ترم ها به این صورت تعریف می شود:

$$\mu_{LOW}(d_i) = \frac{1}{1 + e^{(d_i - m)/\sigma\alpha}} \quad (1-2)$$

$$\mu_{HIGH}(d_i) = \frac{1}{1 + e^{(m - d_i)/\sigma\alpha}} \quad (2-2)$$

$$\mu_{MEDIUM} = \frac{1}{(1 + e^{(d_i - m - \alpha/2)/\sigma\alpha})(1 + e^{(m - \alpha/2 - d_i)/\sigma\alpha})} \quad (3-2)$$

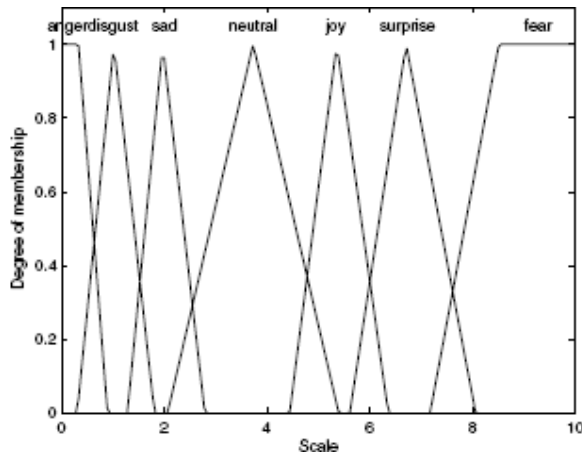
m مقدار میانگین d_i و با استفاده از مجموعه دیتا محاسبه می شود. σ انحراف معیار و α مقدار ثابت با اندازه ی ۰,۱ است.

¹ real life

² LOW ,MEDIUM ,HIGH

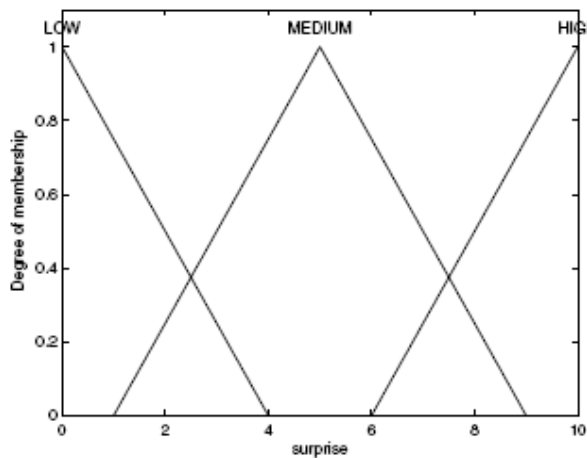
۲-۳-۳-۲- نمایش فازی احساسات

سری احساسات بعنوان مجموعه فازی با هفت بخش جهانی تعریف می شود. این احساسات روی باند حسی از صفر تا ۱۰ مقیاس می شوند. در شکل ۲-۱۱ مکان هر حس در طول محور افقی براساس نتایج تحقیقات روان شناسی مشخص شده است.



شکل (۲-۱۱) توابع عضویت احساسات مختلف

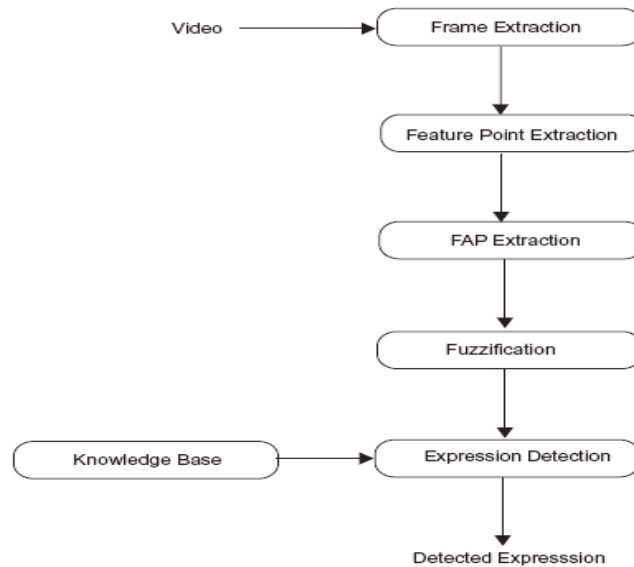
سه مقدار شدت در سیستم فازی برای هر حس با همان پایین، میانه، و بالا مشخص می شوند. این سه بخش با استفاده از مجموعه های فازی از طریق توابع عضویت مثلثی روی مقیاس صفر تا ۱۰ تعریف می شوند. به عنوان مثال شکل ۲-۱۲ درجه عضویت بخش های مختلف با حالت تعجب را نشان می دهد.



شکل (۲-۱۲) تابع عضویت حالت تعجب

۲-۳-۴- توضیح الگوریتم سیستم فازی

با استفاده از روش بالا ممکن است هر ترکیب FAP را به نزدیکترین حالت ممکن نگاشت کرد. شکل ۲-۱۳ الگوریتم سیستم تشخیص حالت فازی را نشان می دهد.



شکل (۲-۱۳) شمای کلی سیستم تشخیص احساس فازی

۲-۳-۵- نتایج سیستم مبتنی بر MPEG4 و کلاسیفایر فازی

در این روش برای هر فریم استخراج شده مجموعه‌ای از فواصل نقاط ویژگی اندازه‌گیری و به مقادیر FAP نگاشت می‌شود. مدل FAP مجموعه‌ای از درجه عضویت FAP را می‌سازد. مدل حالت، مدل FAP خروجی را با هر یک از قوانین مرجع مقایسه می‌کند و مجموعه‌ای از «توابع عضویت» احساسات ایجاد می‌کند. پس از فازی سازی، عملیات فازی زدایی انجام می‌گردد تا احساس مشخص شده استخراج شود. شکل ۲-۱۴ چند حالت و جدول ۲-۶ خروجی سیستم را نشان می‌دهد.



شکل (۲-۱۴) چند حالت آزمایش شده

جدول (۲-۶) نتایج تشخیص سیستم تشخیص احساس فازی

Case	Membership Grade					Output
	Joy	Surprise	Fear	Disgust	Anger	
a	0.5	0.8	0.4	0.2	0.2	Surprise
b	0.6	0	0.1	0.1	0	Joy
c	0.3	0.1	0	0.5	0.2	Disgust
d	0.9	0.4	0.1	0.1	0	Joy
e	0.3	0.6	0.4	0	0.1	Surprise

۲-۴- تشخیص احساس چهره از طریق روش مدل-پایه ی گراف های کلی و محلی

۲-۴-۱- معرفی سیستم

یکی از روش های مهم تشخیص احساس چهره استفاده از گراف های متفاوت حرکت چهره روی کل چهره یا نواحی خاصی از آن مثل چشم و دهان است. گراف های حرکت چهره بر اساس نقاط ویژگی و حرکات ماهیچه ای تنظیم می شوند [2,53]. در اینجا روش مدل-پایه گراف محلی - کلی (LGG^1) را برای تشخیص احساسات چهره از تصاویر استاتیک و در شرایط روشنایی متفاوت ارائه می دهیم که نسبت به تغییرات مقیاس، چرخش و تغییر روشنایی مقاوم است. در ابتدا برای تشخیص پوست پروسه ی ثبات رنگ عصبی^۲ را روی تصاویر رنگی پیاده می کنند. ابتدا مهمترین ویژگی های چهره از قبیل چشم و بینی و دهان استخراج می شود. روش LGG دو گروه اطلاعات محلی (شکل ویژگی های چهره با گراف محلی در هر گره) و اطلاعات کلی (توپولوژی چهره) را استخراج می کند. تشخیص احساسات چهره به کمک مقایسه ی گرافهای محلی - کلی با گرافهای احساسی موجود که در دیتابیس LGG حاضرند انجام می شود.

برای شناسایی نواحی ویژگی چهره یک روش جدا سازی شبه فازی برای جداسازی نواحی مختلف پوست استفاده می شود. یک پروسه ی تطبیق منحنی روی نواحی جدا شده پیاده شده و یک مجموعه ی پاره خط که بعداً برای تعریف گراف محلی و شکل ناحیه به کار می رود، ایجاد می گردد. گراف کلی - محلی که هندسه ی مکانی ویژگی های چهره را یکپارچه می کند بعداً ساخته می شود. برای تشخیص یک احساس، گراف ساخته شده با گراف کلی محلی حاضر در دیتابیس با یک تکنیک تطابق گراف مقایسه می شود.

۲-۴-۲- گراف ناحیه ای ویژگی های چهره (FRG^3)

کاربرد روش جداسازی FRG مجموعه ای از نواحی رنگ ایجاد می کند. روی این نواحی یک پروسه ی تطبیق منحنی به کار می رود که مجموعه ای از پاره خط های متصل می آفریند که شکل هر ناحیه را تعریف می کند. شکل ناحیه در تولید گراف ناحیه ی محلی و سنتز نواحی همسایه به کار می رود. نمایش ناحیه با مجموعه ای از پاره خط ها کافی نیست، چرا که برای توضیح دقیق، شکل هندسی هر ناحیه و شکل ویژگی های چهره خیلی مهم است. یک راه حل مناسب برای این مسئله ایجاد یک گراف محلی است که رابطه ی مکانی این بخش های خطی را کدگذاری کند. در این مرحله شکلی از ناحیه با گراف محلی، با تکه خط در گره ها و اتصالات بین آنها، تعریف می شود.

¹ Local Global Graph

² Neural Color Constancy

³ Facial Feature Region Graph

۲-۴-۳- تطبیق ویژگی های صورت با گراف محلی

در روش گراف کلی محلی، هر ناحیه صورت با گراف محلی نمایش داده می شود. برای تشخیص هر ناحیه از چهره، نیاز است گراف محلی هر بخش را با گراف محلی مدل ذخیره شده در دیتابیس تطبیق دهیم. اگر ناحیه ی ویژگی مدل $f(t)$ و ناحیه ی نامزد را با $g(t)$ نشان دهیم باید برای تطبیق تبدیل هندسی، بهترین نگاشت از $g(t)$ به $f(t)$ ، شامل انتقال، چرخش و مقیاس را نشان دهیم. انتقال بین دو منحنی بسته $f(t)$ و $g(t)$ به این صورت تعریف می شود:

$$g(t) = S \cdot R \cdot f(t) + T \quad (۴-۲)$$

$$\begin{bmatrix} x'(t) \\ y'(t) \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & s_x \\ 0 & 1 & s_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x(t) \\ y(t) \\ 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (۵-۲)$$

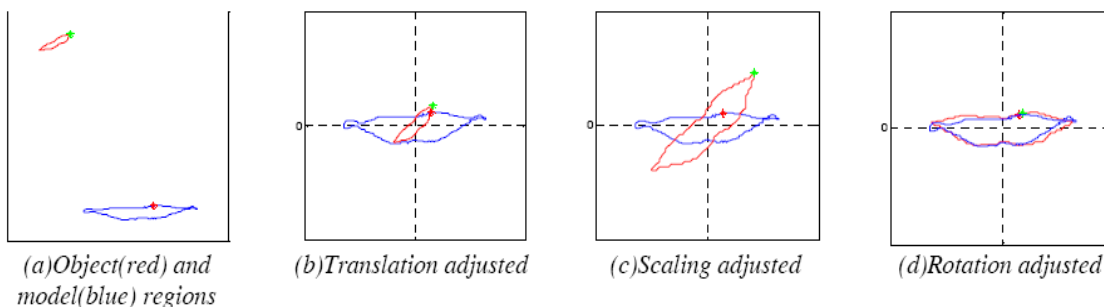
بیشتر تصاویر فاکتور مقیاس یکسانی در دو جهت x و y دارند و به این دلیل، در تبدیل هندسی، چرخش و مقیاس، مرکز ثقل را تغییر نمی دهد و انتقال به این صورت تعریف می شود:

$$T = F_{cen}(g(t)) - F_{cen}(f(t)) \quad (۶-۲)$$

$F_{cen}(f(\cdot))$ تابعی است که مرکز ثقل منحنی را مشخص می کند. پس از محاسبه ی مرکز ثقل هر ناحیه، آن را به نقطه اصلی حرکت می دهند. برای یافتن پارامتر مقیاس، ممتم را محاسبه می کنند. ممتم اندازه ای از توزیع جرم است و به این صورت تعریف می شود.

$$Mom = \frac{1}{N} \sum_{i=1}^N m_i \cdot \|p_i - P_{centroid}\|^2 \quad (۷-۲)$$

N تعداد نقاط منحنی و m_i وزن جرم در نقطه P_i است. انتقال و چرخش شکل شیء را تغییر نمی دهد، بنابراین ممتم می تواند قبل و بعد از انتقال و چرخش تعریف شود. شکل ۲-۱۵ نمونه ی پیچیده ای از تطابق ناحیه ای را نشان می دهد.






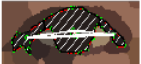




















شکل (۲-۱۵) نمونه ی پیچیده ای از تطابق ناحیه ای

۲-۴-۴- توسعه گراف برای تشخیص احساسات چهره

مراحل کار تشخیص چهره، تشخیص ویژگی و سپس تشخیص حالات است. در روش LGG بعد از مرحله تشخیص چهره، ویژگی های بسیاری به کمک پاسخ گره های گراف کلی-محلی قابل دستیابی است. مرحله ی آخر مقایسه ی بین تصویر گراف LGG با گراف های موجود حالات است. یک دیتابیس خاص شامل گراف های کلی محلی احساسات است. حسن این روش نسبت به روش های دیگر این است که هر شکل ویژگی چهره برای فرد خاص با نمایش گرافی در جزئیات خیلی زیادی حاضر می شود که اجازه می دهد حالت ها را به دقت بیشتری شناسایی کنیم. برای تشخیص حالت، هر گره در گراف LG به این شکل اصلاح می شود.

$$node = \{(x, y), color/texture, L, size, border, LG_{EXPR_i}, \dots, LG_{EXPR_i}\} \quad (۸-۲)$$

ویژگی های استفاده شده ابروهای چپ و راست، چشم چپ و راست و دهان است. شکل ۲-۱۶ گراف LG حالات شادی، عصبانیت، ترس و حالت عادی را نشان می دهد.

Features Cons. Expressions	EyeBrowL (LG_{EXPR_EBL})	EyeBrowR (LG_{EXPR_EBR})	EyeL (LG_{EXPR_EL})	EyeR (LG_{EXPR_ER})	Mouth (LG_{EXPR_M})
 Neutral					
 Happy					
 Angry					
 Scream					

شکل (۲-۱۶) گرافهای LG حالات مختلف

در شماتیک گراف LG بعد از مرحله ی تشخیص چهره، یک تخمین مکانی از ویژگی های چهره که در طول مرحله ی تطبیق گراف LG حاصل شده می تواند به دست آید. یک بخش از ناحیه ی تصویر اطراف ویژگی چهره لحاظ می شود. ناحیه به کمک روش جداسازی FRG بخش بخش شده و گراف کلی محلی حالات ساخته می شود. برای تشخیص حالت، گراف هر ویژگی با گراف حالات همان ویژگی مقایسه می شود. از ترکیب گرافهای کلی و محلی حالات برای شناسایی یک حالت خاص، حداقل خطای کل احساس خاصی را نشان می دهد. از دیگر روش های مطرح مدل-پایه استفاده از رویه های توپوگرافیک می باشد که به علت پیچیدگی های محاسباتی چندان مقبول نیفتاده است.

۲-۵- تشخیص احساس چهره از طریق روش تصویر- پایه ی مدل ظهور فعال

۲-۵-۱- معرفی سیستم

مدل ظهور فعال یا AAM^۱ یکی از روش های توانمند در استخراج ویژگی های چهره به منظور تشخیص احساسات است که به خوبی و با دقت بالایی این ویژگی ها را استخراج می کند و در سال های اخیر بسیار مورد توجه و کاربرد دانشمندان قرار گرفته است [54,55]. در اینجا روشی برای تشخیص حالات چهره بر اساس مدل ظهور فعال با حداقل سازی درجه دو و کلاس بندی شبکه عصبی بررسی می شود [56]. حداقل سازی درجه دو این امکان را می دهد که همگرایی صحیحی با کمترین افت نرخ فریم برای AAM حاصل شود. به کمک روش AAM می توان به درستی شکل^۲ چهره را استخراج کرد. به علاوه بردارهای ویژگی با بعد بالا از شش احساس مختلف شامل شکل چهره و بافت^۳ می تواند به پرسپترون چند لایه وارد شود.

حداقل سازی درجه دو اجازه می دهد که شکل چهره خیلی دقیق از فریم تصاویر ورودی با سرعتی در حدود ۱۱ فریم بر ثانیه جدا شود. از آنجا که تعریف ویژگیهای کلی و فیزیکی هر احساس از شکل و بافت کار دشواری است، سیستم پیش نهادی از پرسپترون چند لایه به عنوان مدول واحد تشخیص استفاده می کند که شکل و بافت یا ضرایب اجزا اساسی آنها را به کار می برد. تصاویر قبل از ورود به پرسپترون چند لایه به شکل میانگین AAM نرمالایز می شوند چرا که تصاویر نرمالایز شده کجی و انحراف کوچکی دارند ولی شکل های اصلی انحراف زیادی از نظر مکان، چرخش و مقیاس دارند. خروجی های پرسپترون چند لایه کدهای باینری پنج رقمی هستند که حالت ها را مشخص می کنند و نرخ تشخیص آن ۹۸٪ است.

۲-۵-۲- استخراج ویژگی های چهره

در این جا چگونگی کاربرد روش حداقل سازی درجه دوم در AAM همان طور که در شکل ۲-۱۷ نشان داده شده است توضیح داده می شود.

اختلاف بین مدل ظهور فعال کنونی و مدل ظهور فعال با الگوریتم ترکیبی معکوس، ژاکوبین تابع وارپینگ^۴ است. مدل ظهور فعال قبلی تنها از ژاکوبین $J(0)$ در $p=0$ استفاده می کند ولی مدل

ظهور فعال با حداقل سازی درجه دوم از ژاکوبین $J' = \frac{1}{2} \{J(p) + J(0)\}$ بهره می برد. شکل ۲-۱۸

یک نمونه از یافتن نقاط ویژگی چهره با AAM را نشان می دهد. ستون اول نتیجه حاصل از الگوریتم ترکیبی معکوس درجه یک و ستون دوم الگوریتم حداقل سازی درجه دو است.

¹ Active Appearance Model

² Shape

³ Texture

⁴ Warping

Pre-compute:

- A. Evaluate gradient $\nabla \mathbf{T}$ of the mean texture \mathbf{T}
- B. Evaluate the Jacobians $\frac{\partial \mathbf{w}}{\partial \mathbf{p}}_{\mathbf{p}=\mathbf{0}}$ and at $\mathbf{p} = \mathbf{0}$

Iterate:

- A. Warp \mathbf{I} with $\mathbf{w}(\mathbf{x}; \mathbf{p})$ to compute $\mathbf{I}(\mathbf{w}(\mathbf{x}; \mathbf{p}))$
- B. Compute $\frac{\partial \mathbf{w}}{\partial \mathbf{p}}_{\mathbf{p}_c}$, \mathbf{J}' , and \mathbf{H}^{-1} as (10)
- C. Compute $\Delta \mathbf{p}$ as (9)
- D. Update $\mathbf{w}(\mathbf{x}; \mathbf{p}) \rightarrow \mathbf{w}(\mathbf{x}; \mathbf{p}) \circ \mathbf{w}(\mathbf{x}; -\Delta \mathbf{p})$

شکل (۲-۱۷) حداقل سازی درجه دوم در AAM

ستون اول نشان می دهد روش درجه اول با مقداری نویز یا حرکت سریع صورت همگرا نمی شود. چرا که این شرایط نامطلوب، تعداد زیادی مینیمم محلی می سازد که روش کاهش گرادیان را گیج می کند، اما روش دوم می تواند بر این شرایط غلبه کند چون مشتق مرتبه دوم بر مینیمم های محلی فائق می آید. [57,58]

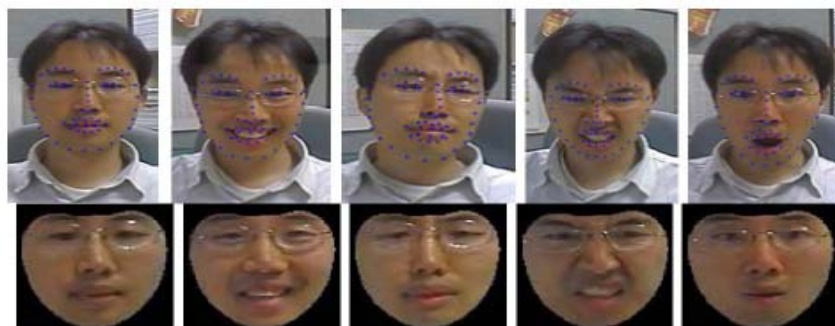
۲-۵-۳- شبکه عصبی برای کلاسه بندی احساس صورت

در این سیستم می توان یک ساختار نرمالایز شده ۳۲ در ۳۲ و یک شکل تبدیل نشده ۶۶ نقطه ای از چهره در طول دنبال کردن چهره توسط AAM به دست آورد. چنین مثالی در ردیف دوم و سوم شکل ۲-۱۹ نشان داده شده است.





شکل (۲-۱۸) نمونه هایی از دنبال کردن چهره با AAM



شکل (۲-۱۹) نمونه ای از شکل و بافت برای پنج حس مختلف

تصویر قبل از ورود به پرسپترون چند لایه به شکل میانگین AAM نرمالایز می شود. تصاویر نرمالایز شده کجی و انحراف کوچکی دارند ولی شکل های اصلی انحراف زیادی از نظر مکان، چرخش و مقیاس دارند. از آنجا که شکل منتقل نشده با چرخش حول محور عمود بر صفحه ی تصویر تغییر نمی کند، انتقال و مقیاس به عنوان ویژگی ثابت به شمار می رود.

ورودی پرسپترون چند لایه بردار یک بعدی ترکیبی از بافت نرمالایز شده و شکل تبدیل نشده کلی و خروجی، برداری پنج بعدی است. بردار نرمالایز شده به دست آمده از آنالیز اجزا اصلی برای ایجاد میانگین صفر و انحراف معیار یک، به جای بردار یک بعدی اصلی به کار می رود. روش های کلاسه بندی مانند SVM و شبکه های دیگر عصبی نیز قابل استفاده اند [78, 79, 82].

۲-۶- تشخیص احساس چهره از طریق روش تصویر-پایه ی مدل ظهور فعال و ویژگیهای مکانی

۲-۶-۱- معرفی سیستم

این روش تشخیص احساس چهره بر اساس مدل ظهور فعال برای استخراج نقاط ویژگی و ترکیب چند ویژگی از نواحی محلی برای تشکیل یک کلاس بندی است [59]. برای بهبود روش مدل ظهور فعال از الگوریتم آدابوست^۱ که الگوریتم تشخیص بسیار قدرتمندی است استفاده می شود تا مکان چشم ها را برای راه اندازی مدل ظهور فعال پیدا کند. توسط مدل ظهور فعال نقاط مهم چهره از قبیل گوشه های چشم و دهان مشخص شده و نقاط ویژگی استخراج می شود. بعد از استخراج نقاط ویژگی تغییرات مکانی چهره را آنالیز کرده و از یک سری ویژگی های ساده برای تشکیل کلاسه بندی کننده موثر استفاده خواهد شد. در آخر تشخیص احساس توسط آنالیز محلی ویژگی های چهره انجام می شود.

۲-۶-۲- استخراج نقاط ویژگی

برای استفاده ی کارآ از الگوریتم تشخیص نیاز است که مکان دقیق بعضی نقاط ویژگی مشخص شود. برای شناختن این نقاط با استفاده از روش های معرفی شده توسط عزیز یومیت باتور^۲ یک مدل شکلی ساخته می شود [60]. یکی مدل شکلی می تواند با کاربرد PCA^۳ به بردارهای x به دست آید.

$$x = \bar{x} + E_x c_x \quad (9-2)$$

یک مدل بافت یا ساختار می تواند با کاربرد PCA روی بافت های نرمال به دست آید.

$$g = \bar{g} + E_g c_g \quad (10-2)$$

\bar{x} شکل میانگین، \bar{g} بافت میانگین در یک محدوده ی میانگین گیری شده است. بعد از حصول مدل های شکل و بافت، یک مدل ظهور ترکیب شده برای ضرایب شکل و بافت خواهیم داشت. این مدل ظهور ترکیبی پارامتری به نام C برای کنترل شکل و بافت دارد:

$$x = \bar{x} + \Phi_x c \quad (11-2)$$

$$g = \bar{g} + \Phi_g c \quad (12-2)$$

Φ_x و Φ_g ماتریس های توضیح مدهای تغییر از مجموعه ی آموزش هستند.

$$\begin{bmatrix} \Lambda c_x \\ c_g \end{bmatrix} = E_a c_a \quad (13-2)$$

¹ Adaboost


² Aziz Umit Batur

³ Principal Component Analysis

ماتریس قطری مقیاس برای اصلاح اختلاف واحدهای ضرایب شکل و بافت است.

$$\Phi_x = E_x \left[\Lambda^{-1} \mid \mathbf{0}_{N_{\epsilon_x} \times N_{\epsilon_g}} \right] E_a \quad (14-2)$$

$$\Phi_g = E_g \left[\mathbf{0}_{N_{\epsilon_g} \times N_{\epsilon_x}} \mid I_{N_{\epsilon_g}} \right] E_a \quad (15-2)$$

مسئله‌ی مورد نیاز برای توضیح عملکرد مدل ظهور فعال بسیار به مکان اولیه وابسته است. مکان اولیه اشتباه به نتایج نادرست و واگرایی منجر می‌شود. برای راه اندازی صحیح مدل ظهور فعال از آدابوست^۱ برای مکان دهی چشمها که مهمترین ویژگی های در چهره اند استفاده می‌شود. این روش یک ناحیه از صورت را با ایجاد مجموعه بزرگی از ویژگی های ساده شبه ویولت-هار^۲ مدل می‌کند، و سپس از آدابوست برای انتخاب ویژگی ها که بین کلاس شی و یک مجموعه از تصاویر پس زمینه اختلاف می‌اندازند استفاده می‌کند. ویژگی های انتخاب شده به تنهایی کلاس بندی ضعیفی می‌سازند اما آدابوست به کلاس بندی ویژگی ها برای تولید یک کلاس ترکیبی قوی وزن می‌دهد. ناحیه اطراف چشمها از هر صورت در مجموعه آموزش با آدابوست استخراج می‌شود. اولین ویژگی آموزش داده شده با آدابوست  است و مکان آن در شکل ۲-۲۰ نشان داده شده است که در کل پنج ویژگی آموزش داده می‌شود.



شکل (۲-۲۰) اولین ویژگی آموزش داده شده توسط آدابوست

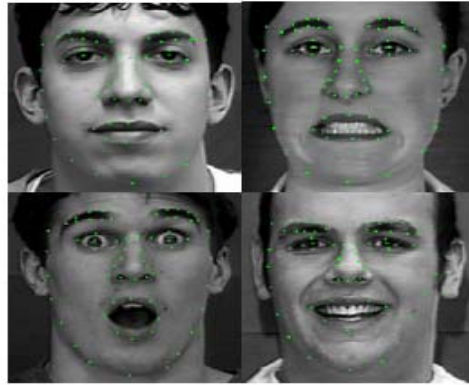
در این جا ۶۳ نقطه روی چهره برای نمایش چهره تعریف می‌شود. بعضی نمونه ها در شکل ۲-۲۱ آورده شده است. به کمک مدل ظهور فعال ما به صورت اتوماتیک می‌توانیم این نقاط را روی چهره مشخص سازیم.

۲-۶-۳- کلاسه بندی و تشخیص حالات چهره

بعد از استخراج ویژگی ها نوبت به تشخیص حالت می‌رسد. یک روش استفاده مستقیم از مختصات نقاط ویژگی یا بردارهای ویژه ی آنها به عنوان بردار ورودی شبکه عصبی برای تشخیص حالات است. به هر حال این روش نتایج خوبی در اختیار نمی‌گذارد چرا که افراد مختلف کانتورهای چهره متفاوت و هم چنین مکان ویژگی های چهره متفاوت دارند.

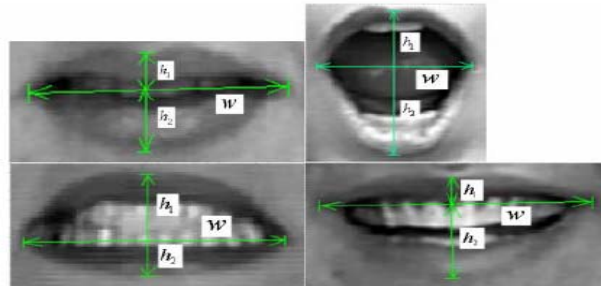
¹ Adaboost

² Wavelet-Haar



شکل (۲-۲۱) مکان نقاط ویژگی بر چهره

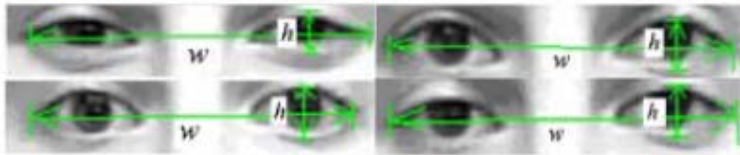
در اینجا یک روش جدید شناسایی احساس با آنالیز مستقل نقاط ویژگی مربوط به هر بخش چهره و ترکیب آنها استفاده شده است که بر اساس سیستم کد کردن AUهای چهره به روش اکمن می باشد. یک احساس ممکن است با احساس دیگری مغشوش شود اما تفاوت هایی در بعضی نواحی چهره برای هر حس وجود دارد. دهان ناحیه ی بسیار پر کاربرد و مهمی است چون در هر چهار حس وضعیت آن به صورت آشکارا تغییر می کند. مطابق شکل ۲-۲۲ محور x خط واصل گوشه ی راست و چپ لب ها و محور y را خط واصل بین بالا و پایین نقاط ویژگی لب تعریف می شود. نقطه ی تلاقی را نقطه ی مرکز و h_1 طول بین نقطه ویژگی بالای لب و نقطه ی مرکز، h_2 طول بین نقطه ی ویژگی پایین لب و نقطه ی مرکز و w عرض دهان است.



شکل (۲-۲۲) پارامتر های دهان

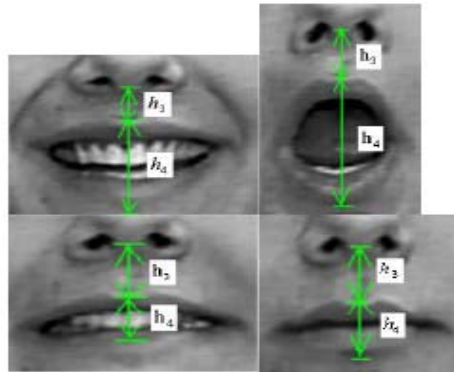
در شکل ۲-۲۲ می بینیم که $w/(h_1 + h_2)$ و h_1/h_2 برای حالات مختلف متفاوت اند پس می تواند فاکتوری برای شناختن احساسات محسوب شود. بسیار آسان است که از $\theta_1 = w/(h_1 + h_2)$ استفاده شود تا حالت تعجب از سه حس دیگر باز شناخته شود. برای حالت عادی یا شادی و ترس $\theta_2 = h_1/h_2$ فاکتور مناسبی است. نرخ شناسایی ۸۰٪ فقط از کاربرد این دو فاکتور به دست می آید. برای چشمها مطابق شکل ۲-۲۳ می بینیم که چشمها همیشه در شادی باریک می شوند اما تغییرات چشمها به آشکاری تغییر دهان نیست اما برای شناسایی حالت شادی از چشمها استفاده می شود. از h/w به

عنوان ویژگی دیگری در کلاس بندی استفاده شده است که h ارتفاع چشم و w فاصله ی دو گوشه چشم است.



شکل (۲-۲۳) پارامترهای چشم

عامل دیگر که گاهی مناسب به نظر می آید h_3 / h_4 است که مطابق شکل ۲-۲۴ می بینیم h_3 طول بین نقطه ی بالای بینی و نقطه ی بالای دهان و h_4 ارتفاع دهان است. این ویژگی ها حدود ۸٪ به نرخ تشخیص می افزایند. ترکیب این ۴ عامل و دادن وزنه‌های مناسب طبق مجموعه ی آموزش در کلاس بندی کمک می کند تا نرخ تشخیص نهایی ۸۸٪ را داشته باشیم.



شکل (۲-۲۴) ویژگی های دهان و بینی

۲-۷- تشخیص احساس چهره از طریق روش تصویر-پایه ی ماتریس انتقال حالت

۲-۷-۱- معرفی سیستم

بازدهی سیستم های تشخیص احساس در موارد رزولوشن پایین به طور قابل توجهی افت می کند و روش های تا به حال بررسی شده هیچ یک نمی توانند با این مسئله به خوبی کنار بیایند [80]. در این قسمت روش تصویر-پایه انتقال حالت برای شناسایی احساسات در تصاویر با وضوح کم معرفی می شود که به نسبت بازدهی خوبی دارد [61]. براساس روش آماری دو روش SVD^1 و نگاشت مستقیم برای محاسبه ماتریس انتقال حالت استفاده می شود [62]. تصاویر تبدیل شده با یک مجموعه ماتریس های انتقال حالت با حالات شناخته شده چهره مقایسه می شوند و حالت صحیح مشخص می گردد. این روش دقت و بازده خیلی بالایی ندارد.

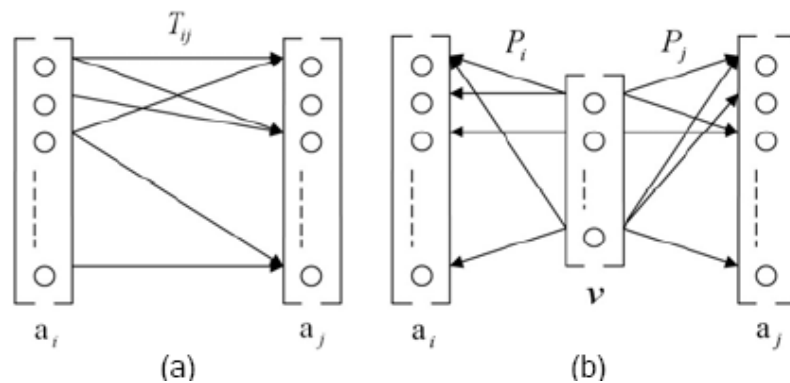
² Singular Value Decomposition

۲-۷-۲- ماتریس انتقال حالت

نواحی موثر چهره به صورت دستی بر اساس جای چشم ها و بالای دهان در فاز آموزش داده شده جدا شده است. برای هر تصویر نرمال شده، یک بردار ویژگی حالت، از اسکن تصویر از بالا-چپ به زیر-راست به دست می آید:

$$\mathbf{a} = [a_{0,0} \quad a_{0,1} \quad \dots \quad a_{n,n}]^T \quad (۱۶-۲)$$

$a_{x,y}$ مقدار پیکسل در نقطه‌ی $B_i = [a_i^1, a_i^2, \dots, a_i^k]$ است. مطابق شکل ۲-۲۵ از دو روش نگاشت مستقیم و SVD برای محاسبه‌ی ماتریس انتقال حالت استفاده می شود.



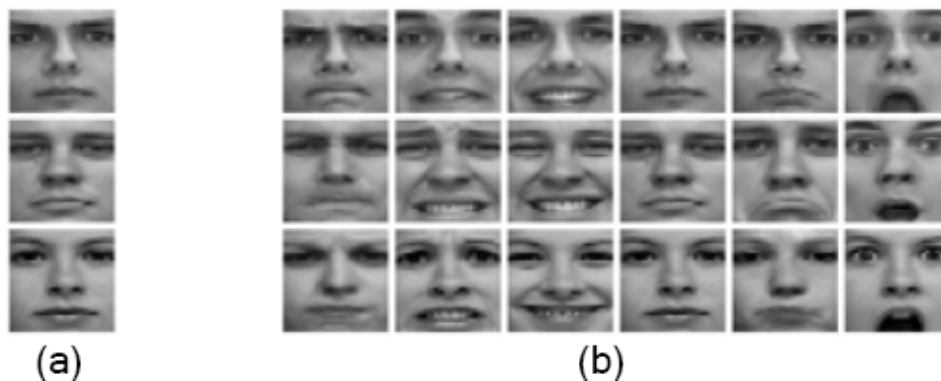
شکل (۲-۲۵) ماتریس انتقال حالت. (a) مپینگ مستقیم (b) SVD

۲-۷-۳- تشخیص احساس چهره

یک ماتریس انتقال حالت انتقالی از تصویر با حالت شناخته شده به تصویری با حالت مطلوب تحویل می دهد و شش حالت را از طریق آن می شناسیم. برای استخراج نواحی مهم چهره از الگوی تطبیقی ساده و سریعی برای پیدا کردن مکان چشمها و بالای دهان استفاده می شود. سپس مقیاس اندازه به منظور بهبود نواحی موثر چهره به کار رفته و در پایان روش هیستوگرام برای برداشتن اثر تغییرات روشنایی به کار می رود و بردار ویژگی \mathbf{a} استخراج می شود. بر اساس ماتریس انتقال حالت می توانیم تصویر را از حالتی شناخته شده به حالتی دیگر تبدیل کنیم.

$$\mathbf{a}_j = T_{ij} \mathbf{a}_i \quad (۱۷-۲)$$

\mathbf{a}_i بردار ویژگی حالت i و T_{ij} ماتریس انتقال از حالت i به j و \mathbf{a}_j بردار ویژگی تبدیل شده است. شکل ۲-۲۶ تصاویر ورودی با حالت خنثی و تصاویر تبدیل شده توسط ماتریس انتقال خنثی - عصبانیت، خنثی - ترس، خنثی - شادی، خنثی - خنثی، خنثی - ناراحتی، خنثی - تعجب به ترتیب است.



شکل (۲-۲۶) انتقال تصاویر (a) حالت خنثی و (b) حالات تبدیل شده

توسط یک تصویر ورودی با حالتی ناشناخته می توان تصاویر تبدیل شده از حالت خنثی با مجموعه ماتریس های انتقال حالت را به تمام حالتها به دست آورد. سپس روش تطابق همبستگی برای محاسبه ی ضرایب همبستگی بین تصاویر تبدیل شده و رجیستر شده از حالت خنثی اعمال می شود. در پایان حالت چهره ی درست با بالاترین ضریب همبستگی انتخاب می شود و نرخ تشخیص حدود ۸۰٪ است که برای تصاویر رزولوشن پایین بازده خوبی محسوب می شود.

۲-۸- تشخیص احساس چهره از طریق روش تصویر-پایه بر اساس توابع ویولت هار^۱

۲-۸-۱- معرفی سیستم

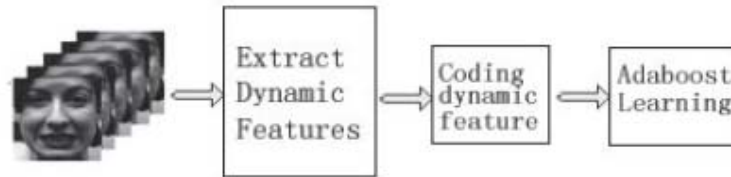
کاملاً طبیعی است که چگونگی استخراج ویژگی های دینامیکی عامل مهمی در آنالیز چهره به روش ویدیو-پایه باشد. در ایجا روش جدید شناسایی AUها و تشخیص احساس بر اساس ویژگی های دینامیکی کد شده ارائه می شود [12]. برای استخراج ویژگی های دینامیکی چهره، ویژگی های دینامیکی شبه هار به منظور نمایش تغییرات زمانی حالات چهره طراحی می شود. در مقایسه با ویژگی های گابر^۲ این ویژگی ها از نظر قیمت محاسبات بسیار ساده ترند و فقط شامل اعمال جمع و منها هستند. ویژگی های شبه هار برای آنالیز موثر و بیشتر به پترن های بانیری کد می شوند. در آخر بر اساس ویژگی های دینامیکی کد شده، آدابوست^۳ برای آموزش ترکیب ویژگی های متمایز بهینه، برای کلاس بندی و تشخیص احساس چهره به کار می رود. شکل ۲-۲۷ ساختار کاری سیستم را که سه بخش دارد نشان می دهد. سه بخش شامل استخراج ویژگی های دینامیکی، کد کردن ویژگی های دینامیکی و آموزش آدابوست است. برای استخراج ویژگی های دینامیکی، ویژگی های شبه هار دینامیکی طراحی شده است تا

¹ Haar

² Gabor

³ Adaboost

تغییرات زمانی AUها و حالات چهره را نشان دهد. با الهام از کد کردن پترن باینری، ویژگیها آنالیز گردیده و برای آن ها کتاب کد تعریف می شود. طبق این کتاب کد ویژگی های دینامیکی شبه هار به ویژگی های پترن باینری نگاشت می شوند. در آخر هم برای آموزش ویژگی های کد شده از آدابوست استفاده کرده اند.



شکل (۲۷-۲) ساختار سیستم مبتنی بر هار

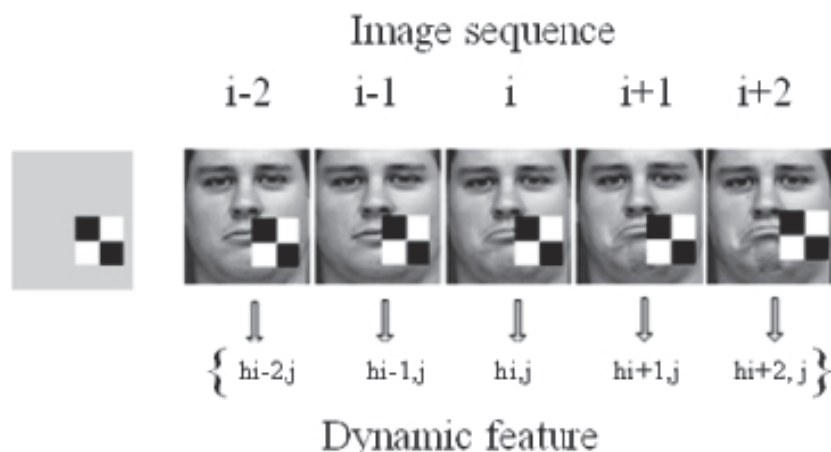
۲-۸-۲- نمایش ویژگی های دینامیکی شبه هار

ویژگی های شبه هار نقش مهمی در تشخیص چهره دارند. آنها در مقایسه با ویژگی های گابریل حجم محاسباتی خیلی کمی دارند و از آنجا برای نمایش ویژگی های دینامیکی چهره مناسب اند. هزاران ویژگی شبه هار از هر فریم تصویر استخراج می شود که ویژگی های یکسان در فریم های متوالی به عنوان ویژگی های دینامیکی با هم ترکیب می شود. شکل ۲۸-۲ پیاده سازی برخی ویژگی های شبه هار را نشان می دهد.



شکل (۲۸-۲) نمونه هایی از ویژگی های شبه هار پیاده شده روی چهره

برای سادگی یک تصویر I را با n فریم در نظر می گیرند و هر فریم با برچسب I_i لحاظ می شود. H مجموعه ویژگی های شبه هار یک تصویر در نظر گرفته می شود. برای هر ویژگی I_i به آن برچسب $h_{i,j}$ می زنند که i اندیس فریم و j اندیس ویژگی شبه هار در مجموعه ویژگی H است. بر اساس هر ویژگی شبه هار $h_{i,j}$ ، یک واحد ویژگی شبه هار دینامیکی $u_{i,j}$ در حوزه ی زمان ساخته می شود: $u_{i,j} = \{h_{i-k,j}, h_{i-k+1,j}, \dots, h_{i+k,j}\}$. شکل ۲۹-۲ این واحد را نشان می دهد. هر $u_{i,j}$ را یک واحد ویژگی دینامیکی نامند. تغییرات زمانی AUهای چهره می تواند به طور کامل توسط $u_{i,j}$ ها توضیح داده شود.



شکل (۲-۲۹) نمونه ای از یک واحد ویژگی دینامیکی در سکانس تصویر

۲-۸-۳- کد کردن ویژگی های دینامیکی

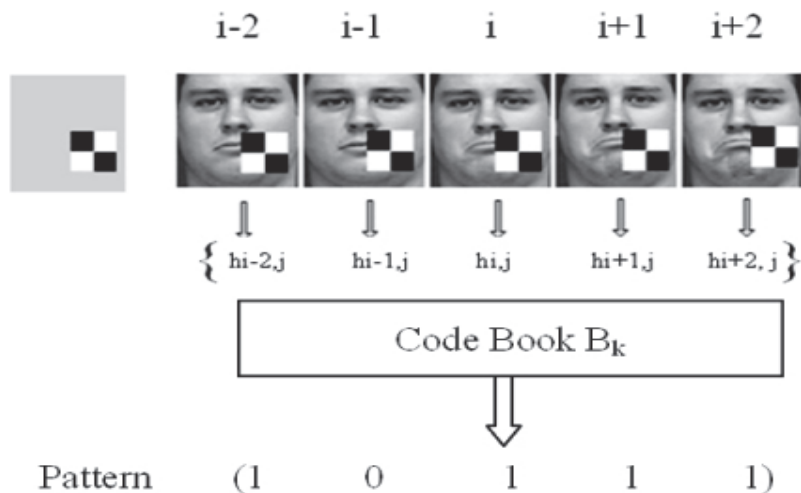
گفتیم یک واحد ویژگی دینامیکی شامل مجموعه ای از ویژگی های شبه هار در مکان یکسانی است. در واقع یک واحد ویژگی دینامیکی یک بردار ویژگی است. در این سیستم $u_{i,j}$ به پترن باینری تبدیل می شود. این سیستم کدینگ دو مزیت دارد: اول این که آموزش آدابوست با یک ویژگی، ساده تر از آموزش با یک بردار ویژگی است. دوم این که کدینگ باینری بر اساس توزیع آماری داده های آموزش است، بنابراین نسبت به نویز مقاوم است.

برای هر حالت چهره و هر AU بر اساس داده های آموزش، مجموعه کدی مورد نیاز است. ابتدا توزیع هر واحد ویژگی u_j را تحت هر حالت یا AU آنالیز کرده و میانگین و واریانس را تخمین می زنند. توزیع گوسین $N_j(\mu, \sigma)$ با تخمین توزیع هر واحد ویژگی u_j سازگار است. سپس کتاب کد $B_k(N_1(\mu_1, \sigma_1), N_2(\mu_2, \sigma_2), \dots, N_m(\mu_m, \sigma_m))$ را برای واحدهای ویژگی به دست می آورند. k برچسب حالتهای مختلف، j اندیس واحد ویژگی و m اندازه ی مجموعه ویژگی H است. بر اساس کتاب کد می توان هر $h_{i,j}$ دینامیکی را به پترن $\{1,0\}$ نگاشت کرد.

$$C_{i,j} = \begin{cases} 0: & \text{if } \frac{\|h_{i,j} - \mu_j\|}{\sigma_j} > T \\ 1: & \text{if } \frac{\|h_{i,j} - \mu_j\|}{\sigma_j} < T \end{cases} \quad (18-2)$$

بر اساس رابطه ی بالا می توان یک واحد ویژگی دینامیکی شبه هار $u_{i,j}$ را به یک پترن باینری $T_{i,j}$ نگاشت کرد. شکل ۲-۳۰ پروسه ی ایجاد ویژگی کد شده ی $T_{i,j}$ را نشان می دهد.

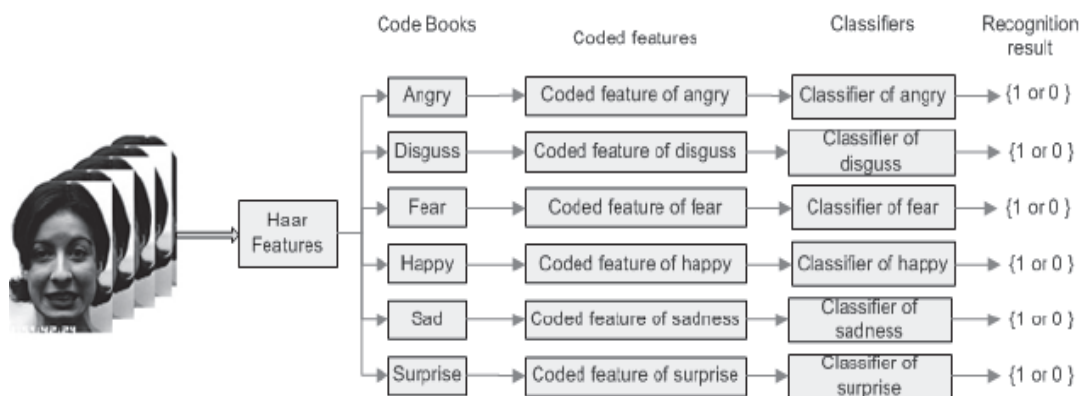
$$T_{i,j} = \{C_{i-L,j}, C_{i-L+1,j}, \dots, C_{i,j}, \dots, C_{i+L-1,j}, C_{i+L,j}\} \quad (18-2)$$



شکل (۲-۳۰) نمونه ای از کدینگ یک واحد ویژگی دینامیکی

۲-۸-۴- آموزش آدابوست

از آنجا که هزاران ویژگی کد شده ی دینامیکی داریم برای ایجاد کلاس بندی نهایی باید مجموعه ای از ویژگی های متمایز انتخاب شود. در این جا از آموزش آدابوست بهره می برند. توالی تصویر از یک حالت را به عنوان نمونه های مثبت و سکانس تصاویر با حالت های دیگر را نمونه های منفی می گذارند و یک کلاس بندی روی هر حالت بر اساس ویژگی های کد شده پاسخ انجام می دهند. پروسه ی تست در شکل ۲-۳۱ نشان داده شده است.



شکل (۲-۳۱) پروسه ی تشخیص احساس

۲-۹- تشخیص احساس چهره از طریق روش تصویر-پایه ی استفاده از فیلترهای گابر

۲-۹-۱- معرفی سیستم

در سال های اخیر روش توانا و ساده ای که در استخراج ویژگی های چهره در پروسه ی تشخیص احساس بسیار مورد توجه قرار گرفته است مبتنی بر فیلترهای گابر^۱ است. این فیلترها را می توان روی کل چهره یا بخش های ویژه ای از آن اعمال کرد و نتایج برای احساسات مختلف متفاوت و به راحتی قابل کلاسه بندی اند [63,64]. در این قسمت درباره ی فیلترهای گابر برای استخراج ویژگی ها در ترکیب با LVQ^۲ برای تشخیص هفت احساس چهره بحث می کنیم [8]. نتایج حاصل شده از این روش در چندین زمینه نسبت به روشهای دیگر در زمینه تشخیص احساس بهتر است. کلاس بندی ویژگی ها در تشخیص احساس ترس نسبت به روش MLP^۳ نتایج بهتری دارد.

پروفسور لیونز^۴ در ۱۹۹۸ روشی برای کدینگ^۵ احساسات چهره توسط ویولت گابر ارائه کرد [42,43] و ژانگ در ۱۹۹۸ نشان دادند کدینگ حالات چهره بر اساس فیلتر بانک های گابر برای استخراج ویژگی و کلاس بندی ویژگی ها بر اساس MLP نتایج بهتری نسبت به تشخیص حالات چهره بر اساس ویژگی های هندسی دارد [5]. در این قسمت روش استخراج ویژگی پیشنهاد شده توسط لیونز در ۱۹۹۸ تطابق یافته است و PCA برای کاهش طول بردار ویژگی استفاده می شود. هر چند شبکه عصبی به طور گسترده ای برای کلاس بندی و تشخیص احساس به کار می رود اما استفاده از شبکه عصبی در تشخیص چهره چندین مشکل به همراه دارد. چندین ساختار مختلف شبکه عصبی وجود دارد که هر یک نقاط ضعف و قدرت خود را دارد. عملکرد خوب یک ساختار در مسئله ای خاص متضمن نتایج مشابه در مسئله ای دیگر نیست. در اینجا محاسن استفاده از LVQ برای تشخیص حالات چهره نسبت به MLP مورد بررسی قرار می گیرد و نتایج LVQ روی دیتا بیس JAFFE با عملکرد MLP گزارش شده در کارهای ژانگ مقایسه می شود.

۲-۹-۲- استخراج ویژگی

ابتدا تصاویر با استفاده از دوربین گرفته و پیش پردازش می شود تا تغییرات محیطی در تصاویر مختلف حداقل شود. پیش پردازش تصاویر شامل عملکردهایی از قبیل تنظیم مقیاس و روشنایی تصاویر، تنظیم

¹ Gabor

² Learning Vector Quantization

³ Multi Layer Perceptron

⁴ Lyons

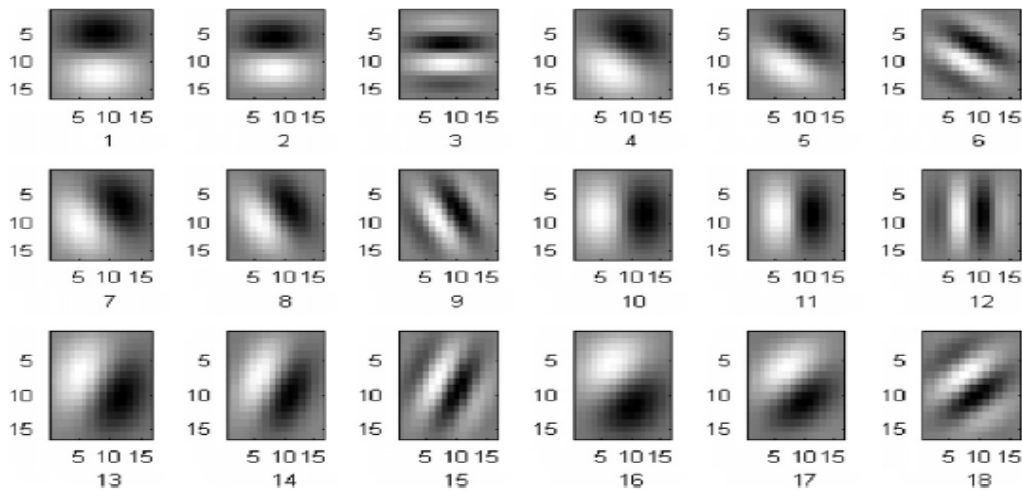
⁵ Coding

کنتراست^۱ و دیگر عملیات بهبود تصویر است. یک تابع دو بعدی گابریل موج صفحه ای با فاکتور موج K است که توسط تابع پوش گوسین با پهنای نسبی σ محدود شده است.

$$\psi(k, x) = \frac{k^2}{\sigma^2} \exp\left(-\frac{k^2 x^2}{2\sigma^2}\right) \left[\exp(ik \cdot x) - \exp\left(-\frac{\sigma^2}{2}\right) \right] \quad (19-2)$$

مقدار σ برای تصاویر با رزولوشن ۲۵۶ در ۲۵۶ برابر با π است. این مجموعه ی گسسته با شماره موج $K = \pi/4, \pi/8, \pi/16$ و شش چرخش متمایز از صفر تا ۱۸۰ درجه در گام های ۳۰ درجه که فیلتر بانکی از ۱۸ فیلتر گابریل مختلف می سازد تشکیل شده است. شکل ۲-۳۲ هجده کرنل مختلف فیلتر گابریل را نشان می دهد.

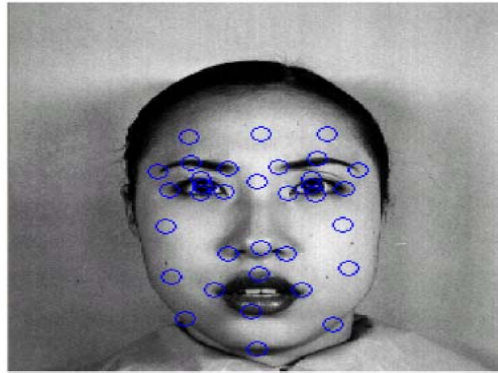
فیلترهای گابریل برای هر یک از تصاویر استفاده می شود و پاسخ های فیلتر تنها در نقاط ثابت از پیش تعریف شده به دست می آید. به منظور مقایسه ی عملکرد LVQ و MLP، ۳۴ نقطه ی ثابت برای دستیابی به پاسخ فیلتر بانک گابریل استفاده می شود. این نتایج در یک بردار ویژگی با طول ۶۱۲ حالت تصویر ورودی را نشان می دهد که حاصل اثر ۱۸ فیلتر بر ۳۴ نقطه ی ثابت است.



شکل (۲-۳۲) کرنل های فیلتر گابریل استفاده شده برای حصول بردار ویژگی

شکل ۲-۳۳ مکان ۳۴ نقطه ی ثابت در چهره را از جایی که پاسخ های فیلتر گابریل ساده می شوند نشان می دهد. این نقاط می توانند توسط ویژوال بیسیک به طور خودکار و یا به صورت دستی مشخص شوند. مختصات هندسی نقاط هر تصویر پس از استخراج برای پردازش بیشتر وارد فیلتر گابریل می شود.

¹ Contrast



شکل (۲-۳۳) مکان نقاط ثابت روی چهره

۲-۹-۳- آنالیز اجزا اصلی

آنالیز اجزا اصلی تکنیکی است برای کاهش بعد فضای ویژگی که مجموعه ای از نقاط دیتا را می گیرد و زیرفضای خطی با بعد کمتری می سازد که تغییرات این نقاط دیتا را از میانگینشان بهتر توضیح می دهد. آنالیز اجزا اصلی تبدیلی خطی است که معمولا برای ساده کردن یک مجموعه ی دیتا با کاهش بعد مجموعه ی دیتا استفاده می شود. با استفاده از آنالیز اجزا اصلی کاهش بعد مجموعه ی دیتا در حالی است که دیتای اطراف واریانس را می دهد، اجزای اصلی درجه پایین را نگه می دارد و از اجزا درجه بالا چشم می پوشد. آنالیز اجزا اصلی ذاتا تبدیل خطی بهینه ای است که زیرفضایی که بزرگترین واریانس را دارد نگه می دارد. برخلاف دیگر تبدیل های خطی، آنالیز اجزا اصلی مجموعه ی ثابتی از بردارهای پایه ندارد و بردارهای پایه آن بستگی به مجموعه ی دیتا دارد. طول بردار ویژگی به تدریج از ۱۰ تا جایی که به نرخ تشخیص صدمه نزند، افزایش می یابد.

۲-۹-۴- کلاس بندی به کمک شبکه عصبی کوهنن خود سازمان ده

شبکه عصبی کوهنن خود سازمان ده یا LVQ¹ که توسط کوهنن^۲ در سال ۲۰۰۱ توسعه یافت یکی از پرکاربردترین الگوریتم های خوشه بندی بدون نظارت است. چندین نوع از LVQ و LVQ-I در اینجا مورد استفاده قرار گرفته است. LVQ-I دو لایه دارد: لایه ی رقابت و لایه ی خروجی. نرون ها در لایه ی رقابت به نام زیرکلاس هم خوانده می شوند. هر زیرکلاس بردار وزنی مشابه بردار ورودی دارد. هنگامی که یک بردار ورودی برای یک LVQ به کار می رود بهترین تطابق در لایه رقابت جستجو شده و به بهترین تطابق لایه برنده گویند. وقتی نرون خاصی در لایه رقابت برنده می شود خروجی خاص مربوط به کلاس آن نرون بالا می رود. چندین نرون در لایه رقابت ممکن است به کلاس یکسانی در لایه خروجی پاسخ

¹ Learning Vector Quantization

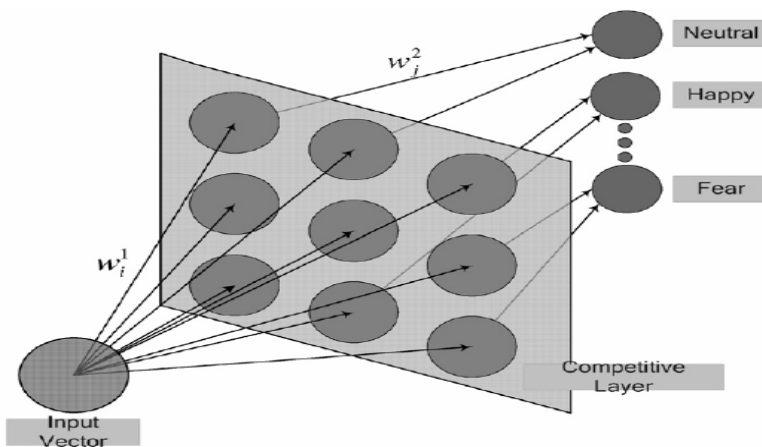
² Kohonen

دهند اما یک نرون در لایه رقابت تنها به کلاس خاصی مربوط می شود. این به این خاطر است که نرون ها در لایه رقابت به نام زیرکلاس ها شناخته می شوند.

روش آموزشی که معمولا برای LVQ استفاده می شود قانون آموزش رقابتی است که برای هر الگوی آموزش داده شده، نرون لایه رقابتی که به ورودی نزدیکترین است مشخص می شود و نرون خروجی پاسخ به نام نرون برنده خوانده می شود. وزن های اتصالات این نرون ها با فرمول های زیر تطابق می یابد:

$$w_i^1(n) = \begin{cases} w_i^1(n-1) + \alpha(p - w_i^1(n-1)) & \text{if classification is correct} \\ w_i^1(n-1) - \alpha(p - w_i^1(n-1)) & \text{otherwise} \end{cases} \quad (2-20)$$

که w_i وزن لایه ورودی است، p بردار ورودی و α نرخ آموزش است. با توجه به این معادله جهت تطابق وزنی بستگی به این دارد که کلاس پترن آموزش و کلاس تعیین شده به بردار مرجع یکسانی هست یا نه. اگر یکسان باشد بردار مرجع به نزدیکی پترن آموزش حرکت می کند. در غیر این صورت دور می شود. این حرکت بردار مرجع توسط نرخ آموزش کنترل می شود.



شکل (۲-۳۴) ساختاری از شبکه LVQ

این حالت بخشی از فاصله تا پترن آموزش را به خود می گیرد که بردار مرجع چقدر حرکت کرده است. معمولا نرخ آموزش در زمان کاهش می یابد. بنابراین تغییرات ابتدایی نسبت به تغییرات ساخته شده در مبدا های بعدی پروسه آموزش بزرگترند. شکل ۲-۳۴ ساختاری از شبکه LVQ را می بینید.

فصل سوم

مروری بر تبدیل موجک و گابر

۳-۱- مقدمه

مرحله اول در تشخیص احساسات چهره استخراج اطلاعات و الگوهای مناسب است تا در مرحله بعد با استفاده از کلاسه بندی مناسب موفق به تشخیص احساس شویم. استخراج این الگوها و اطلاعات به روش های گوناگون انجام می شود که به نوع و مشخصات کیفی تصویر وابسته است. روش های مناسب استخراج ویژگی در موارد استاتیک و دینامیک، تصاویر رنگی و سیاه سفید، دید از روبرو یا با چرخش سر متفاوت است. در رابطه با موضوع این رساله از جمله روش های قدرتمند استخراج ویژگی استفاده از تبدیل موجک و فیلترهای گابر است که در ادامه به شرح مختصری درباره آن ها می پردازیم.

۳-۲- تاریخچه موجک ها^۱

از نقطه نظر تاریخی ریاضیات موجک به کار جوزف فوریه در قرن نوزدهم بر می گردد. فوریه به پایه گذاری تئوری آنالیز فرکانسی اقدام نمود، که مشخص شد که بسیار مهم و تاثیر گذار است. توجه محققان تدریجاً از آنالیز براساس فرکانس به آنالیز بر اساس مقیاس تغییر پیدا کرد، و این وقتی بود که روشن شد

¹Wavelet.

که روشهایی با حساسیت کمتری به نویز وجود دارند که میانگین تغییرات در مقیاسهای مختلف را اندازه گیری می کنند [65].

اولین نتایج ثبت شده از چیزی که ما آن را موجک می نامیم به نظر میرسد در سال ۱۹۰۹ در پایان نامه دکترای آقای آلفرد هار باشد. اگر چه ایده ها و ساختارهای تئوری موجک از اوایل قرن بیستم مطرح بوده است، اما اخیرا به صورت یک تئوری واحد و در برگیرنده تعداد زیادی از تکنیکهایی که به طور مستقل گسترش یافته اند، رشد و توسعه یافته و در بسیاری از حوزه های مختلف مانند ریاضیات، فیزیک و مهندسی پایدار گشته است. در اواسط دهه ی هشتاد پژوهشگران دانشگاه فرانسه با راهنمایی های یک محقق فیزیک نظری و دو ریاضیدان با نامهای مورلت^۱، گراسمن^۲ و میر^۳ پایه و شالوده های ریاضی تئوری موجک را بنا نهاده و به طور قابل ملاحظه ای ارتباط آن را با حوزه های دیگر بیان کردند. به محض با خبر شدن انجمن پردازش سیگنال، مالات^۴ و دابچیز^۵ اساس تئوری موجک را در زمینه های مهندسی، خصوصا پردازش سیگنال پایه ریزی کردند. مالات از مفاهیم آنالیز مولتی رزولوشن سیگنال ها در تعریف موجک ها استفاده کرد، در حالیکه دابچیز موجک های اورتونرمال را بر پایه تکرار فیلترهای دیجیتالی بنا نهاد. پس از آن متفکران زیادی با همکاری یکدیگر در این خصوص کار کردند و در نتیجه این موضوع به سرعت رشد و توسعه یافت. بالاخره در چند سال اخیر دریافته اند که تئوری موجک تکنیک های زیادی را مانند پردازش مولتی رزولوشن سیگنالها، بینایی کامپیوتر^۶، کد کردن زیرباندها و شناسایی لبه ها^۷، فشرده سازی صوت و تصویر و بسط سری های موجک که در ریاضیات به کار برده می شود شامل شده و همگی این نظریه های مختلف یک تئوری واحد به نام تئوری موجک را می سازد [74].

هم چنین از موجک ها در آشکارسازی و آنالیز سیگنال های گذرا استفاده شده است و اخیرا نیز این تئوری از دیدگاهی مشابه تبدیل فوریه زمان کوتاه^۸ و به دلیل شباهت زیادی که به آن دارد در آنالیز سیگنال های غیر ایستان کاربرد فراوانی یافته است. به عبارت دیگر تبدیل موجک را نیز می توان همانند تبدیل فوریه زمان کوتاه برای نمایش زمان فرکانس چنین سیگنال هایی استفاده کرد. در پیوست ۱ درباره ی آنالیز فوریه و تبدیل فوریه ی زمان کوتاه و تبدیل موجک و انواع آن توضیحاتی آورده شده است.

۳-۳- کاربرد موجک در پردازش تصویر

کارایی تبدیل موجک در پردازش تصویر در سال های اخیر به خوبی ثابت شده است [71]. موجک ها ویژگی های مهمی از قبیل تعامد طبیعی^۹، حمایت فشرده^۱، نمایش هم زمان زمان و فرکانس و الگوریتم

¹ Morlet

² Grossmann

³ Meyer

⁴ Mallat

⁵ Daubechies

⁶ Computer Visison

⁷ Edge detection & sub band coding

⁸ Short-Time Fourier Transform (STFT)

⁹ Biorthogonality

سریع را در خود گرد آورده اند. نمایش محتوای فرکانسی هر ناحیه بر حسب مقیاس چارچوب ایده آلی برای آنالیز ویژگی های تصویر فراهم کرده است. مثلا در جایی که ویژگی های مورد نظر دارای اندازه های متفاوتی اند، می توانند بر حسب ویژگی های فرکانسی شان با هم مقایسه شوند. اگر چه فرم استاندارد موجک بر اساس انتقال و مقیاس محدودیت هایی از قبیل فقدان شیفت اینواریانس دارد لیکن به نحو مطلوبی فرکانس و مقیاس را به هم مربوط می سازد که در یافتن پترن خاصی در تصویر، نویز زدایی، فشرده سازی تصاویر و دیگر موارد بسیار کاربرد دارد. موجک با فیلتر کردن تصویر در سطوح مختلف امکان بررسی بیشتر تصویر را به ما می دهد [75].

ویژگی های موجک نشان می دهد که محتوای فرکانسی یک تصویر به اندازه ی آن مربوط نمی شود و این ویژگی مهمی است که در دنبال کردن پترن و مقایسه تصاویر استفاده می شود. یک فرم از موجک به ما نمایش کاملی از دامنه فرکانسی در هر مقیاس را می دهد و ما را توانا می سازد هر ناحیه را روی طیفی از فرکانس آنالیز کنیم و باعث انعطاف پذیری بیشتر می شود. این خاصیت در بسیاری از مسائل از قبیل استخراج ویژگی های خطی و منحنی، آنالیز بافت، و ردیابی یک الگوی مشخص کاربرد دارد. حسن اصلی موجک در این است که انعطاف پذیری بیشتری در طراحی مدل ها و تخمین پروسه برای آنالیز مسائل به ما می دهد. روش موجک تعیین کردن مدل های دامنه فرکانسی مناسب و ایجاد فرضیه هایی برای تست برای کلاس های مختلف از ویژگی است و هم چنین مقیاس مناسبی که مدل ها را صحیح می سازد مشخص می کند. این ویژگی ها به اندازه کافی موثر و کارا هستند تا در طیف وسیعی از مسائل پردازش تصویر به کار روند. موجک در فشرده سازی تصاویر و بررسی تصاویر پزشکی و پردازش تصویر در اندازه های بالا کاربرد دارد. یکی از روش های فشرده سازی تصاویر استفاده از الگوریتم های مولتی رزولوشن هرمی معرفی شده توسط بورت^۱ و آدلسن^۲ است. در این روش تصویر ابتدا توسط یک فیلتر گوسی پایین گذر فیلتر شده و تفاضل نقطه به نقطه ی تصویر حاصل از تصویر اصلی محاسبه می گردد. در نتیجه مولفه های فرکانس بالای تصویر اصلی در این تصویر تفاضلی حفظ می شود.

هم چنین موجک فیلترهای بالا گذر و پایین گذر را روی تصویر اعمال می کند و تصویر اصلی را به دو زیر باند تخمین^۳ در باند پایین تر و جزییات^۴ در باند بالاتر تقسیم می کند. می توان با کاربرد تعداد فیلترهای بیشتر به تعداد تصاویر زیر باند افزود. دو دسته ی کلی بازسازی موجک بازسازی استاندارد و غیر استاندارد است که در شکل ۱-۳ می بینیم.

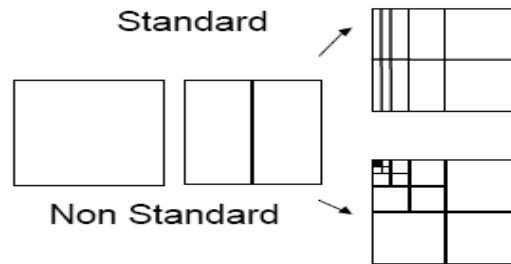
¹ Compact Support

² Burt

³ Adelson

⁴ Coarse Scale Approximation

⁵ Detail



شکل (۱-۳) نمایش دو روش بازسازی در موجک

در بازسازی استاندارد تبدیل موجک یک بعدی روی تمام سطرها اعمال می شود و تبدیل موجک از فیلترهای دوبانچی و فیلترهای تپ بیو ارتوگونال^۱ استفاده می کند. در فرم غیر استاندارد موجک یک بعدی اول روی سطرها و سپس روی ستون ها اعمال می شود که چهار بخش تصویر فیلتر شده را می سازد و سپس همین مراحل را برای بخش تخمین^۲ در دو جهت سطر و ستون به کار می برد. در کل فرم غیر استاندارد برای تحلیل تصویر موثرتر است.

فشرده سازی تصاویر زمینه ی مهم و پرکاربرد موجک است. برای مثال در فشرده سازی تصاویر پزشکی تشخیص هنگامی موثر است که تکنیک فشرده سازی تمام اطلاعات مهم و مفید تصویر را حفظ کند و به کمک تبدیل موجک ما به این مقصود می رسیم. پارامتر مهم موجود در تصویر بافت آن است که موجک به خوبی آن را حفظ و آنالیز می کند.

هدف اصلی در فشرده سازی تصاویر جابه جایی افزونگی^۳ در یک تصویر برای یافتن نمایشی فشرده تر است. روش رایج در فشرده سازی تصاویر کدینگ انتقال نامیده می شود که تصویر را در فضای متفاوتی نسبت به فضای خود تصویر نمایش می دهد. ثابت شده است که بازسازی موجک چند رزولوشنی تصویر را توسط توابع مقیاس به زیر فضاهایی منتقل می کند. پس تصویر روی توابع مقیاس پایه تخمین هایی از سیگنال و تصویر روی پایه های موجک اختلاف بین تخمین های دو سطح تنظیم شده از رزولوشن را به ما می دهد. بنابراین موجک ها قابلیت این را دارند که به خوبی در فشرده سازی تصاویر به کار روند. نمونه ای از استفاده از تبدیل موجک در فشرده سازی معروف لنا در شکل ۳-۲-الف تا ۳-۲-د آورده شده است که نتایج بسته به تعداد سطح خاکستری استفاده شده در آنالیز کوآکیورنس^۴ در مرحله ی اول فشرده سازی تغییر می کنند [102].

¹ tap biorthogonal

² coarse scale approximation

³ Redundancy

⁴ Cooccurrence



شکل (۳-۲-الف) تصویر اصلی لنا



شکل (۳-۲-ب) تصویر بازسازی شده از ۱۶ سطح کوانتیزه ی موجک

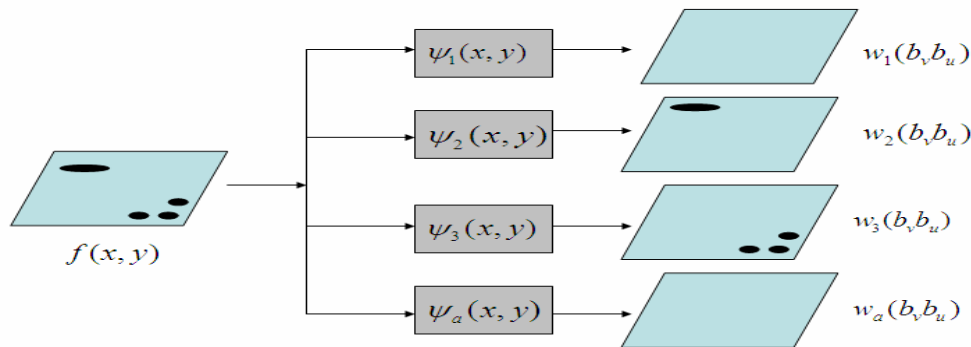


شکل (۳-۲-ج) تصویر بازسازی شده از ۶۴ سطح کوانتیزه ی موجک



شکل (۳-۲-د) تصویر بازسازی شده از اعمال معکوس موجک در کل تصویر

از دیگر کاربردهای موجک در پردازش تصاویر استخراج موارد خاص از تصویر است. برای درک این مطلب برای مثال فرض کنید که تصویر موجود در شکل ۳-۳ شامل اشیاء دایروی با اندازه‌های متفاوت باشد و موجک اساسی انتخاب شده تنها به اشیاء دایروی با شعاع واحد پاسخ دهد. به این ترتیب بررسی تصاویر خروجی فیلترهای میان‌گذر می‌تواند اطلاعات مضاعفی در رابطه با مکان اشیاء دایروی موجود در تصویر با شعاع معین در اختیار قرار دهد. زیرا هر شیئی به شعاع معین، تنها در تصویر خروجی متناظر به اندازه‌اش ظاهر می‌شود.



شکل (۳-۳) : بانک فیلتری برای تبدیل موجک پیوسته تصویر

۳-۴- توضیح مختصری درباره تبدیل گابر

دنیس گابر^۱ در سال ۱۹۴۶ تبدیل فوریه ای برای آنالیز تنها یک بخش کوچک از سیگنال در یک زمان مشخص اصلاح نمود که این تکنیک را پنجره کردن سیگنال نامید. روش تطبیقی گابر به نام تبدیل فوریه زمان کوتاه نامیده شد و یک سیگنال را به یک تابع دو بعدی از زمان و فرکانس نگاشت می‌کرد. تبدیل فوریه زمان کوتاه نوعی مصالحه بین نمایش زمانی و فرکانسی سیگنال ارائه می‌دهد و اطلاعاتی درباره ی اینکه چه زمانی و در چه فرکانسی یک رخداد خاص برای سیگنال رخ می‌دهد نمایش می‌دهد. بعد از گابر اصلاحاتی روی تابع گابر رخ داد و به آن شکل و کاربرد تازه ای بخشید. برای مثال فیلترهای گابر برای تعیین لبه، استخراج ویژگی‌های بافت و شبیه سازی سلول‌های ساده و پیچیده، تشخیص کانتور شی و توضیح اثرات مفهوم بینایی به کار می‌رود. رابطه ی ۳-۱۰ کانولوشن تکی یا چند گانه ی یک تصویر را تابع گابر دو بعدی نشان می‌دهد. خروجی فیلتر گابر همان اندازه ی ابتدایی تصویر را دارد و

¹ Denis Gabor

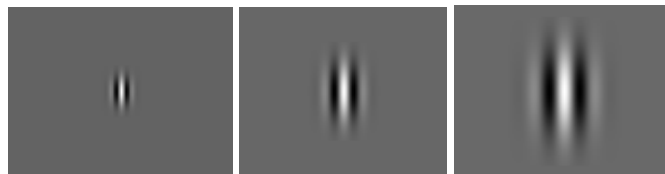
بسته به میزان پارامترهای تابع گابری که توضیح برخی از آن‌ها در زیر داده می‌شود، ویژگی‌های خاصی از تصویر را نمایش می‌دهد. توابع گابری دو بعدی پیشنهاد شده توسط داگمن^۱ برای مدل کردن ویژگی‌های مکانی سلول‌های ساده، به طور گسترده در پردازش تصویر، بینایی ماشین، نروساینس^۲ و روان‌شناسی کاربرد دارد [77,81].

$$g_{\lambda, \theta, \phi, \sigma, \gamma}(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \phi\right) \quad (3-10)$$

$$x' = x \cos \theta + y \sin \theta$$

$$y' = -x \sin \theta + y \cos \theta$$

λ همان طول موج فاکتور کسینوس کرنل فیلتر گابری است. مقادیر قابل قبول برای λ اعداد حقیقی بزرگتر از ۲ هستند. طول موج برابر با ۲ باید در ترکیب با فاز آفست -90° یا $+90^\circ$ به کار رود. برای جلوگیری از رخداد اثرات ناخواسته در مرزهای تصویر مقدار طول موج باید کوچکتر از یک پنجم اندازه‌ی تصویر ورودی باشد. شکل ۳-۴ اثر تغییر طول موج را در باقی شرایط یکسان نشان می‌دهد [86].



شکل (۳-۴) اثر تغییر طول موج را در شرایط یکسان، λ های ۵ و ۱۰ و ۱۵ از چپ به راست

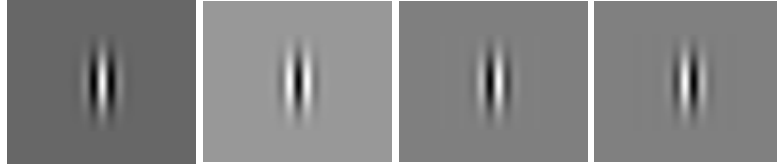
θ یا چرخش میزان چرخش از نوارهای حالت نرمال تصویر را به ما نشان می‌دهد. مقادیر قابل قبول برای θ از صفر تا 360° است. شکل ۳-۵ اثر تغییر θ را در باقی شرایط یکسان نشان می‌دهد.



شکل (۳-۵) اثر تغییر θ را در باقی شرایط یکسان، زاویه چرخش 0° و 45° و 90° از چپ به راست

ϕ یا فاز آفست در فاکتور کسینوس تابع گابری به صورت درجه مشخص می‌شود و مقادیری بین -180° و $+180^\circ$ می‌گیرد. مقدار صفر و -180° به ترتیب به مرکز تقارن مرکز سفید^۳ و مرکز سیاه^۴ پاسخ می‌دهد، در حالی که -90° و $+90^\circ$ مربوط به توابع آنتی متقارن و باقی موارد مربوط به توابع غیر متقارن است. در شکل ۳-۶ اثر تغییر ϕ را در باقی شرایط یکسان می‌بینید.

¹ Daugman
² Neuroscience
³ center_on
⁴ center_off

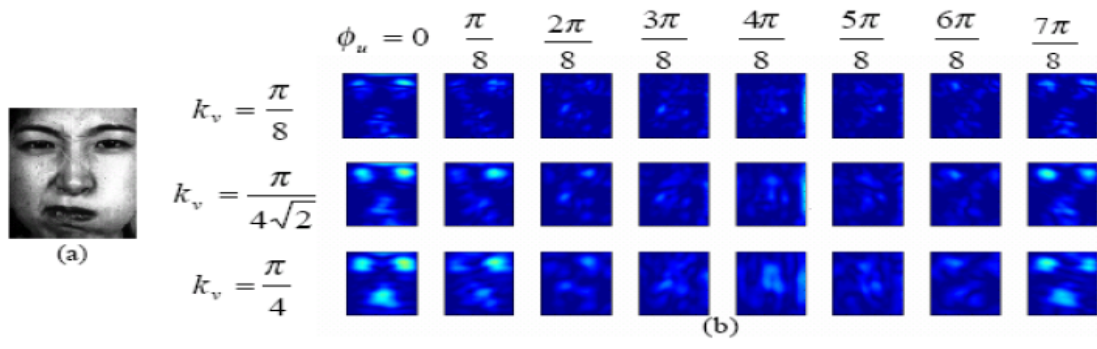


شکل (۳-۶) اثر تغییر φ را در باقی شرایط یکسان، φ های ۰ و ۱۸۰ و ۹۰- و ۹۰ از چپ به راست در بینایی ماشین تجزیه چند فرکانسی کانال به واسطه ی مفهوم چند رزولوشنی تفسیر می شود. به طور کلی آنچه که ما مایل به تشخیص آن هستیم رزولوشن های مکانی متفاوتی دارد و در عمل غیر ممکن است که از قبل رزولوشن بهینه ای را برای تصویر داده شده تعریف کنیم. در تشخیص احساس چهره، آنالیز احساسات به ویژگی های متفاوت که حاصل رزولوشن های متفاوتی است نیاز دارد. به عبارت دیگر مجموعه ای از کرنل های گابر چند فرکانسی، چند چرخشی نمایش ویژگی های مکانی، فرکانسی و چرخشی را برای تصویر و احساس چهره حمایت می کنند. در واقع ویژگی های گابر با اطلاعات رزولوشنی متفاوتی کاراکتر بندی می شود. این یکی از دلایلی است که ویژگی های گابر به صورت موفق در تشخیص چهره و احساس چهره به کار می روند. در هر کانال فرکانس خاصی از یک تصویر چهره، ویژگی ها توسط مجموعه ای از فیلترهای گابر با فرکانس خاص و چرخش های متفاوت استخراج می شوند. نتایج آزمایشات نشان می دهد که ویژگی های استخراج شده به این روش عملکرد و بازدهی بیشتر و حجم محاسبات کمتری نسبت به روش های قدیمی دارند.

ویولت گابر یک تابع گوسین است که توسط توابع نمایی پیچیده ای که بهترین مصالحه را بین زمان-رزولوشن و فرکانس-رزولوشن ایجاد می کنند مدوله می شود. فرم تابع کلی فیلترهای گابر دو بعدی در رابطه 3-11 مشخص شده است.

$$\varphi_{u,v}(x) = \frac{|k_{u,v}|}{\sigma^2} \exp\left(-\frac{|k_{u,v}|^2 |x|^2}{2\sigma^2}\right) (\exp(ik_{u,v}x) - \exp(-\frac{\sigma^2}{2})) \quad (3-11)$$

x مکان پیکسل در حوزه ی مکانی، $k_{u,v}$ مقیاس و چرخش پنجره گوسین و σ پهنا ی باند فیلتر است. ویژگی های گابر از یک تصویر چهره از طریق کانالو آن با گروهی از فیلتر های گابر به دست می آید. تجزیه ی چند رزولوشنی به عبارتی تجزیه تصویر در کانال های فرکانسی با پهنا ی باند ثابت روی یک مقیاس لگاریتمی است. همان طور که گفته شد فیلتر های گابر به سادگی برای استخراج اطلاعات مکانی و فرکانسی قابل تنظیم اند. یک سری از تجزیه تصویر را با استفاده از الگوریتم هرمی گابر در شکل 3-7 می بینید.



شکل (3-7) تجزیه تصویر را با استفاده از الگوریتم هرمی گابر. (a) تصویر چهره ورودی، (b) نمایش اندازه کرنل های گابر بر روی تصویر چهره با سه مقیاس و هشت چرخش

فصل چهارم

مروری بر شبکه های عصبی و آنالیز مولفه های اساسی

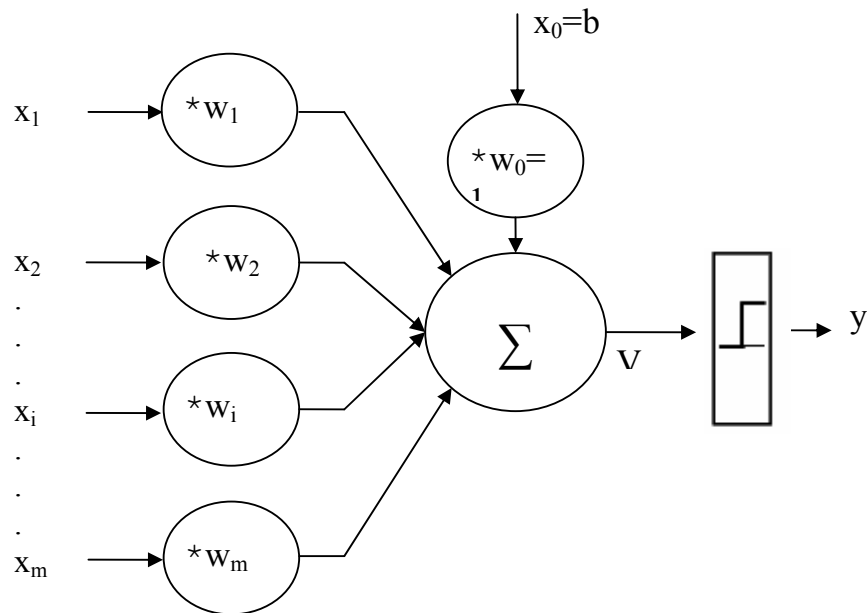
۴-۱-مقدمه

بعد از بررسی های بسیار دریافتیم که اغلب کلاسه بندی احساسات توسط یک الگوریتم بر پایه ی شبکه های هوش مصنوعی انجام می شود. روش کلاس بندی غالب در اکثر کارهای انجام شده در زمینه ی تشخیص احساس شبکه های عصبی می باشد. ولی این روش با مشکلاتی هنگام یک پیاده سازی موفق رو به رو است که مهمترین آن ناتوانی شبکه در پردازش داده ها با حجم بالا است. لیکن در این فصل ابتدا نگاهی به شبکه های عصبی مصنوعی که با توجه به ساختار عصبی بیولوژیکی طراحی شده اند می اندازیم و در آخر آنالیز مولفه های اساسی را که روش موثری برای کاهش حجم داده هاست به طور مختصر توضیح می دهیم.

۲-۴-نگاهی به انواع شبکه های مصنوعی عصبی

۱-۲-۴- شبکه های عصبی پرسپترون

فلسفه اصلی محاسبات شبکه های عصبی این است که با مدل کردن ویژگی های عمده مغز و نحوه عملکرد آن بتوان کامپیوترهایی را ساخت که اکثر ویژگیهای مفید مغز را از خود نشان دهند [90,87]. مغز را می توان به عنوان مجموعه بسیار متصل و شبکه ای از عناصر پردازشی نسبتاً ساده در نظر گرفت. به مدلی نیاز داریم که بتواند ویژگیهای مهم سیستم های عصبی را کسب کند، تا بتواند رفتار مشابهی از خود بروز دهد. ابتدا به مدل سازی یک نرون تنها می پردازیم تا بعداً با در کنار هم قرار دادن تعدادی از این نرون ها یک شبکه عصبی ایجاد کنیم. یک مدل ارائه شده برای نرون به شکل ۱-۴ است.



شکل (۱-۴) مدلی برای نرون

که در آن x_i ورودی i ام است و بردار ورودی را به شکل $X^T = [x_1 \ x_2 \ \dots \ x_i \ \dots \ x_m]$ نمایش می دهیم. $x_0=b$ را بایاس می نامیم و فرض می کنیم ثابت است و w_0 را نیز همیشه برابر یک فرض می کنیم. w_i ها ضرایب وزنی هستند.

در این مدل خروجی دو حالت دارد: فعال (یک) یا غیرفعال (صفر). در این مدل مجموع وزنی ورودی ها محاسبه می شود و سپس با سطح آستانه داخلی مقایسه می شود و چنان که این مجموع وزنی از سطح آستانه تجاوز کرد، خروجی فعال می شود و در غیر این صورت غیر فعال باقی می ماند. چون ورودی ها برای تولید خروجی از میان نرون عبور می کنند به این سیستم پیش خور^۱ می گوئیم. نرون های مدل که

^۱ - Feed-forward

به طریقی ساده به یکدیگر متصل اند در سال ۱۹۶۲ توسط فرانک روزن بلات به نام پرسپترون^۱ نامگذاری شد. او برای نخستین بار نرون های مدل را در کامپیوترهای دیجیتالی شبیه سازی کرد و آنها را به طور رسمی تحلیل نمود.

۴-۲-۲- فراگیری در مدل نرون:

چون این فراگیری از طریق در دست داشتن نتیجه مطلوب راهنمایی می گردد به این نوع آموزش فراگیری با سرپرست می گوئیم. روشی که در اینجا مطرح می کنیم به شکل زیر است.

۱- ضرائب وزنی و مقادیر آستانه را به طور تصادفی تعیین کنید.

۲- یک ورودی را به مدل ارائه دهید.

۳- مقدار خروجی را با توجه به مقایسه ضرائب وزنی ورودی ها و مقدار آستانه محاسبه کنید.

۴- ضرائب وزنی را برای تقویت تصمیمات درست و تضعیف تصمیمات نادرست تغییر دهید، به عبارت دیگر خطا را کاهش دهید.

۵- ورودی بعدی را به مدل ارائه دهید.

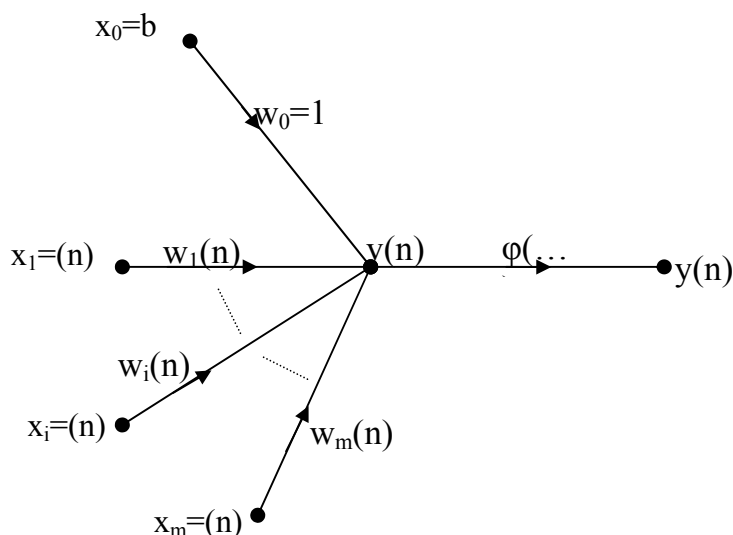
روش فراگیری فوق را می توان توسط الگوریتمی به نام پرسپترون پیاده سازی کرد- برای مشاهده یک مثال به پیوست 2 مراجعه شود- اما ناتوانی الگوریتم پرسپترون در حل مسائل ساده ای چون XOR اولین بار توسط مینسکی و پاپرت در سال ۱۹۶۹ در کتاب "پرسپترون ها" برملا شد. افشای این امر که پرسپترون ها تنها قادر به حل مسائل جدایی پذیری خطی هستند باعث شد که اکثر دانشمندان این رشته را کنار بگذارند. در این حوزه از سال ۱۹۶۹ به بعد که کتاب فوق الذکر به چاپ رسید تقریباً اتفاق مهمی روی نداد، تا اینکه در سال ۱۹۸۶ رومل هارت و مک کلند پرسپترون بهبود یافته را معرفی کردند و میزان علاقه به این رشته را به شدت افزایش دادند. حال به توضیح پرسپترون چند لایه^۲ (MLP) می پردازیم.

۴-۲-۳- پرسپترون چند لایه ای

مدل نرون در حالت پرسپترون چند لایه را می توان به صورت شکل ۴-۲ نمایش داد.

^۱ - Perceptron

3-Multi Layer Perceptron

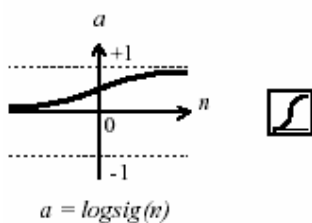


شکل (۲-۴) مدل نرون با تابع انتقال $\varphi(\dots)$

و با توجه به شکل (۲-۴) داریم:

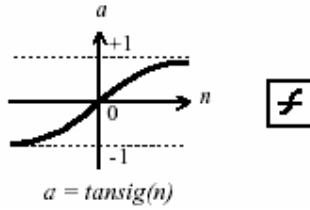
$$y(n) = \varphi(v(n)) = \varphi\left(\sum_{i=0}^m x_i(n)w_i(n)\right) \quad (1-4)$$

در این روش، خروجی هر نرون دیگر تنها یک یا صفر نیست. مدل نرون در این حالت دارای تابع انتقال $\varphi(\dots)$ می باشد. شبکه های عصبی چند لایه ای عموماً از تابع "LOG-SIGMOID" به عنوان تابع انتقال استفاده می کنند. نمودار تابع logsig به صورت شکل ۳-۴ است.



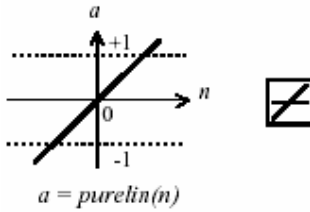
شکل (۳-۴): تابع انتقال Log-Sigmoid

تابع logsig در ازای ورودی های از منفی بی نهایت تا مثبت بی نهایت خروجی های بین صفر و یک را در خروجی نرون ایجاد می کند. به جای تابع logsig می توان از تابع "TAN-SIGMOID" استفاده کرد که در شکل ۴-۴ می بینید.



شکل (۴-۴) تابع انتقال Tan-Sigmoid

تابع دیگری که گاهی به عنوان تابع انتقال نرون استفاده می شود "Linear" است. نمودار این تابع به صورت شکل ۴-۵ است.



شکل (۵-۴) تابع انتقال خطی

چند نمونه از توابع فعال سازی که اغلب به کار می روند در زیر آمده است.

Linear

$$f(\langle w, x \rangle) = w^T x$$

Piecewise Linear

$$f(\langle w, x \rangle) = \begin{cases} 1 & \text{if } \langle w, x \rangle > 1 \\ \langle w, x \rangle & \text{if } |\langle w, x \rangle| \leq 1 \\ -1 & \text{if } \langle w, x \rangle < -1 \end{cases} \quad (۲-۴)$$

Hard Limiter

$$f(\langle w, x \rangle) = \text{sign}(w^T x)$$

Unipolar Sigmoidal

$$f(\langle w, x \rangle) = 1 / (1 + \exp(-w^T x))$$

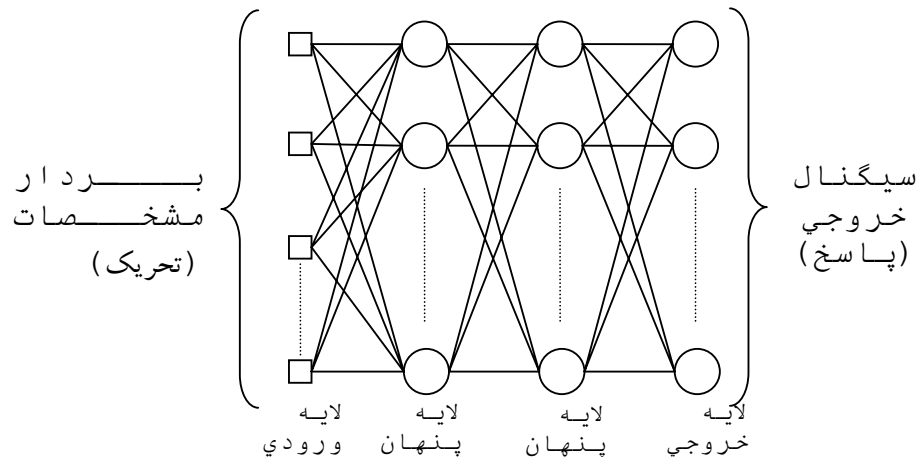
Bipolar Sigmoidal

$$f(\langle w, x \rangle) = \tanh(w^T x)$$

اگر در لایه خروجی یک شبکه MLP از توابع انتقال سیگموئیدی استفاده شود خروجی ها به یک بازه کوچک محدود خواهند شد در صورتی که اگر از تابع خطی استفاده شود خروجی شبکه هر مقداری را می تواند اختیار کند. توابع مذکور رایج ترین توابع مورد استفاده در پس انتشار^۱ هستند. ولی در صورت لزوم هر تابع مشتق پذیری را می توان به عنوان تابع انتقال در نرون های شبکه پس انتشار استفاده کرد. با استفاده از تعدادی از این نرون ها می توان به یک شبکه پرسپترون چند لایه ای دست یافت. شکل ۴-۶ ساختار یک پرسپترون چند لایه ای با دو لایه پنهان^۲ را نمایش می دهد.

¹ Back Propagation

² Hidden Layer



شکل (۴-۶) یک شبکه MLP با دو لایه پنهان

لایه پنهان لایه‌ای است که بین لایه ورودی و لایه خروجی قرار می‌گیرد. ممکن است در یک شبکه پرسپترون چند لایه‌ای یک، دو و یا چند لایه پنهان داشته باشیم. لایه ورودی تغییری بر روی ورودی‌ها اعمال نمی‌کند و تنها از چند گره تشکیل شده است. نحوه عملکرد پرسپترون چند لایه مشابه پرسپترون تک لایه است. بدین صورت که الگویی به شبکه عرضه می‌شود و خروجی آن محاسبه می‌شود، مقایسه خروجی واقعی و خروجی مطلوب باعث می‌گردد که ضرائب وزنی شبکه تغییر یابد به طوری که در دفعات بعد خروجی درست‌تری حاصل گردد.

ضرائب وزنی هر واحد، جداگانه میزان می‌شود و میزان خطا کاهش می‌یابد این عمل در مورد واحدهای لایه خارجی ساده است، زیرا خروجی واقعی و مطلوب آنها را می‌دانیم ولی در مورد لایه میانی چندان روشن نیست و نمی‌دانیم خروجی مطلوب یک نرون مربوط به یک لایه پنهان چگونه است، پس خطا را در آن نمی‌توانیم به طور مستقیم تعیین کنیم تا با توجه به آن ضرائب وزنی مربوط به آن نرون را در راستای کاهش خطا اصلاح کنیم. این مشکلی است که آن را به کمک روشی به نام انتشار به عقب^۱ یا پس انتشار حل می‌کنیم. در قاعده استفاده شده، مقدار تابع خطا محاسبه می‌شود و آن به عقب از یک لایه به لایه پیشین آن انتشار داده می‌شود عبارت پس انتشار به این علت است. در پیوست ۳، تعاریف، قراردادهای، روش تعیین اصلاحات در جهت بیشترین کاهش خطا و روش انتشار به عقب تشریح شده است [88].

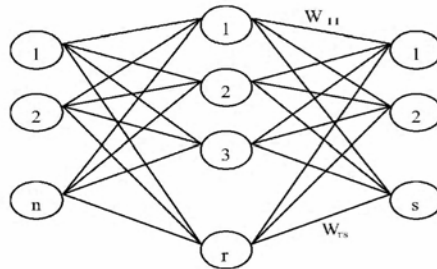
۴-۲-۴- شبکه‌های عصبی RBF^۲

ساختار یک شبکه عصبی RBF در شکل ۴-۷ نشان داده شده است. ساختار یک شبکه عصبی RBF شامل یک لایه ورودی، یک لایه پنهان و یک لایه خروجی با یک معماری پیش‌خور است. لایه ورودی

^۱ Back Propagation

^۲ Radial Basis Function

شبکه مجموعه ای از n واحد است که اجزای بردار مشخصات ورودی را دریافت می کند. واحدهای ورودی کاملاً متصل به r واحد موجود در لایه پنهان می باشند.



شکل (۴-۷): ساختار شبکه عصبی RBF

ارتباطات بین ورودی و لایه پنهان ضریب وزنی واحد دارند و بنابراین لازم نیست آموزش ببینند. در این ساختار، واحدهای پنهان، واحدهای RBF نامیده می شوند. هدف واحدهای RBF دسته کردن داده ها و کاهش ابعاد با استفاده از یک تبدیل غیر خطی و نگاشتن داده ها به یک فضای جدید است. واحدهای RBF به لایه های خروجی کاملاً متصل هستند. لایه خروجی شامل s واحد است که یک ترکیب خطی را در این فضای جدید پیاده سازی می کند. شبکه های عصبی RBF کلاسی از شبکه های عصبی هستند که در آن فاصله ی بین بردار ورودی و یک بردار نمونه، تابع فعال سازی واحدهای پنهان را تعیین می کند. تابع فعال سازی واحدهای RBF به این شکل معادله (۴-۳) بیان می شود.

$$R_i(x) = R_i\left(\frac{\|x - c_i\|}{\sigma_i}\right), \quad i = 1, 2, \dots, r \quad (۳-۴)$$

که در آن x بردار ویژگی n بعدی ورودی، c_i بردار n بعدی است که مرکز واحد RBF نامیده می شود، σ_i عرض واحد RBF و r تعداد واحدهای RBF است. یکی از متداول ترین تابع های فعال سازی برای واحدهای RBF تابع گوسین با بردار میانگین c_i و بردار واریانس σ_i به شکل معادله (۴-۴) زیر است.

$$R_i(x) = \exp\left(-\frac{\|x - c_i\|^2}{\sigma_i^2}\right) \quad (۴-۴)$$

واحدهای خروجی خطی هستند و بنابر این پاسخ z امین واحد خروجی برای ورودی x به صورت معادله (۴-۵) خواهد بود.

$$y_i(x) = \sum_{i=1}^r R_i(x) w_{ij} \quad (۵-۴)$$

که در آن w_{ij} وزن اتصال i ام از واحد RBF به واحد j ام از لایه خروجی است.

۴-۲-۵- شبکه های بازگشتی^۱

در ادبیات شبکه های عصبی، شبکه های عصبی ای که یک یا بیش از یک حلقه فیدبک دارند شبکه های بازگشتی نامیده می شوند. فیدبک در شبکه های بازگشتی که باعث می شود شبکه یک شبکه ترتیبی باشد به شبکه اجازه می دهد رفتار گذرا از خود نشان دهد. در اینجا به دو نمونه از شبکه های بازگشتی یعنی شبکه المان و شبکه هاپفیلد می پردازیم [89].

۴-۲-۵-۱- شبکه های المان

شبکه های المان، شبکه های دو لایه پس انتشاری به علاوه یک اتصال فیدبک از خروجی لایه پنهان به ورودی آن هستند. این مسیر فیدبک به شبکه المان اجازه می دهد که بیاموزد تا الگوهای موقتی همچنین الگوهای فضایی را بشناسد و تولید کند.

۴-۲-۵-۲- شبکه هاپفیلد

شبکه هاپفیلد از تعدادی گره تشکیل شده است که هر گره به تمام گره های دیگر متصل است. هر گره مانند پرسپترون دارای یک سطح آستانه و یک تابع پلکانی است. گره ها مانند پرسپترون مقدار ورودی وزنی را محاسبه کرده، از آن مقدار آستانه را کم می کنند و سپس از تابع پلکانی می گذرانند تا خروجی خود را تعیین کنند. در این شبکه هر گره دقیقاً مانند گره های دیگر است. ورودی های شبکه در یک زمان واحد به کلیه گره ها وارد می شوند. مقدار اولیه هر گره تعیین می شود و سپس شبکه به حال خود رها می شود تا از یک حالت به حالت دیگر چرخش کند. این عمل تا جایی ادامه می یابد که شبکه به حالت ثابتی همگرا شود. در این صورت مقادیر خروجی گره ها دیگر تغییر نمی کند. مجموعه خروجی گره ها در حالت ثابت و پایدار به عنوان خروجی شبکه تلقی می شود.

در این شبکه چون کلیه گره ها به یکدیگر متصل هستند، خروجی هر گره بر تمامی گره های دیگر اثر می کند و ورودی هر گره سعی می کند بر دیگر گره ها اثر کند. به تدریج که شبکه از یک حالت به حالت دیگر حرکت می کند، سعی می کند بین گره ها حالت سازشی را ایجاد کند. نهایتاً در حالت پایدار شبکه به بهترین جواب سازشی دست می یابد. نحوه عمل این شبکه کاملاً با پرسپترون متفاوت است. به شبکه پرسپترون یک بردار ورودی داده می شود و شبکه جواب را به صورت خروجی تولید می کند. لیکن در شبکه هاپفیلد، خروجی اول به صورت ورودی بعدی، مجدداً به سیستم داده می شود که خود خروجی دیگری را باعث می شود و مجدداً این عمل تکرار می شود. جواب نهایی هنگامی حاصل می شود که از یک دوره به دوره دیگر تغییری روی ندهد.

۴-۲-۶- شبکه های با فراگیری بدون سرپرست

همانطور که در روش های پس انتشار خطا دیدیم آموزش با سرپرست به پاسخ های آموزشی خارجی (پاسخهای مطلوب) که برای کلیه داده های ورودی مجموعه آموزشی معلوم است تکیه می کند. در بسیاری از کاربردها بهتر است از شبکه بخواهیم که داده های آموزشی را به تنهایی طبقه بندی کند. برای این منظور باید دو فرض اصلی را در مورد شبکه در نظر بگیریم. اول اینکه عضویت در یک طبقه به طور کلی به معنای داشتن ویژگیهای مشترک است. دیگر اینکه شبکه می تواند این ویژگی های مشترک را در گستره ی داده های ورودی تمیز دهد. برای نمونه ای از شبکه های عصبی که با روش فراگیری بدون سرپرست کار می کنند می توان به شبکه های خود سازمان ده کوهنن^۱ اشاره کرد.

۴-۲-۶-۱- شبکه های خود سازمان ده کوهنن

در این شبکه ها بردار ورودی با بردارهای وزن نرون ها سنجیده می شود و برنده نرونی است که بردار وزن لینک های ورودی آن به بردار ورودی فعلی شبکه نزدیک تر باشد. آموزش به روش دسته بندی نظارت نشده و بر اساس خوشه بندی داده های ورودی استوار است. خوشه بندی به معنای گروه بندی اشیای مشابه و جدا سازی نمونه های متفاوت از هم باشد.

هدف الگوریتم آموزش کوهنن یافتن بردار وزن مطلوب برای m نرون خروجی است. معیار اصلاح وزن به صورت رابطه (۴-۶) می شود.

$$\|x - w_r\| = \min \|x - w_i\| \quad i = 1, \dots, m \quad (۴-۶)$$

که r اندیس نرون برنده است. به عبارت دیگر در این الگوریتم نرون برنده نرونی است که بردار وزن آن، بزرگترین حاصل ضرب داخلی یا همان همسویی را با ورودی فعلی ایجاد کند. رابطه ی (۴-۷) روی این مفهوم تاکید می کند.

$$\langle w_r, x \rangle = \max \langle w_i, x \rangle \quad i = 1, \dots, m \quad (۷-۴)$$

این در حالی است که تمام بردار های وزن یکه هستند. پس از شناسایی نرون برنده باید بردار وزن آن نیز اصلاح شود، تا در هر گام آموزش فاصله $\|x - w_r\|$ کاهش یابد. از این رو گرادیان فاصله را محاسبه می کنیم و داریم:

$$\frac{d\|x - w\|^2}{dw} = -2(x - w) \quad (۸-۴)$$

پس برای قاعده اصلاح وزن خواهیم داشت:

$$w_r(n+1) = w_r + \eta(x - w_r) \quad (۹-۴)$$

^۱ Kohonen

و $\eta < 0$ نرخ یادگیری است و مقداری کوچک معمولاً بین ۰,۱ و ۰,۷ دارد. باید توجه شود که عمل نرمال سازی وزن ها $w_r = \frac{W_r}{\|W_r\|}$ در هر مرحله الزامی است. بقیه نرون ها به حال خود رها می شوند و تغییری در بردار وزن آنها صورت نمی گیرد.

در پایان مرحله آموزش بردارهای وزن نهایی در راستای مرکز جرم کلاس های ورودی خواهند بود. به عبارتی هر نرون بردار وزنی خواهد داشت که مشخص کننده ی یکی از کلاس های ورودی ست. از این رو آموزش زمانی قابل اجراست که کلاس ها یا خوشه های الگوهای ورودی به صورت خطی قابل جداسازی از کلاس های دیگر باشند.

در صورتی که دانش اولیه درباره الگوی ورودی وجود داشته باشد می توانیم الگوریتم کوهنن را به صورت نظارت شده ای در آورد تا بر کارایی آن افزوده شود. در این حالت اصلاح وزن با ضریب η وابسته به پاسخ شبکه خواهد بود. $\eta > 0$ برای پاسخ های صحیح و $\eta < 0$ برای تشخیص های غلط اعمال می شوند. اصلاح دیگر الگوریتم کوهنن این است که بردارهای وزن هم برای نرون برنده و هم برای نرون بازنده تصحیح گردد. معیار این تصحیح وزن باید سطح پاسخ نرون ها باشد. این روش که بازنده ها هم بسته به ارزش خود در بازی شرکت داده می شوند برای خوشه بندی الگوهای پیچیده مفید واقع می شود.

۴-۳- تحلیل مولفه های اساسی (PCA)

۴-۳-۱- مقدمه

تحلیل مولفه های اساسی تکنیکی است که می تواند به عنوان ابزاری برای ساده سازی یک مجموعه دیتا به کار رود. به عبارت دیگر PCA یک تبدیل فضای برداری است که برای کاهش بعد مجموعه داده به ابعاد کمتر از آن، با حذف اجزا غیر اصلی استفاده می شود. این روش در سال ۱۹۰۱ توسط کارل پیرسن^۲ پیشنهاد شد و امروزه از آن به عنوان ابزاری پرکاربرد در تحلیل داده استفاده می شود. تحلیل مولفه های اساسی یک روش تبدیل خطی است که داده ها را به یک سیستم مختصات جدید می برد، به طوری که بزرگترین واریانس داده بر روی اولین محور مختصات قرار می گیرد و دومین بزرگترین واریانس بر روی محور دوم و به همین ترتیب به هر واریانس یک محور مختصات نسبت داده می شود [91,93,94]. این روش آماری در تشخیص چهره و فشرده سازی تصویر کاربرد فراوان دارد. کار اصلی PCA یافتن الگویی برای داده ها در ابعاد بالاست. مزیت PCA این است که بعد از یافتن الگوها، امکان فشرده سازی داده با کاهش تعداد بعد، بدون از دست دادن اطلاعات اساسی وجود دارد.

¹ Principal Component Analysis

² Karl Pearson

۴-۳-۲-مختصری درباره الگوریتم تحلیل مولفه های اساسی

اگر بخواهیم به صورت مختصر پیاده سازی آنالیز مولفه های اصلی را توضیح دهیم می توانیم در شش مرحله آن را خلاصه کنیم .

گام ۱: جمع آوری داده

گام ۲: کسر میانگین از داده ها

گام ۳: محاسبه ی ماتریس کواریانس

گام ۴: محاسبه ی بردارهای ویژه^۱ و مقادیر ویژه^۲ ماتریس کواریانس

گام ۵: کاهش بعد، انتخاب مولفه ها و تشکیل بردار ویژگی^۳

گام ۶: به دست آوردن مجموعه داده های جدید و کاهش بعد در صورت نیاز با حذف بردارهای مربوط به مقادیر ویژه پایین

در این مرحله بردارهای ویژه با بزرگترین مقادیر ویژه را به عنوان اجزا اصلی مجموعه دیتا در نظر می گیریم. چرا که با پیدا کردن بردارهای ویژه و مقادیر ویژه از ماتریس کواریانس، بردارهای ویژه با بزرگترین مقادیر ویژه ابعادی را ایجاد می کند که قوی ترین میزان همبستگی را در مجموعه دیتا دارد و این بردارهای ویژه اجزای اصلی فضای داده ای ما محسوب می شوند [92,97,98,99].

۴-۳-۳-توضیح الگوریتم

در اینجا به توضیح بسیار مختصری برای روشن شدن بیشتر الگوریتم می پردازیم.

واریانس پراکندگی دیتا را در مجموعه داده نشان می دهد. برای محاسبه ی واریانس از رابطه ی ۴-۱۰ سود می جوئیم.

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)} \quad (10-4)$$

کواریانس میزان همبستگی بین دو بعد مختلف را نشان می دهد. کواریانس بین یک بعد و خودش همان واریانس است. به عبارت دیگر کواریانس نشان می دهد هر یک از بعدها چه مقدار از مقدار میانگین با احترام به ابعاد دیگر تغییر می کنند.

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)} \quad (11-4)$$

¹ Eigenvector

² Eigenvalue

³ Feature Vector

ماتریس کواریانس برای سه بعد به شکل زیر است:

$$C = \begin{bmatrix} \text{cov}(X, X) & \text{cov}(X, Y) & \text{cov}(X, Z) \\ \text{cov}(Y, X) & \text{cov}(Y, Y) & \text{cov}(Y, Z) \\ \text{cov}(Z, X) & \text{cov}(Z, Y) & \text{cov}(Z, Z) \end{bmatrix} \quad (12-4)$$

اگر x_i را i امین بردار از مجموعه ی آموزش بدانیم و X مجموعه ای از گروه آموزش باشد که از اتصال بردار های ستونی حاصل شده است، خواهیم داشت:

$$\psi = \left(\frac{1}{N} \right) \sum_{j=1}^N x_j \quad (13-4)$$

$$\psi_i = x_i - \psi \quad (14-4)$$

$$X = \{ \psi_1, \psi_2, \dots, \psi_N \} \quad (15-4)$$

ψ همان میانگین بردارهای ورودی و ψ_i ورودی با میانگین صفر و N تعداد نمونه های مجموعه ی آموزش است. مرحله ی بعد پیدا کردن یک مجموعه از M بردار متعامد است که به بهترین حالت توزیع دیتای ورودی را در یک فضای حداقل خطای مربع بیان می کند. این بردارها به عنوان بردارهای ویژه ماتریس کواریانس X در نظر گرفته می شوند که آنها را با V نشان می دهیم. از تجزیه ی یک بردار ویژگی روی V داریم:

$$V.e_j = \lambda_j.e_j \quad (16-4)$$

که e_j و λ_j به ترتیب بردار ویژه و مقدار ویژه را نشان می دهند. مقادیر ویژه ی بزرگتر به ما واریانس بیشتری مربوط به بردار ویژه می دهد. پس بردار ویژه با مقدار ویژه ی کمتر اطلاعات چندانی ندارد و می تواند حذف شود تا به نمایش ابعادی کمتری دست یابیم. در این حالت لازم به ذکر است که اطلاعات اساسی دیتای ورودی را حفظ کرده ایم.

فرض کنیم M تعداد بردارهای ویژه مربوط به بزرگترین مقادیر ویژه باشد. به منظور تصویر کردن یک بردار دیتای جدید از رابطه زیر استفاده می کنیم:

$$w = e^T (j - \psi) \quad (17-4)$$

که j بردار دیتای جدید برای نگاشت در زیر فضاست و w تصویر را در فضای آیکن¹ نشان می دهد.

۴-۳-۴- کاربرد تحلیل مولفه های اساسی در پردازش تصویر

¹ Eigen Space

استفاده از آنالیز مولفه های اصلی برای فشرده سازی تصاویر به عنوان هتلینگ^۱ یا تبدیل کارونن^۲ و لوو^۳ شناخته شده است. کاهش حجم دیتا یک تکنیک رایج در پردازش تصویر است. تبدیل تصویر از تصاویر رنگی به سیاه سفید از رایج ترین موارد استفاده از PCA است که تصویری با سه رنگ به تصویر تک رنگ حاوی حداکثر اطلاعات تبدیل می شود [95,96].

زمینه ی دیگر استفاده از PCA در پردازش تصویر تعیین چرخش جسم است. ویژگی های PCA اجازه می دهد جهت و یا چرخش قسمت خاصی را در تصویر تعیین کنیم. کاربرد دیگر PCA را می توان یافتن پترن مورد نظری در تصویر قلمداد کرد. در اینجا در شکل ۴-۸ نمونه ای از کاربرد PCA را در فشرده سازی تصویر می بینیم [103].



شکل (۴-۸-الف) تصویر اصلی و اجزا سه رنگش



¹ Hotelling

² Karhunen

³ Leove

شکل (۴-۸-ب) تصاویر حاصل از PCA روی تصویر مورد نظر

Weighted Sum of RGB



1st Component



شکل (۴-۸-ج) مقایسه دو روش مهم در فشردن سازی تصاویر. سمت چپ تصویر حاصل از مجموع وزنی

رنگ های R,G,B و سمت راست تصویر حاصل از پیاده سازی PCA

فصل پنجم

روش های پیشنهادی و پیاده سازی آن ها

۱. 5-1- مقدمه

چهره انسان دارای ساختار بیولوژیکی پیچیده ای است که شکل کلی آن به وسیله ی استخوان جمجمه سر و فک تعیین می شود. با وجود این پیچیدگی، نیاز به مدل کردن کامپیوتری چهره به شدت احساس می شود. تعیین موقعیت چهره در یک تصویر یا فیلم و تشخیص احساس فرد کار بسیار دشواری است و با این وجود در بسیاری از کاربردها ضروری است. احساسات بخش مهمی از ارتباطات روزمره ی ما را تشکیل می دهند. در تعریف احساس اختلاف نظرهای زیادی وجود دارد. احساس تابعی از حافظه و محرک های محیطی است. یک شخص با تغییر شرایط محیطی یا دریافت یک تحریک مشخص عکس العمل نشان می دهد. این عکس العمل فرآیندی است که شامل واکنش های متعدد فیزیولوژیکی است که نماد بیرونی آنها شامل تغییر وضعیت ماهیچه های بدن است که در حالات چهره و آهنگ صدا به نحو قابل توجهی آشکار است. باید دقت کرد بین احساسات، عکس العمل های غیر ارادی و وضعیت روانی تفاوت وجود دارد. عکس العمل های غیر ارادی خیلی کوتاه، کمتر از نیم ثانیه هستند و نمی توان از بروز آنها جلوگیری کرد. وضعیت روانی دارای زمان طولانی تری نسبت به احساس است و اشاره به گرایش شخص در بروز یک احساس دارد. احساسات دارای دوره ی محدودی از ۱ تا ۴ ثانیه اند و یک حالت چهره برای احساس برای چند دقیقه یا ساعت نمی تواند ادامه داشته باشد. پیدا کردن یک مدل عمومی برای چهره و احساسات آن که برای تمام کاربردهای آن مناسب باشد، با در نظر گرفتن این پیچیدگی ها غیر ممکن به نظر می رسد. بنابراین برای هر کاربرد خاص بایستی مدلی متناسب در نظر گرفته شود. ما در این جا به تشخیص احساس روی تصاویر استاتیک چهره می پردازیم و برای استخراج ویژگی های اصلی چهره از روش ویژگی-پایه^۱ استفاده کرده و نقاط برجسته ی تصویر را با تبدیل موجک و الگوریتم اصلاح

¹ Image Based

شده ی لوپپاس به دست می آوریم و اجزا اصلی چهره یعنی چشم و دهان را استخراج می کنیم و پس از استخراج ویژگی، به کلاسه بندی ویژگی های استخراج شده می پردازیم. برای این کار چهار روش پیشنهادی جداگانه که همه آنها هدف یکسان و مسیر نسبتاً مشابهی را دنبال می کنند، پیاده سازی شده است. دو روش اول را روی پایگاه داده FG_Net و دو روش آخر را روی پایگاه داده JAFEE اجرا می کنیم. به طور کلی این چهار روش پیشنهادی به صورت زیر خلاصه می شوند.

الف- نقاط برجسته ی تصویر را با تبدیل موجک توسط الگوریتم اصلاح شده ی لوپپاس به دست می آوریم و اجزا اصلی چهره یعنی چشم و دهان را استخراج می کنیم. برای استخراج ویژگی تحلیل مولفه های اساسی^۱ را روی نواحی استخراج شده اعمال کرده و با مقایسه فاصله ی نگاشت هر تصویر با گروه مبنا به کلاس بندی تصاویر می پردازیم. یک بار هم تحلیل مولفه های اساسی را روی کل چهره اعمال کرده و نتایج را با هم مقایسه می کنیم.

ب- نقاط برجسته ی تصویر را با تبدیل موجک توسط الگوریتم اصلاح شده ی لوپپاس به دست می آوریم و اجزا اصلی چهره یعنی چشم و دهان را استخراج می کنیم. برای استخراج ویژگی از تحلیل مولفه های اساسی^۲ یک بار با یک گروه پایه و بار بعد با دو گروه پایه استفاده کرده، برای کلاس بندی از یک شبکه عصبی پرسپترون چند لایه کمک می گیریم. یک بار هم آنالیز مولفه های اساسی و شبکه عصبی را روی کل چهره اعمال و احساسات را کلاس بندی می کنیم.

ج- نقاط برجسته ی تصویر را با تبدیل موجک توسط الگوریتم اصلاح شده ی لوپپاس به دست می آوریم و اجزا اصلی چهره یعنی چشم و دهان را استخراج می کنیم. برای استخراج ویژگی روی اجزای استخراج شده فیلترهای گابری عبور می دهیم. کرنل های مشخصی از خروجی فیلتر گابری را انتخاب کرده و سپس برای کلاسه بندی حالات مختلف چهره، ابتدا با استفاده از تحلیل مولفه های اساسی^۳ حجم پارامترهای به دست آمده از مرحله ی قبل را کاهش داده و سپس از یک شبکه عصبی پرسپترون چند لایه برای مشخص کردن هر حالت استفاده می کنیم.

د- نقاط برجسته ی تصویر را با تبدیل موجک توسط الگوریتم اصلاح شده ی لوپپاس به دست می آوریم و اجزا اصلی چهره یعنی چشم و دهان را استخراج می کنیم. در مرحله بعد برای کلاسه بندی از یک الگوریتم تلفیقی آماری-شبکه عصبی استفاده می کنیم.

5-2- مشخصه های احساسات شش گانه

سه ناحیه اصلی در چهره که احساسات باعث تغییر در آنها می شوند شامل قسمت بالای چهره (ابروها و چشم)، و قسمت پایین چهره (دهان) می باشد. هر احساس با تغییرات خاص این نواحی در چهره مشخص می شود. مثلاً در ناراحتی تاج ابروها و گوشه ی لب پایین می آید و گوشه ی داخلی پلک بالا می رود و در احساس شادی گوشه های لب به عقب و بالا کشیده می شود. لب ها برای لبخند (شادی کم) بسته و برای خنده (شادی زیاد) باز می شوند. چین و چروک های اطراف بینی از بالا و عقب رفتن لب ها ایجاد

¹ PCA

³ PCA

می شود. گونه ها و پلک پایین ممکن است بالا رود و چروک ایجاد شود. اما شادی بیشتر قسمت پایین چهره را تحت تاثیر قرار می دهد. در اینجا ما به استخراج نواحی اصلی در بالا و پایین چهره، یعنی چشم و دهان می پردازیم.

۲. ۵-۲-۱- استخراج ویژگی های چهره

با توجه به مطالب گفته شده در فصل سوم در این جا با استفاده از تبدیل موجک و روش پیشنهاد شده توسط لویپاس^۱ نقاط برجسته ی چهره را تشخیص می دهیم و اجزا اصلی را جدا می کنیم که این اجزا اصلی همان چشم و دهان هستند [104]. یافتن اجزا اصلی یا نقاط برجسته در یک تصویر یک الگوریتم تکراری موجک است. در مرحله ی ابتدایی، الگوریتم از ضرایب در سطح j شروع می شود و تا رسیدن به بهترین سطح از تصویر ورودی ادامه می یابد. ضرایب ویولت افقی، عمودی و قطری در سطح j را به عنوان نقطه ابتدایی C_j در نظر می گیریم که از زیر مجموعه ی نقاط در سطح $j-1$ محاسبه می شود. این زیر مجموعه از نقاط در سطح $j-1$ فرزندان C_j خوانده می شوند. فرض شده است که فرزندی با مقدار مطلق بالاتر مهمترین فرزند برای C_j باشد و به عنوان نقطه ی ابتدایی برای سطح $j-1$ انتخاب شود. این عملکرد در هر سطح به صورت تکرار شونده انجام می شود تا به سطح یک یا همان تصویر ورودی برسد. به روشی مشابه، ضرایب موجک در سطح یک از یک زیر مجموعه از نقاط در تصویر اصلی به نام فرزندان محاسبه می شوند. در این زمینه فرزندی با گرادین بالاتر به عنوان مهمترین فرزند یا C_0 در نظر گرفته خواهد شد. در اینجا C_0 به عنوان نقطه ی برجسته از نقاط داده شده ی در سطح j شناخته می شود و اندازه ی برجستگی به اندازه زیر است:

$$S = \left(\sum_{k=1}^{j-1} |C_k| \right) + |gradient C_0| \quad (1-5)$$

نقاط برجسته تر نقاطی هستند که ارزش برجستگی بالاتر و مطابق رابطه ۵-۱ مقدار S بزرگتری دارند. با انجام عملکردهای آماری روی نقاط برجسته با زیربنای تبدیل موجک، اجزاء چهره دقیقاً قابل شناسایی است. پیاده سازی تابع برجستگی S بسیار پیچیده است، به این صورت که در ابتدا تبدیل ویولت گسسته دوبعدی را که ویولت مادر آن تابع هار است، بر روی تصویر اصلی پیاده سازی می کنیم و ضرائب تخمین $cA1$ ، و سه بردار ضرائب جزئیات افقی $cH1$ ، عمودی $cV1$ و و قطری $cD1$ در سطح اول به دست می آید. در مرحله بعدی یعنی در سطوح دوم و سوم تبدیل ویولت گسسته دو بعدی را بر روی تخمین به دست آمده از مرحله قبل پیاده می کنیم و ضرائب تخمین و جزئیات سطوح بعدی را بدست می آوریم (معادله ۵-۲). نکته مهم این است که در الگوریتم پیشنهادی توسط لویپاس انتخاب بهترین تعداد سطح و هم چنین بهترین ویولت مادر اشاره نشده است. در این پایان نامه برای رفع ابهام در این مساله استفاده از معیار آنتروپی برای تعیین بهترین تعداد سطح تجزیه پیشنهاد می شود. به این صورت که هر گاه مجموع آنتروپی تخمین ها و جزئیات در سطح بعدی بیشتر از آنتروپی تخمین سطح قبلی باشد عملیات تجزیه متوقف می شود (معادله ۵-۳). همچنین بررسی مقدار ضریب همبستگی به عنوان بهترین معیار برای

¹ Loupias

انتخاب بهترین موجک مادر پیشنهاد می شود. هر گاه میزان همبستگی تصویر و موجک به یک نزدیکتر باشد نوع موجک مادر بهتر است (معادله ۴-۵).

$$[C1A, C1H, C1V, C1D] = dwt[C0, Mw] \quad (۲-۵)$$

$$[C2A, C2H, C2V, C2D] = dwt[C1A, Mw]$$

⋮

$$[CjA, CjH, CjV, CjD] = dwt[Cj-1A, Mw]$$

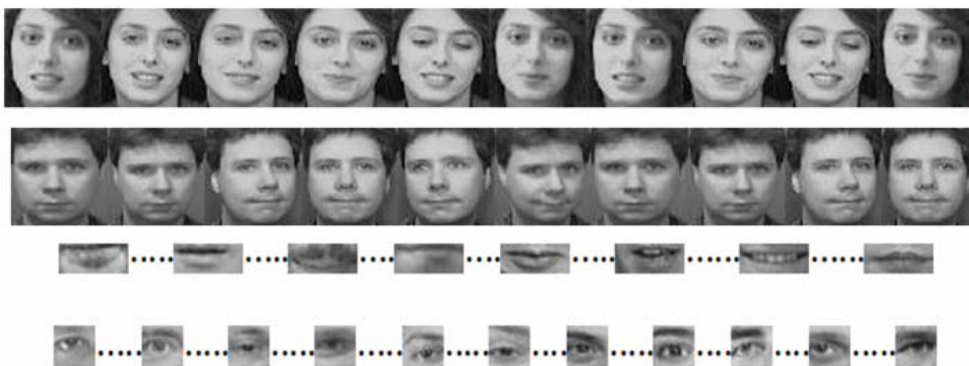
$$E_{SH}(C_j(k)) = -\sum_m C_j^2 \log(C_j^2) \quad (۳-۵)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad r(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2 \right]^{1/2}} \quad (۴-۵)$$

بعد از به دست آوردن جزئیات افقی در سطوح مختلف نقاط برجسته هر سطح و سطح اول را که همان تصویر اصلی است، به دست می آوریم. به این صورت که تابعی به شکل مربع های یک در یک، کل صفحه مقادیر جزئیات در هر سطح را جاروب می کند و مقادیر حداکثر همه آنها را با حفظ مختصاتشان با هم جمع می کند. در مرحله بعد تابعی به شکل مربع های یک در یک، روی گرادیان تصویر اصلی، پیرامون نقطه حداکثر پیدا شده در مرحله قبل، اعمال می کنیم تا در مربع های یک در یک نقطه ای با مقدار حداکثر را تشخیص دهیم. مختصات نقطه مذکور را نگه داشته و مقدار حداکثر به دست آمده را با مقادیر به دست آمده قبلی جمع می کنیم. مقدار Sها را در ستون اول یک ماتریس و مختصات نقاط برجستگی را در ستون دوم و سوم همان ماتریس قرار می دهیم. به این روش ما نقاط برجستگی را روی موجک افقی به دست آورده و ماتریس را مرتب می کنیم.

می دانیم که اجزا اصلی چهره یعنی همان چشم و دهان در تصویر نقاطی هستند که مقدار برجستگی بزرگتری نسبت به سایر نقاط چهره دارند. طبق درصدی که چشم و دهان در چهره اختیار می کنند درصدی از بزرگترین نقاط برجستگی را انتخاب می کنیم و در ماتریسی قرار می دهیم و از بقیه نقاط به دست آمده صرف نظر می کنیم. در مرحله بعد بررسی می کنیم که نقاط برجستگی انتخاب شده در ناحیه ی چشم و دهان که خود به طور دستی و به کمک میانگین آماری حاصل از بررسی تعداد زیادی تصویر تعریف کرده ایم، قرار می گیرند یا نه. از نقاطی که در این نواحی قرار نمی گیرند صرف نظر می کنیم و نقاط باقی مانده را که در ناحیه مذکور قرار گرفته اند، در ماتریس های چشم چپ، چشم راست و دهان قرار می دهیم. نقاط برجستگی دیگری را از جزئیات عمودی و قطری تصویر به دست می آوریم. مراحل طی شده برای حصول این نقاط مشابه مراحل اجرا شده در جزئیات افقی است. نکته قابل توجه در این قسمت این است نقاط برجستگی به دست آمده از طریق جزئیات عمودی و قطری در تشخیص چشم ها به ما کمک می کنند اما نتایج به دست آمده نشان داد که این نقاط در تشخیص دهان مفید نیستند. در مرحله ی آخر برای انتخاب مناسب ترین نقاط برجستگی از انحراف معیار استاندارد استفاده می کنیم. در ابتدا نقطه ی میانگین نقاط برجستگی به دست آمده را پیدا می کنیم. سپس انحراف معیار استاندارد xها و yهای ماتریس های چشم چپ، چشم راست و دهان را به دست می آوریم. نقاطی را که اختلاف

آنها از میانگین به دست آمده کمتر از انحراف معیار استاندارد باشد به عنوان نقاط نهایی نگه می داریم که این نقاط با صحت بالایی چشم و دهان را برای ما مشخص می کند و بقیه نقاط را دور می ریزیم. ما در این بررسی از پایگاه داده تشخیص احساس و ژست FG_Net¹ و پایگاه داده JAFEE استفاده می کنیم. در پایگاه داده تشخیص احساس و ژست FG_Net تلاش شده است که تصاویری که از افراد تهیه می شود در زمانی باشد که احساسات واقعی و نه ساختگی نشان می دهند. به این منظور تصاویری به صورت متوالی و سکانس تصاویر از فرد هنگامی که با تماشای کلیپ های ویدئویی در موقعیت احساس مورد نظر قرار گرفته است تهیه شده است و شامل چرخش محدود سر در تمام جهات است و هفت احساس جهانی را نشان می دهد. شکل ۵-۱ بعضی نمونه های چهره و مکان اجزا تشخیص داده شده را نشان می دهد.



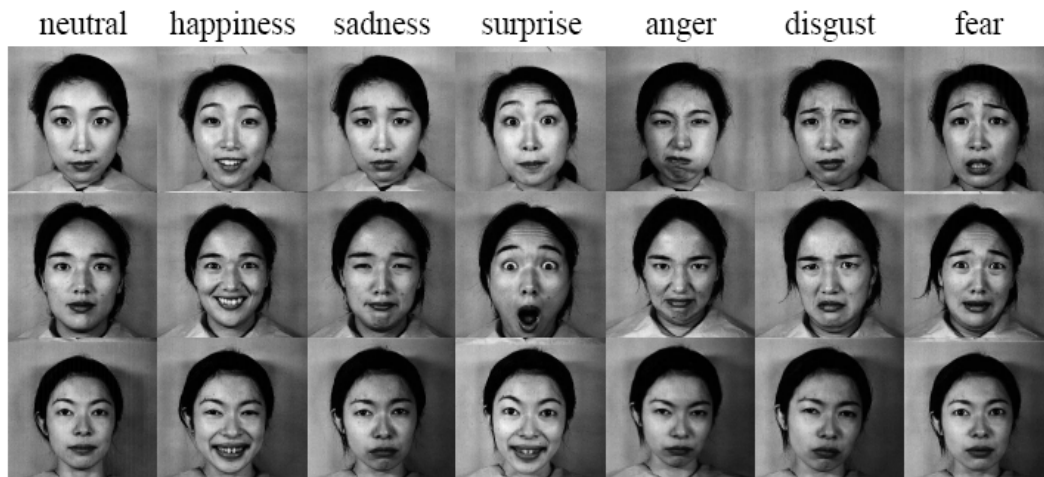
شکل (۵-۱) نمونه هایی از تصاویر پایگاه داده FG_Net و چشم و دهان استخراج شده از طریق تبدیل موجک پایگاه داده JAFEE شامل ۲۱۳ تصویر ۲۵۶ در ۲۵۶ پیکسل از احساسات چهره زنان ژاپنی است که توسط کاماچی^۲ و گیوبا^۳ از دانشگاه کیوشو^۴ تهیه شده است. برای تهیه این پایگاه داده ها از ده نفر خواسته شده است که در مقابل دوربین احساسات مختلفی از خود نشان دهند و وقتی آماده بودند از خود عکس بگیرند. تصاویر اصلی جوری تغییر مقیاس داده شد و برش خورد که چشم افراد تقریباً در یک مکان و با فاصله ی ۶۰ پیکسل از کنار عکس قرار بگیرد. از هر فرد حدود ۲۰ تصویر از هفت احساس مختلف تهیه شد که در شکل ۵-۲ نمونه ای از آن تصاویر را می بینید.

¹ Face and Gesture Recognition Research Network

² Kamachi

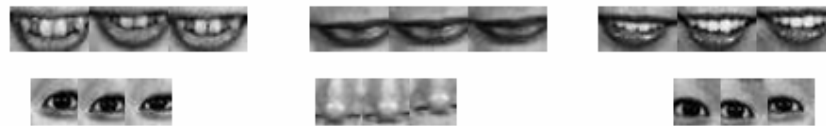
³ Gyoba

⁴ Kyushu



شکل (۵-۲) نمونه ای از آن تصاویر پایگاه داده JAFEE

این تصاویر سیاه سفید بوده و قاب چهره را در دید از روبرو نشان می دهند. استخراج نقاط اصلی چهره روی این تصاویر به نحو بهتری پاسخ می دهد هر چند نتایج حاصل از پایگاه داده FG_Net نیز از صحت بالا و خطای کمی برخوردار است. در شکل ۵-۳ چند نمونه چشم، بینی و دهان های استخراج شده توسط تبدیل موجک از تصاویر پایگاه داده JAFEE را می بینید.



شکل (۵-۳) چند نمونه چشم، بینی و دهان های استخراج شده توسط موجک از پایگاه داده JAFEE در حالت شاد

۵-۳- روش های پیشنهادی جهت تشخیص احساسات

۳-۵-۱- روش تشخیص احساسات توسط آنالیز مولفه های اساسی و معیار فاصله

طبق شرح مختصری که درباره تحلیل مولفه های اساسی در فصل پنجم عرضه شد دیدیم که PCA روشی است برای معرفی کردن الگوهایی در یک مجموعه داده که به کمک این الگوها شباهت ها و تفاوت های موجود بین داده ها را می توانیم بهتر بررسی کنیم. از آنجا که پیدا کردن الگو در داده هایی با بعد بالا دشوار است، PCA ابزاری توانمند برای آنالیز داده ها محسوب می شود. مزیت اصلی دیگر PCA این است که هنگامی که الگوها را در مجموعه داده پیدا کردیم با کاهش تعداد بعد مجموعه داده، بدون از دست دادن اطلاعات اساسی قادر به فشرده سازی داده خواهیم بود. این تکنیک در فشرده سازی تصاویر به شدت استفاده می شود.

در بخش قبل به کمک تبدیل موجک یک سری نواحی اساسی یعنی چشم و دهان را مشخص کردیم. با استفاده از PCA برای هر یک از این نواحی می توانیم مقادیری به دست آوریم که این مقادیر، ویژگی های استخراج شده برای کاربرد در مرحله ی کلاسه بندی است. آنالیز مولفه های اساسی را می توانیم روی کل تصویر چهره و یا روی نواحی استخراج شده چشم و دهان پیاده کنیم.

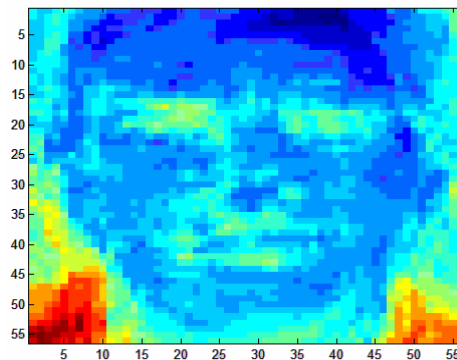
در این جا تصاویر موجود در پایگاه داده FG_Net را در ۴۰ پوشه دسته بندی کرده ایم که هر پوشه حاوی ۱۰ تصویر از فرد است که در وضعیت های متفاوت، حالت های شادی و ناراحتی و سایر حالت ها را نشان می دهد، اما هر پوشه یک حالت را به شدت های متفاوت و با چرخش های محدود سر در تمام جهات گزارش می کند. ما کل حالت ها را به سه دسته ی شادی و ناراحتی و سایر حالت ها برچسب گذاری کرده ایم. در شکل ۴-۵ نمونه ای از تصاویر مورد بررسی را می بینید.



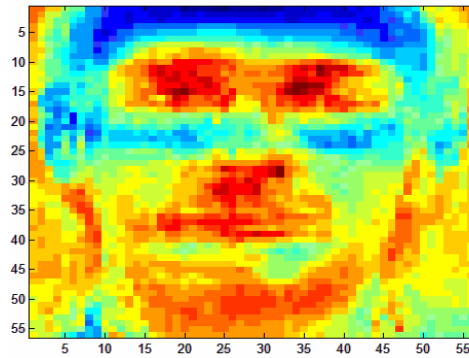
شکل (۴-۵) نمونه ای از تصاویر مورد بررسی از پایگاه داده FG_Net

از ۱۰ پوشه نخست، اولین تصویر را انتخاب کرده که حالت های شاد و غمگین و غیر از آن را شامل می شود و از این ۱۰ تصویر برای به دست آوردن پایه های PCA استفاده می کنیم. از روی ۱۰ تصویر که برای ساختن پایه های PCA انتخاب شده اند بردارهای ویژه و مقادیر ویژه و پایه های مورد نظر را به دست می آوریم.

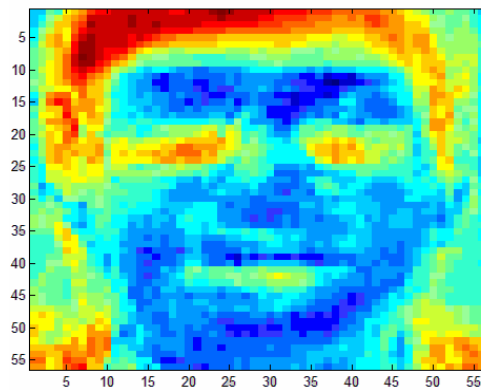
از آنجا که تصاویر ۵۶ در ۵۶ پیکسل هستند کل پایه های موجود برابر با ۵۶×۵۶ می باشد که ما ۱۰۰ پایه نخستین را به عنوان مهمترین پایه ها در نظر می گیریم و تمامی ۴۰۰ تصاویر را روی این ۱۰۰ پایه تصویر یا نگاشت می کنیم که برابر با ضرب داخلی بردار هر تصویر در یکی از بردارهای پایه است. در شکل (۵-۵ الف تا د) تصویر چهار پایه نخست از پایگاه داده FG_Net می بینیم.



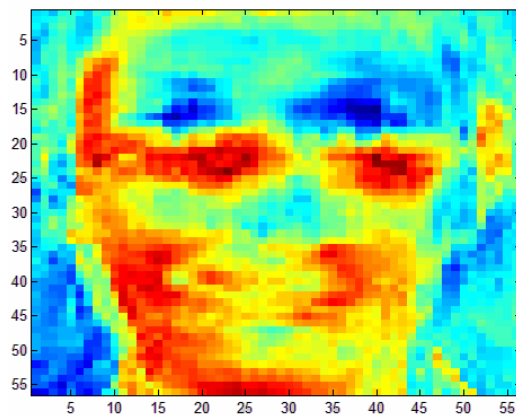
شکل (۵-۵ الف): تصویر پایه ی اول در PCA از پایگاه داده FG_Net



شکل (۵-۵-ب): تصویر پایه ی دوم در PCA از پایگاه داده FG_Net



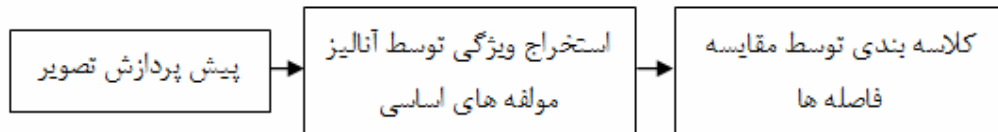
شکل (۵-۵-ج) تصویر پایه ی سوم در PCA از پایگاه داده FG_Net



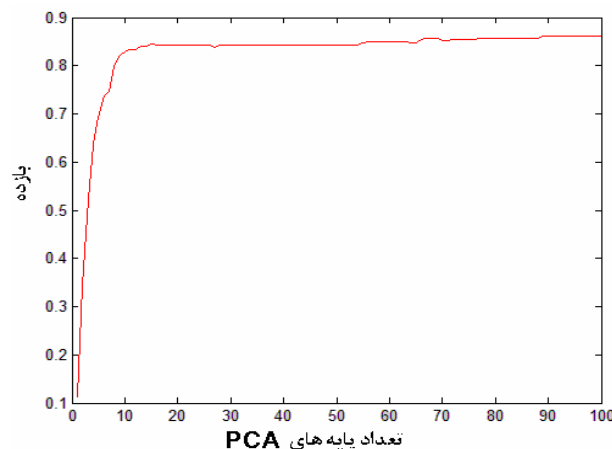
شکل (۵-۵-د) تصویر پایه ی چهارم در PCA از پایگاه داده FG_Net

در مرحله ی اول از هر پوشه تصویر میانگینی به دست می آوریم و نگاشت آن ها را روی بردار های پایه در ماتریسی به نام کلاس ضرایب میانگین ذخیره می کنیم که ماتریسی با ۴۰ سطر در ۱۰۰ ستون است و این ماتریس را به عنوان ماتریس مبنا برای مقایسه در نظر می گیریم. سپس نگاشت حاصل از کل تصاویر موجود را روی بردارهای پایه محاسبه کرده و در ماتریس ضرایب داده ها ذخیره می کنیم که ماتریسی ۴۰۰ در ۱۰۰ خواهد بود. می دانیم که هر سطر این ماتریس متعلق به یک تصویر در پایگاه داده ها است. برای کلاس بندی احساسات محاسبه می کنیم که هر سطر با کدام سطر در ماتریس کلاس ضرایب

میانگین حداقل فاصله را دارد. هر سطر در ماتریس کلاس ضرایب میانگین معرف و متعلق به یک کلاس احساسی است که ما از قبل آن را می دانیم. با یافتن حداقل فاصله مشخص می کنیم که تصویر مورد نظر متعلق به کدام گروه احساسی است و می توانیم تصاویر را کلاسه بندی کنیم که با دقت ۸۶٪ به ما پاسخ صحیح می دهد. در شکل ۵-۶ بلوک دیاگرام سیستم تشخیص احساس با الگوریتم آنالیز مولفه های اساسی روی کل چهره و در شکل ۵-۷ نمودار بازدهی سیستم کلاسه بندی به کمک PCA را می بینید.



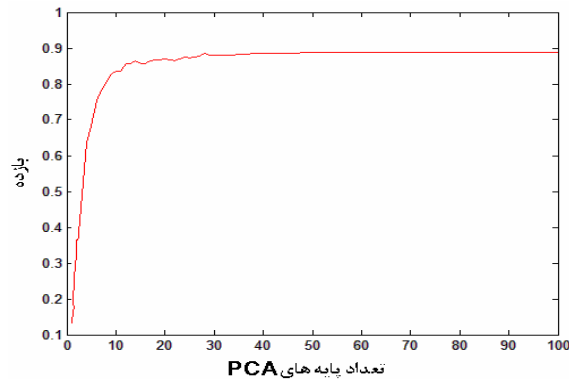
شکل (۵-۶) بلوک دیاگرام سیستم تشخیص احساس با الگوریتم آنالیز مولفه های اساسی روی کل چهره



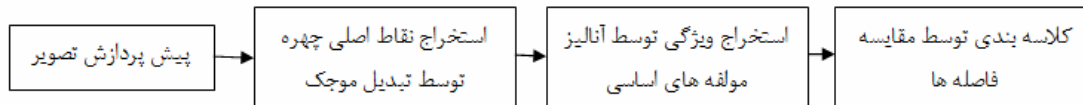
شکل (۵-۷) نمودار بازده کلاس بندی بر حسب تعداد پایه های PCA برای الگوریتم تشخیص توسط PCA روی کل چهره در مرحله دوم به جای آنکه میانگین تصاویر هر پوشه را مبنای مقایسه قرار دهیم تصاویر پنجم از هر پوشه را که احساس آن پوشه را به خوبی نشان می دهد مبنای مقایسه قرار می دهیم و نگاشت آن ها را روی بردارهای پایه در کلاس ضرائب میانگین قرار می دهیم، سپس نگاشت حاصل از کل تصاویر موجود را روی بردارهای پایه محاسبه کرده و در ماتریس ضرائب داده ها ذخیره می کنیم و سطری از ماتریس مبنای که حداقل فاصله را با هر سطر ماتریس دیتا دارد جست و جو می کنیم تا کلاس مربوط به هر تصویر را تشخیص دهیم. کلاسه بندی کننده به این روش با دقت ۸۴٪ به ما پاسخ صحیح می دهد. می بینیم که میزان پاسخ دهی این سیستم نسبت به حالت قبل به جای افزایش، کاهش یافته است و در می یابیم که تصاویر میانگین هر پوشه نسبت به یک تصویر از هر پوشه میزان اطلاعات بیشتری درباره احساس فرد در خود نگه می دارد و چون هر پوشه حاوی تصاویری از یک احساس با شدت های متفاوت است تصاویر میانگین انعطاف پذیری بیشتری به سیستم برای تشخیص احساس صحیح می دهند.

با رسم نمودار بازده کلاسه بندی بر حسب تعداد پایه های PCA همان طور که در شکل ۵-۷ مشخص است، ملاحظه می شود که در PCA موجود در مرحله ی اول و مرحله ی دوم، با حداکثر ۴۰ پایه ی اول، همگرا شده و نیازی به ۱۰۰ پایه نیست و تنها با استفاده از ۴۰ پایه می توانیم حداکثر اطلاعات مورد نیاز خود را نگه داریم و به نحو قابل ملاحظه ای حجم محاسبات خود را کاهش دهیم.

تصاویر چشم ما ماتریسی ۲۱ در ۲۱ و تصاویر دهان ماتریسی ۳۵ در ۲۰ را تشکیل می دهند. ما ۴۰ پوشه از تصاویر افراد مختلف و در حالات مختلف داریم که هر پوشه شامل ۱۰ تصویر از فرد در نماهای متفاوت و چرخش های محدود سر است. تمام مراحل گفته شده ی قبل را روی چشم ها و دهان استخراج شده اعمال می کنیم و برای کلاسه بندی تصاویر می کوشیم. نتایج به دست آمده حاکی از نرخ تشخیص احساس حدود ۸۶٪ برای چشم ها و ۸۹٪ برای لب است که تا حدودی قابل قبول است.



شکل (۸-۵) نمودار بازده کلاس بندی بر حسب تعداد پایه های PCA برای الگوریتم تشخیص توسط PCA روی لب در شکل ۸-۵ نمودار بازده لب را بر حسب پایه های PCA و در شکل ۹-۵ بلوک دیاگرام سیستم تشخیص احساس با الگوریتم آنالیز مولفه های اساسی روی اجزای چهره را می بینید. در مرحله بعد می کوشیم با تغییر دادن نوع سیستم کلاس بندی بازده را افزایش دهیم.



شکل (۹-۵) بلوک دیاگرام سیستم تشخیص احساس با الگوریتم آنالیز مولفه های اساسی روی اجزای چهره

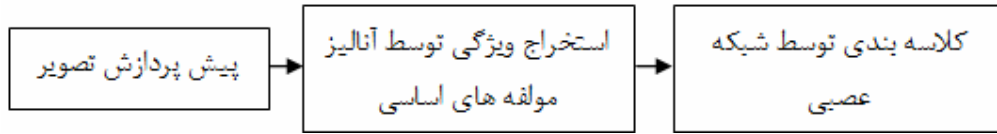
در بررسی بیشتر این روش به علت بازده به نسبت بالا و روش ساده آن متوجه شدیم که این روش در جداسازی احساسات افراد چندان قابل اعتماد نیست. چرا که ما انتظار داریم بردار خروجی PCA شاد و بردار خروجی PCA غمگین از هم فاصله داشته باشند. هم چنین انتظار داریم بردارهای خروجی از حالت شاد کنار هم و بردارهای خروجی از حالت غم نیز در نزدیکی هم قرار بگیرند. اما متوجه شدیم نگاهیست هر ۱۰ تصویر از هر پوشه به شدت اطراف نگاهیست میانگین همان پوشه متمرکز می شود و نتیجه می گیریم که این روش به قیافه و شکل چهره فرد بیشتر از نوع احساسی که نشان می دهد وابسته است و از این رو برای کلاسه بندی احساسات چندان قابل استناد نمی باشد. با این احوالات ما بر آن شدیم که برای کلاس بندی به سراغ شبکه عصبی برویم.

۵-۳-۲- روش تشخیص احساس توسط آنالیز مولفه های اساسی-شبکه عصبی

در این جا دو روش متمایز آنالیز مولفه های اساسی، یکبار با یک گروه پایه و یکبار با دو گروه پایه، و شبکه های مصنوعی عصبی را با هم تلفیق می کنیم و به بررسی راندمان هر یک می پردازیم.

۵-۳-۲-۱- روش آنالیز مولفه های اساسی با یک گروه پایه-شبکه عصبی

در این روش نگاهت ۴۰۰ تصویر چهره را روی پایه های PCA اول که بازده بالاتری دارد به عنوان ورودی های شبکه عصبی استفاده می کنیم و تلاش می کنیم به کمک شبکه عصبی داده های موجود را کلاسه بندی کنیم که بلوک دیاگرام سیستم را در شکل ۵-۱۰ می بینید.



شکل (۵-۱۰) بلوک دیاگرام سیستم تشخیص احساس با الگوریتم آنالیز مولفه های اساسی و شبکه عصبی روی کل چهره ما در این جا از شبکه عصبی پرسپترون چند لایه استفاده می کنیم. شبکه ما یک شبکه عصبی پس انتشار چند لایه و بازخورد از جلوست که مجموعه داده های ورودی را به یک مجموعه داده ی خروجی نگاهت می کند. این شبکه تعمیمی از پرسپترون خطی است که از سه لایه یا بیشتر و توابع فعال سازی غیر خطی استفاده می کند. پرسپترون چند لایه در فاز آموزش الگوریتم انتشار به عقب را که الگوریتمی استاندارد برای تشخیص الگو با ناظر است، به کار می برد. هر گره در لایه های ورودی و پنهان با یک وزن مشخص به تمام گره های لایه های بالاتر متصل می شود. یادگیری در پرسپترون با تغییر وزن ها پس از پردازش هر قطعه از داده، بر اساس مقدار خطای خروجی در مقایسه با نتیجه ی واقعی انجام می گیرد.

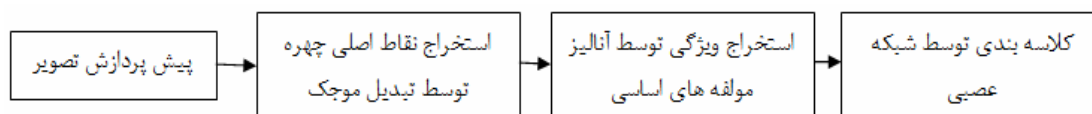
شبکه عصبی ما در این قسمت متشکل از ۴۰ گره در ورودی و ۱ گره در خروجی و دو لایه ی پنهان، با ۸۰ گره در لایه ی پنهان اول و ۴۰ گره در لایه ی پنهان دوم است. در این جا برای آموزش شبکه از روش کراس ولیدیشن^۱ استفاده می کنیم. در ابتدا از ۴۰۰ بردار خروجی PCA به عنوان ورودی شبکه عصبی، ۳۰۰ بردار را برای آموزش شبکه و ۱۰۰ بردار دیگر را برای آزمایش نهایی در نظر می گیریم. ۳۰۰ بردار اختصاص یافته برای آموزش شبکه را به ۶ گروه ۵۰ تایی داده تقسیم می کنیم و در هر بار آموزش از ۵ گروه یعنی ۲۵۰ داده برای آموزش شبکه و از گروه آخر برای آزمایش شبکه سود می جوئیم و بار بعد ۵۰ داده ی دیگر برای آزمایش و ۲۵۰ داده ی باقی مانده برای آموزش شبکه استفاده می شود و به همین ترتیب بعد از ۶ بار آموزش و آزمایش، مطمئن خواهیم بود که همه ی ۳۰۰ داده موجود برای آزمایش استفاده شده است و همین طور این ۳۰۰ داده بارها برای آموزش شبکه به کار رفته اند. این روش به ما کمک می کند که هم دقت و هم تعمیم پذیری شبکه عصبی را بالا ببریم، چرا که هر بار آموزش با گروهی داده و آزمایش با دسته ای دیگر و سپس چرخش اینها به شبکه عصبی توانایی تعمیم پذیری و بهتر پاسخ دهی را می دهد. در هر بار آموزش کردن توسط دسته داده ی ۵۰ تایی ما خطایی به دست می آوریم و پس از ۶ بار انجام آموزش و آزمایش، خطای نهایی را از میانگین گیری خطاهای به دست آمده برای آزمایش محاسبه می کنیم. می توانیم برای افزایش دقت، بهبود آموزش شبکه و افزایش تعمیم پذیری آن، این پروسه را بارها و بارها انجام دهیم. معمولاً این کار را ۱۰۰ بار انجام می دهند. خطای نهایی آموزش از میانگین گیری خطای به دست آمده از این تکرارها حاصل می شود. برای به دست آوردن خطای آموزش روی ۱۰۰ داده آزمایش را انجام می دهیم و میانگین خطا را محاسبه می

^۱ cross validation

کنیم. این کار را هم بارها و بارها انجام می دهیم و خطای آزمایش از میانگین گیری خطای آزمایش به دست آمده در هر مرحله حاصل می شود. پس از محاسبه ی خطای میانگین، انحراف معیار استاندارد آن را هم محاسبه می کنیم.

ما در اینجا یک بار شبکه عصبی را برای حالت شاد، یک بار برای حالت ناراحت و یک بار برای سایر حالت ها آموزش می دهیم. نرخ تشخیص میانگین برای حالت غمگین ۸۵٪ با انحراف معیار استاندارد ۱۶٪ و نرخ تشخیص میانگین برای سایر حالت ها ۸۴٫۲٪ با انحراف معیار استاندارد ۱۲٪ به دست می آید. نرخ تشخیص میانگین برای حالت شاد ۸۲٪ با انحراف معیار استاندارد ۹٪ خواهد بود که کمترین نرخ تشخیص نسبت به دو کلاس دیگر است. با مقایسه نتایج حاصل از شبکه عصبی و نتایج حاصل از کلاسه بندی توسط PCA می بینیم که شبکه عصبی رشد چندان در نرخ تشخیص به ما نداده است هر چند این مقدار بازده در زمینه ی تشخیص احساس چهره مطلوب قلمداد می شود.

برای بهبود بازدهی سیستم در این مرحله ما به پردازش نتایج حاصل از قسمت اول بررسی خود می پردازیم و انتظار داریم که بازده بیشتری به دست آوریم. در شکل ۵-۱۱ بلوک دیاگرام سیستم مورد استفاده را می بینید. در قسمت نخست ما توانستیم به کمک تبدیل موجک و اعمال الگوی لویپاس درباره نقاط برجسته، بخش های اصلی چهره، یعنی چشم و دهان را از تصویر جدا کنیم. با این کار ما بسیاری از نواحی چهره که عملاً فاقد اطلاعات ارزنده ای برای ماست حذف می کنیم و به نحو چشم گیری حجم محاسبات را کاهش می دهیم. هم چنین پردازش اطلاعات موجود در این نواحی اصلی بر دقت پاسخ دهی سیستم ما می افزاید چرا که بسیاری از اطلاعات غیر اصلی مانند اطلاعات مربوط به نواحی مو و گردن و بینی و یا آشفتگی ناشی از پس زمینه که باعث گیج شدن سیستم و کاهش راندمان آن در کلاسه بندی کننده داده ها و تشخیص احساس می شود، حذف می گردد.



شکل (۵-۱۱) بلوک دیاگرام سیستم تشخیص احساس با الگوریتم آنالیز مولفه های اساسی و شبکه عصبی روی اجزای چهره

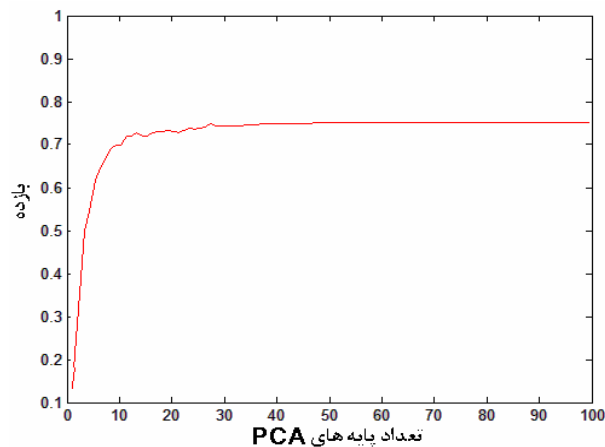
با اعمال شبکه عصبی روی خروجی PCA از ماتریس الحاقی چشم و دهان به نتیجه ی قابل قبول ۸۶٪ می رسیم. با آموزش دادن شبکه با چشم تنها به نتیجه ۸۱٪ و با آموزش شبکه با دهان تنها به نتیجه ی ۸۴٪ می رسیم. طبق آنچه پیش بینی کردیم شبکه عصبی به کمک مجموع اطلاعات چشم و دهان بازدهی بیشتری دارد هر چند سرعت شبکه کند و آموزش و همگرایی آن دشوار می شود.

۵-۳-۲- روش آنالیز مولفه های اساسی با دو گروه پایه-شبکه عصبی

پایگاه داده ها FG_Net شامل تصاویری از جهات مختلف و با چرخش های محدود سر است و این شرایط کار تشخیص احساس را اندکی دشوار تر می کند. ما برای بررسی میزان بازدهی سیستم خود روی تصاویر چهره با دید از روبرو به سراغ پایگاه داده هایی معروف JAFEE می رویم که شامل ۲۱۳ تصویر ۲۵۶ در

۲۵۶ پیکسل از احساسات چهره زنان ژاپنی است که توسط کاماچی^۱ و گیوبا^۲ از دانشگاه کیوشو^۳ تهیه شده است.

در ابتدای کار مانند قسمت قبل می‌کوشیم با اعمال PCA و شبکه عصبی روی کل تصویر چهره به کلاسه بندی تصاویر پردازیم تا بتوانیم نتایج را با مرحله ای بعد که چشم‌ها و دهان را از چهره استخراج و سپس کلاسه بندی می‌کنیم مقایسه کنیم. در این جا ما تصاویر را به سه پوشه ی ۳۰ تایی تصویر تقسیم می‌کنیم که یک پوشه حالت شاد، یک پوشه حالت غمگین و یک پوشه سایر حالت‌ها را نشان می‌دهد. همان طور که در شکل ۵-۱۲ می‌بینیم که ۹۰ درصد اطلاعات با تنها ۱۵ پایه قابل حفظ شدن است و PCA از پایه ۱۵ به بعد همگرا شده است.



شکل (۵-۱۲): نمودار بازده کلاسه بندی بر حسب تعداد پایه‌های PCA برای الگوریتم تشخیص توسط PCA دو پایه روی کل چهره

برای اصلاح نتایج و افزایش بازدهی روش قبل، برای کلاس شاد یک مجموعه پایه ی PCA و برای کلاس غمگین نیز یک مجموعه پایه ی PCA می‌سازیم. از شش تصویر غمگین نشان داده شده در شکل ۵-۱۳ پایه‌های PCA غمگین و از شش تصویر شاد نشان داده شده در شکل ۵-۱۴ پایه‌های PCA شاد را می‌سازیم و تنها ۱۵ پایه ی اول را نگه می‌داریم. نگاشت یا همان ضرب داخلی شش تصویر شاد و غمگین را روی پایه‌های PCA محاسبه می‌کنیم که دو ماتریس ۶ در ۱۵، به نام‌های ماتریس ضرایب شاد برای کلاس شاد و ماتریس ضرایب غمگین برای کلاس غمگین خواهیم داشت. از این دو ماتریس که به عنوان ماتریس‌های مبنا از آن‌ها استفاده خواهیم کرد میانگین گرفته و در دو ماتریس ۱ در ۱۵، به نام‌های ماتریس ضرایب میانگین کلاس شاد برای کلاس شاد و ماتریس ضرایب میانگین کلاس غمگین برای کلاس غمگین ذخیره می‌کنیم.

¹ Kamachi
² Gyoba
³ Kyushu

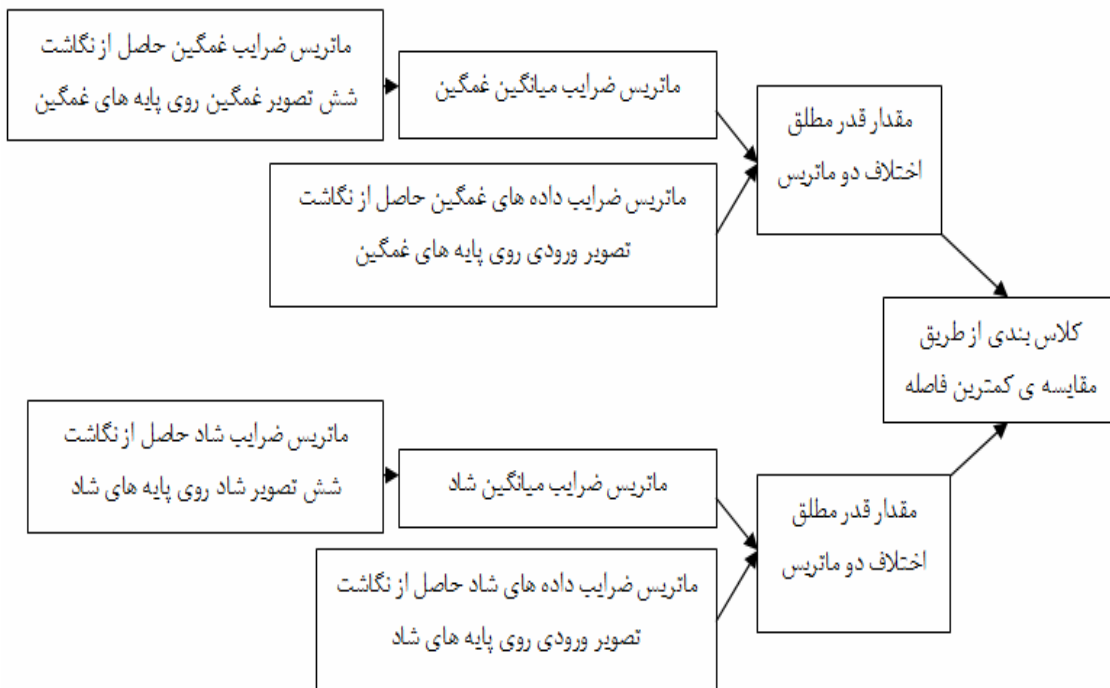


شکل (۵-۱۳): تصاویر استفاده شده برای ساختن پایه ها برای کلاس غمگین



شکل (۵-۱۴): تصاویر استفاده شده برای ساختن پایه ها برای کلاس شاد

برای هر تصویر ورودی نگاشت یا ضرب داخلی آن را یک بار روی ۱۵ پایه کلاس شاد و بار دیگر روی ۱۵ پایه کلاس غمگین به دست می آوریم که به ترتیب در دو ماتریس ۱ در ۱۵ به نام های ماتریس ضرایب داده های شاد و ماتریس ضرایب داده های غمگین ذخیره می کنیم. سپس قدر مطلق مقدار اختلاف بین ماتریس ضرایب میانگین کلاس شاد و ماتریس ضرایب داده های شاد برای کلاس شاد، و هم چنین قدر مطلق مقدار اختلاف بین ماتریس ضرایب میانگین کلاس غمگین و ماتریس ضرایب داده های غمگین برای کلاس غمگین را محاسبه کرده و با هم مقایسه می کنیم. مقدار کمتر اختلاف در یک کلاس نمایان گر تعلق تصویر به آن کلاس احساسی است. این روش ساده که بلوک دیاگرام آن را در شکل ۵-۱۵ می بینید ۷۸٪ بازدهی دارد و همان طور که قبلا دیدیم استفاده از PCA به تنهایی برای کلاس بندی احساسات چهره اگرچه روشی ساده اما چندان مطلوب نیست.



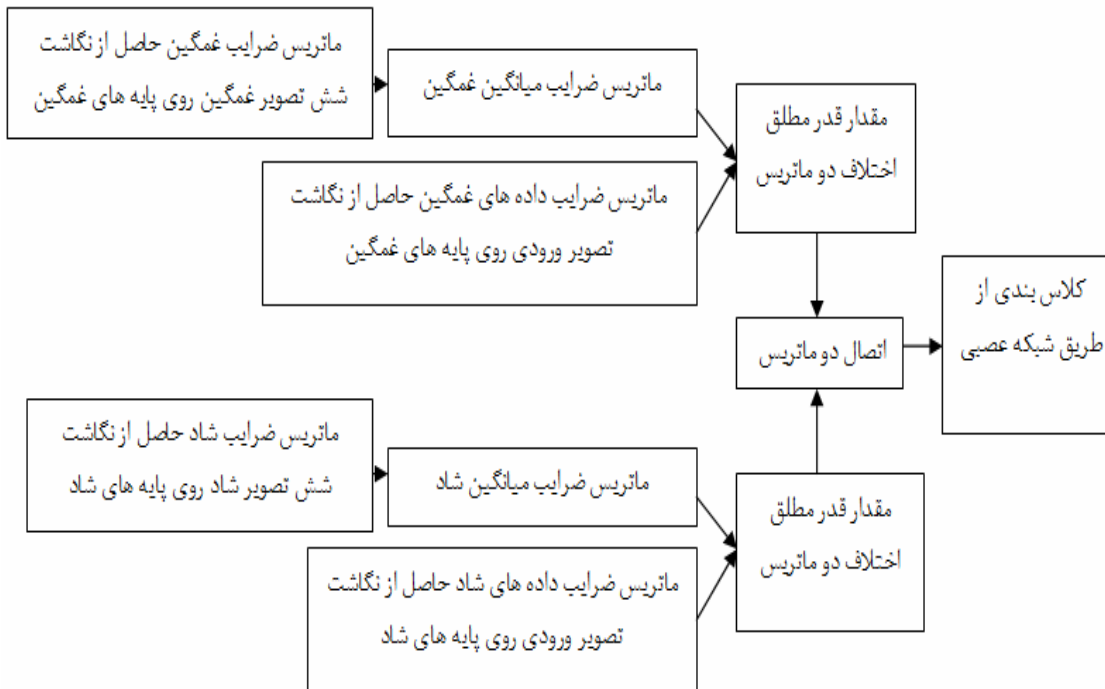
شکل (۵-۱۵) بلوک دیاگرام روش تشخیص احساس توسط آنالیز مولفه های اساسی با دو گروه پایه-معیار فاصله در توسعه ی میزان بازدهی، خروجی های PCA را به شبکه عصبی می دهیم. به این صورت که پس از محاسبه ی قدر مطلق مقدار اختلاف بین ماتریس ضرایب میانگین کلاس شاد و ماتریس ضرایب داده های شاد برای کلاس شاد، و هم چنین قدر مطلق مقدار اختلاف بین ماتریس ضرایب میانگین کلاس غمگین و ماتریس ضرایب داده های غمگین برای کلاس غمگین دو ماتریس ۱ در ۱۵ به دست می آوریم و آن ها را به هم متصل کرده و یک ماتریس ۱ در ۳۰ به نام ماتریس الحاقی می سازیم. سپس این ماتریس را مقیاس^۱ کرده و به عنوان ورودی به شبکه عصبی می دهیم. شبکه عصبی که در این جا استفاده کرده ایم پرسپترون سه لایه ی پس انتشار بازخورد از جلوست و ۳۰ گره در لایه ورودی، ۳۰ گره در لایه پنهان و یک گره خروجی دارد که دو کلاس شاد و غمگین را برای ما جدا می کند.

برای ارزیابی مقدار بازدهی سیستم، آن را توسط کراس ولیدیشن مدل k-fold آزمایش می کنیم. روش کار به این صورت است که در هر دور تصاویر شاد و غمگین ۲۹ نفر، که طبق آنچه گفتیم از هر فرد سه تصویر شاد و سه تصویر غمگین موجود است، را برداشته و برای آموزش شبکه عصبی به کار می بریم و با تصاویر شاد و ناراحت یک نفر باقی مانده آزمایش می کنیم. پس از ۳۰ دور می دانیم که شبکه با تمام داده ها آموزش یافته و با تمام تصاویر متعلق به افراد مختلف نیز آزمایش شده است. برای افزایش دقت، انعطاف پذیری و تعمیم پذیری ما تعداد دور ها را ۱۲۰ می گذاریم تا ۴ بار این پروسه طی شود. در آخر نرخ تشخیص سیستم که بلوک دیاگرام آن را در شکل ۵-۱۶ می بینید ۸۳٪ و انحراف معیار استاندارد آن ۱۱٪ است که نسبت به PCA ی تنها در مرحله قبل پیشرفت داشته و بازده در حدود نتایج به دست آمده از شبکه عصبی روی دیتا بیس آلمانی است.

در مرحله بعد بر آن می شویم که کلاس بندی را روی چشم ها و دهان های جدا شده از چهره اعمال کنیم و انتظار پاسخ دهی مطلوب تری داریم.

از آن جا که بینی کمک خاصی در کلاس بندی به ما نمی کند از آن صرف نظر کرده، چشم چپ و دهان هر فرد را در ماتریسی سطری قرار داده و مانند مرحله ی قبل از ماتریس چشم و دهان همان شش نفر استفاده کرده تا پایه های PCA شاد و غمگین را بسازیم. پس از ساختن پایه های مربوط به کلاس شاد و غمگین، باقی مراحل مانند قسمت قبل می باشد. نرخ تشخیص برای سیستم با ماتریس چشم و دهان ۸۷٪ به دست می آید که نسبت به بازده مرحله قبل حدود ۷٪ افزایش داشته است، هر چند ما انتظار رشد بیشتری در نرخ تشخیص داشتیم.

¹ Scale



شکل (۵-۱۶) بلوک دیاگرام روش تشخیص احساس توسط آنالیز مولفه های اساسی با دو گروه پایه-شبکه عصبی در محله بعد از چشم هم صرف نظر کرده و PCA را فقط روی دهان می زنیم و پایه های کلاس شاد و ناراحت را استخراج می کنیم و به کمک شبکه عصبی کلاسه بندی می کنیم. بازده این بار افزایش می یابد و به ۸۴٪ می رسد. یک بار دیگر سیستم را فقط توسط چشم های استخراج شده آزمایش می کنیم و بازده تشخیص را به دست می آوریم که برابر با ۷۹٪ می شود. نتایج به دست آمده تا حدودی بر خلاف پیش بینی ما است. در تفسیر نتایج می بینیم که چشم ها شبکه عصبی را گیج می کنند و میزان پاسخ دهی سیستم را می کاهند. همان طور که در شکل ۵-۱۷ مشخص است با مقایسه چشم افراد در حالت شاد و غمگین در می یابیم از آن جا که چشم های استخراج شده در حالت شاد و غمگین اختلاف چندان واضحی ندارند و فاقد ابرو هستند که در حالات مختلف تغییرات واضح تری پیدا می کند می توانیم کاهش راندمان سیستم را توجیه کنیم. برای زمانی که چشم و دهان با هم استفاده می شوند و در مقایسه با دهان تنها نرخ تشخیص مقدار کمی افزایش یافته است. تفسیر می کنیم که حضور چشم ها باعث افزایش حجم ماتریس ها شده و ما برای آنکه بتوانیم شبکه عصبی را آموزش دهیم باید حجم ماتریس را کاهش می دهیم و از این رو حضور چشم ها باعث شده است که مقداری از اطلاعات مفید را در فرآیند PCA زدن از دست بدهیم. هم چنین دلیل دیگر می تواند گنگی اطلاعات ترکیبی برای شبکه عصبی باشد که بازدهی آن را چندان رشد نمی دهد.



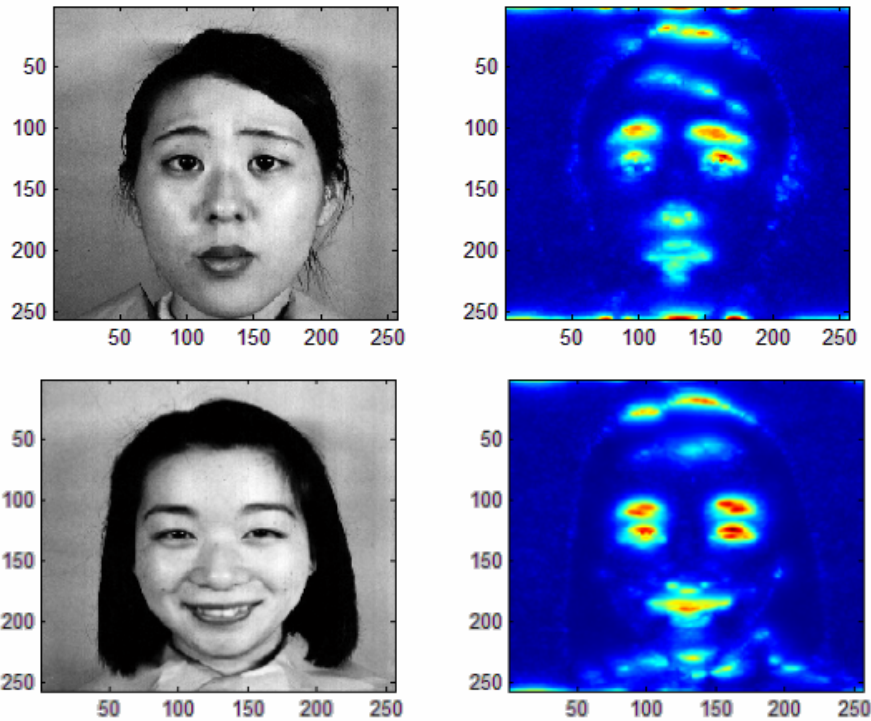
شکل (۵-۱۷): مقایسه چشم افراد در حالت شاد و غمگین

بهترین بازدهی که تا به حال رسیده ایم ۸۷٪ است که در صدد افزایش آن بر می یابیم. برای کمک به کلاس بندی به دنبال ویژگی هایی می گردیم که متمایز کننده دو حالت شاد و ناراحت باشد. می توانیم

از پارامترهایی مانند پهنای چشم یا پهنای دهان استفاده کنیم و یا از فاصله دو پلک و یا از فاصله ی لب بالا و لب پایین، که به طور گسترده ای در کارهای قبل استفاده شده است. روش دیگری که بسیار موثر گزارش شده است استفاده از ویژگی های استخراج شده توسط تبدیل گابر دو بعدی است که در بخش ۵-۳-۳ از چگونگی پیاده سازی آن و نتایج شبیه سازی آن صحبت می کنیم.

۵-۳-۳- روش تشخیص احساس آنالیز مولفه های اساسی - شبکه عصبی با ورودی اصلاح شده توسط گابر

در سال های اخیر استفاده از فیلترهای گابر در پردازش تصویر و آنالیزهای چهره به شدت مورد توجه بوده است که نشانگر قابلیت های زیاد این فیلترها در زمینه پردازش تصویر است. مطابق مختصری که در فصل سوم درباره ی گابر بیان شد این فیلترها با چرخش 'های متفاوت و در گام' های مختلف قابل پیاده سازی اند و هر چرخش و هر گام اطلاعات خاصی از تصویر را استخراج می کند. نمونه ای از خروجی های گابر را در دو چرخش مختلف در شکل ۵-۱۸ می بینید که نشانگر درجه ی بالای همبستگی گابر با اجزای اصلی چهره است.

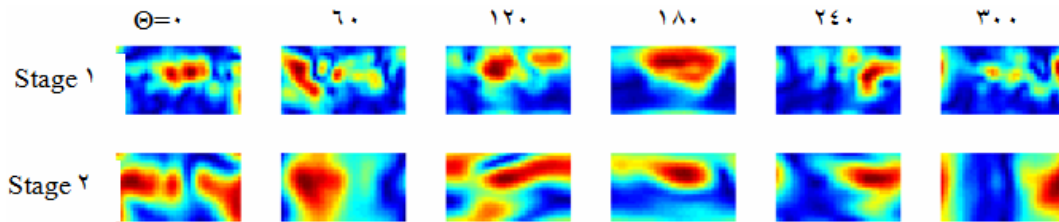


شکل (۵-۱۸) دو نمونه از تصاویر چهره و خروجی فیلتر گابر آن در دو حالت شاد و غمگین به ترتیب در چرخش ۳۶۰ و ۱۸۰ درجه و گام اول

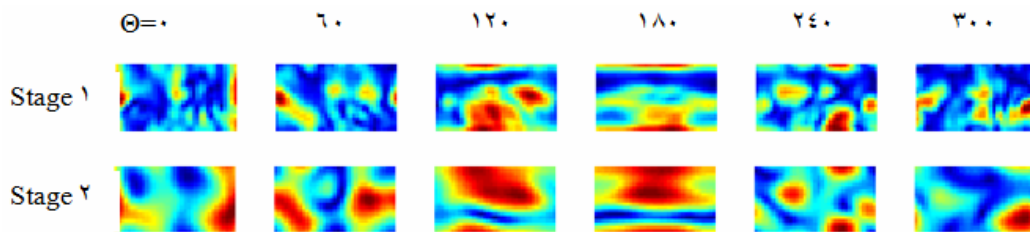
در این جا نیز مانند بخش های قبل با اعمال تبدیل موجک با تابع مادر هار نواحی اصلی چهره مانند چشم و دهان و بینی را استخراج می کنیم. پس از آن این بار روی نواحی استخراج شده فیلترهای گابر

¹ Orientation
² Stage

اعمال می کنیم. دهان ماتریسی ۲۰ در ۳۵ و چشم ماتریسی ۲۱ در ۲۱ است. ما فیلترهای گابر با چرخش ها و گام های متفاوت را روی اجزا اصلی چهره اعمال و نتایج را بررسی کردیم تا بتوانیم کرنل های مناسب گابر را که در احساسات مختلف بیشترین تمایز را با هم دارند و هنگام کلاس بندی به راحتی قابل تفکیک شدن هستند مشخص کنیم. پس از بررسی های زیاد ۱۲ فیلتر گابر با ۶ چرخش^۱ با زاویه های صفر، ۶۰، ۱۲۰، ۱۸۰، ۲۴۰، ۳۰۰ و در دو گام^۲ انتخاب شدند که اطلاعاتی با بیشترین حجم تمایز در کلاس های احساسی شاد و غمگین داشتند. در شکل ۱۹-۵ و ۲۰-۵ و ۲۱-۵ و ۲۲-۵ نمونه خروجی این فیلترها را برای لب و چشم چپ ملاحظه می کنید.

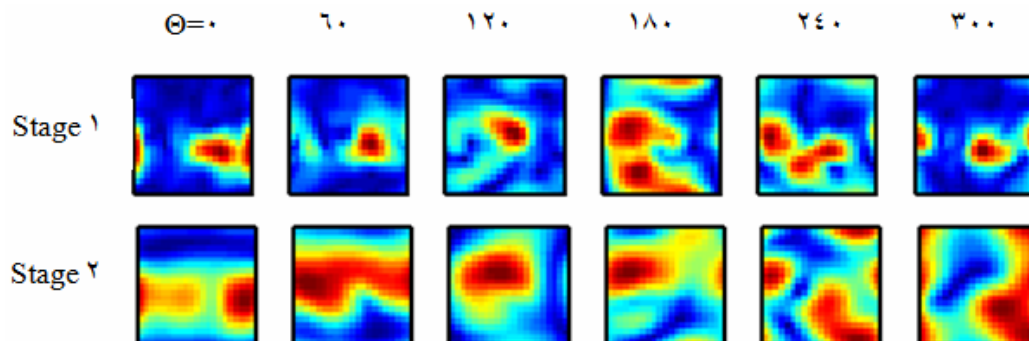


شکل (۱۹-۵) خروجی ۱۲ فیلتر گابر برای لب شاد



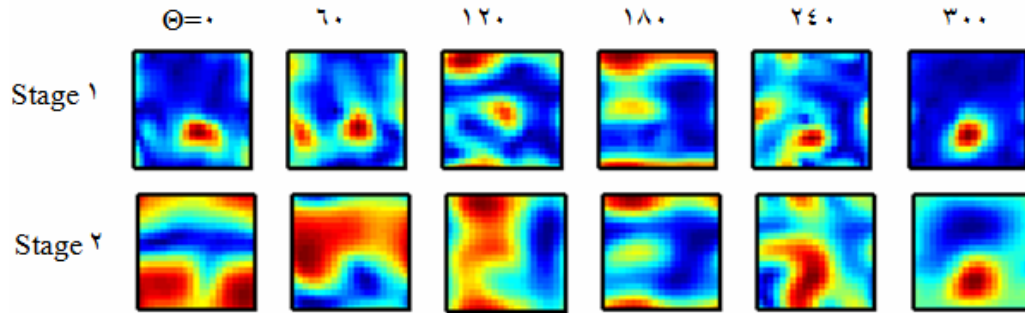
شکل (۲۰-۵) خروجی ۱۲ فیلتر گابر برای لب غمگین

همان طور که در شکل های ۱۹-۵ و ۲۰-۵ می بینید دو کرنلی که بیش از همه برای ما حاوی اطلاعات متمایز کننده احساسات است کرنل های گام دوم با زاویه های ۶۰ و ۲۴۰ است. برای حالت شاد در چرخش ۶۰ نیمه چپ و در چرخش ۲۴۰ نیمه راست خروجی فیلتر حاوی تراکم رنگ آبی است در حالیکه برای حالت غمگین این امر کاملاً برعکس بوده، در چرخش ۶۰ نیمه چپ و در چرخش ۲۴۰ نیمه راست خروجی فیلتر حاوی تراکم رنگ قرمز است.



شکل (۲۱-۵) خروجی ۱۲ فیلتر گابر برای چشم چپ شاد

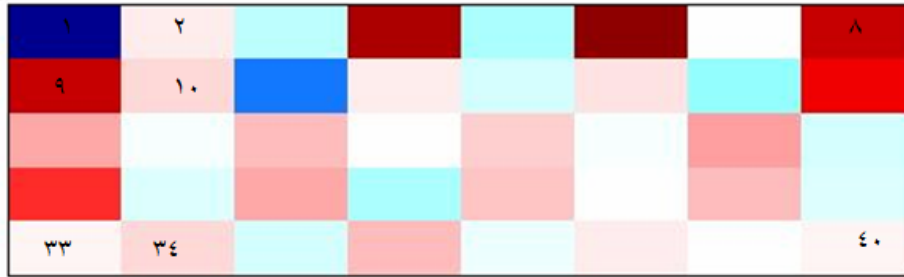
¹ Orientation
² Stage



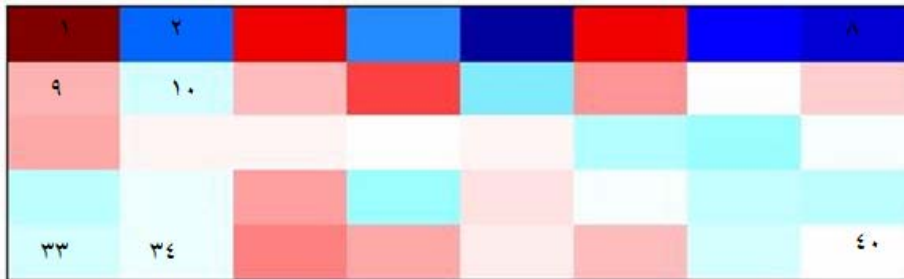
شکل (۲۲-۵) خروجی ۱۲ فیلتر گایر برای چشم چپ غمگین

حاصل اعمال فیلترها اعداد مختلطی شامل دامنه و فاز است که با بررسی میزان مفید بودن اطلاعات دامنه و فاز به طور جداگانه به این نتیجه می‌رسیم که دامنه برای هدف مورد نظر ما مفیدتر است و از نتایج حاصل فاز را حذف کرده و ماتریس دامنه را به دست می‌آوریم و سطری می‌کنیم. پایه‌های PCA را از روی نصف تصاویر شاد و نصف تصاویر غمگین می‌سازیم. البته باید اضافه کرد افزایش یا کاهش تعداد نمونه‌ها برای ساخت پایه‌های PCA تاثیر زیادی در فضای بردارهای ویژه^۱ ندارد. چون فضای بردارهای ویژه برای لب و چشم برش داده شده تقریباً یکسان است و افزایش یا کاهش تعداد نمونه‌ها اختلاف زیادی ایجاد نمی‌کند. تنها نکته‌ی مهم در این است که باید از احساسات مختلف نمونه یا نمونه‌هایی برای ساختن برداریم تا قادر به کلاسه‌بندی احساسات گوناگون باشیم و حضور نمونه از هر احساس مورد بررسی الزامی است. ما از ۸۴۰۰ پایه حاصل از PCA تنها ۴۰ پایه اول را نگه می‌داریم. با مقایسه خروجی PCA برای یک لب شاد (نگاشت روی ۴۰ پایه) و یک لب غمگین متوجه می‌شویم که خواست ما برای کسب اختلافات واضح و تفکیک پذیر بین ویژگی‌های مورد بررسی به خوبی مرتفع شده است. به اختلاف رنگهای حاصل از نگاشت یک لب روی ۴۰ پایه در شکل‌های ۵-۲۳ و ۵-۲۴ توجه کنید. نوع دیگری از نمایش این اختلاف بین دو حالت لب شاد و غمگین را در شکل ۵-۲۵ با نمودار لگاریتمی محدود شده بین ۱- و ۱می بینید. برای مقایسه بیشتر فیلترهای گایر را یک بار روی میانگین تمام لب‌های شاد استخراج شده و بار دیگر روی میانگین تمام لب‌های غمگین استخراج شده اعمال کردیم و نگاشت آن‌ها را روی ۴۰ پایه PCA به دست آوریم. در شکل ۵-۲۶ نمودار لگاریتمی محدود شده بین ۱- و ۱ این میانگین‌ها را می‌بینید. همان‌طور که می‌بینیم میزان اختلاف خروجی PCA برای میانگین‌ها نسبت به لب شاد و غمگین متعلق به یک نفر کاهش یافته است که قابل پیش‌بینی بود. چرا که میانگین‌گیری از تصاویر متعلق به افراد مختلف باعث تلطیف شدن احساس نشان داده شده می‌گردد و اختلاف را با احساسی دیگر کاهش می‌دهد.

¹ Eigen Space

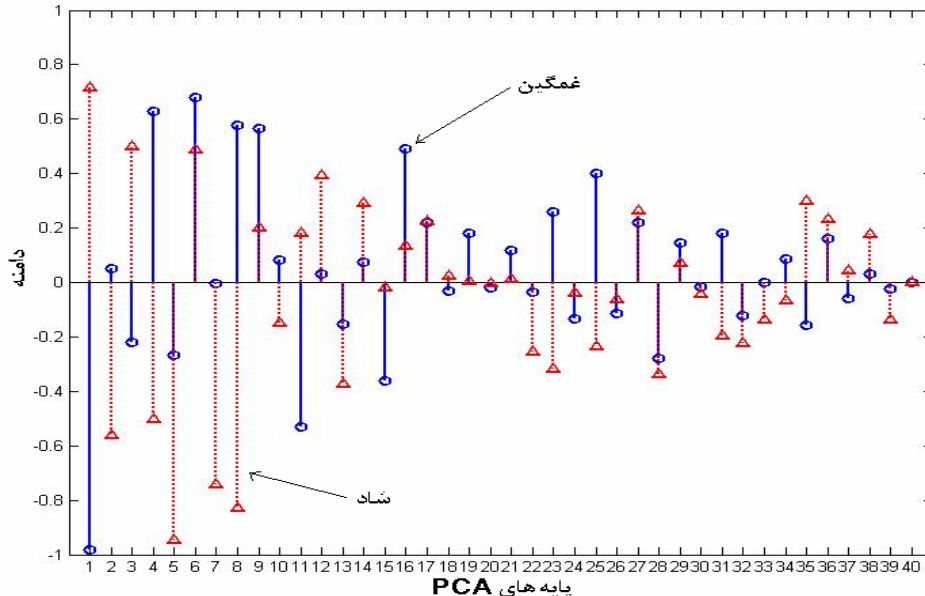


شکل (۲۳-۵) خروجی PCA برای یک لب شاد (نگاشت روی ۴۰ پایه) برای مقایسه (اعداد داخل تصویر شماره پایه اند)

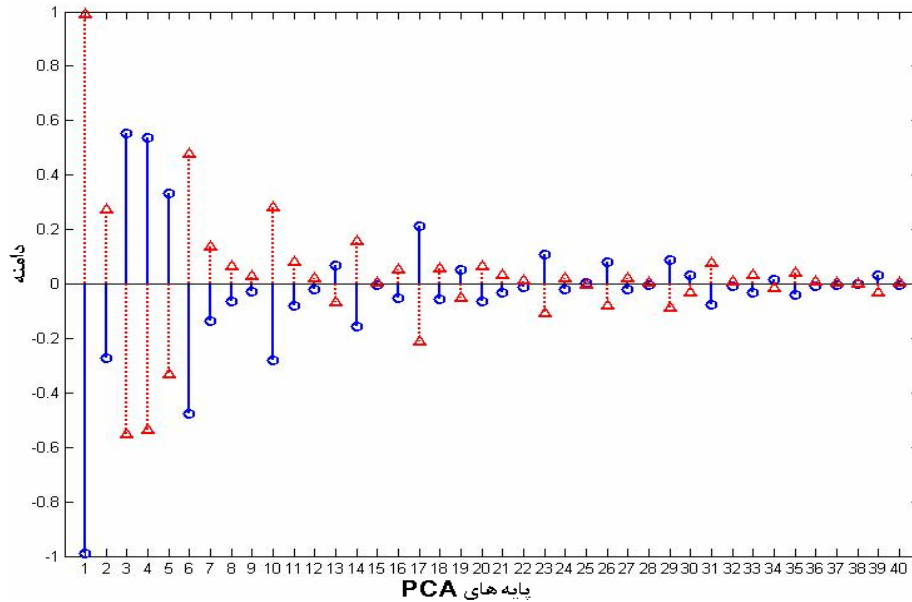


شکل (۲۴-۵) خروجی PCA برای یک لب غمگین (نگاشت روی ۴۰ پایه) برای مقایسه (اعداد داخل تصویر شماره پایه اند)

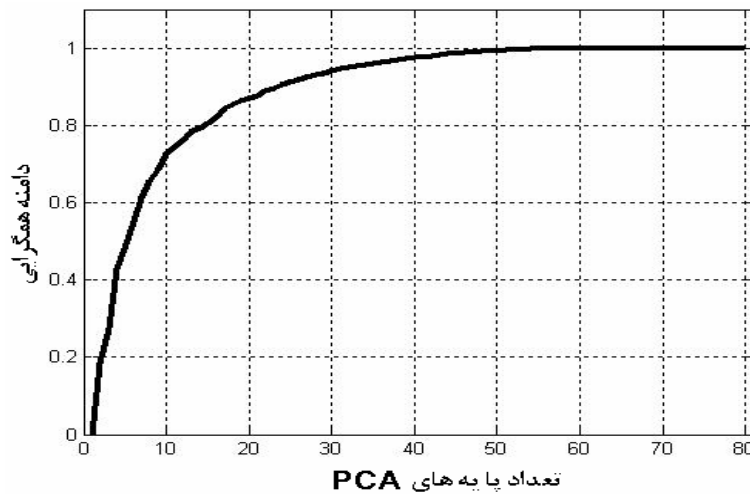
همان طور که با مقایسه ی شکل های ۲۳-۵ و ۲۴-۵ می بینید پایه های نخست اختلاف رنگ واضح تر و شدیدتری با هم داشته که روی پایه های انتهایی این مساله کم رنگ می شود. با مقایسه ی شکل های ۲۵-۵ و ۲۶-۵ نیز همین مسئله را می بینیم. به طوری که پایه های نخست دو حالت شاد و غمگین بیشتر از هم فاصله داشته و در کلاس بندی مفیدترند و روی پایه های آخر میزان اختلاف کم می شود. این موضوع نشان می دهد که ما می توانیم تعداد پایه های PCA را از ۴۰ عدد هم کاهش دهیم.



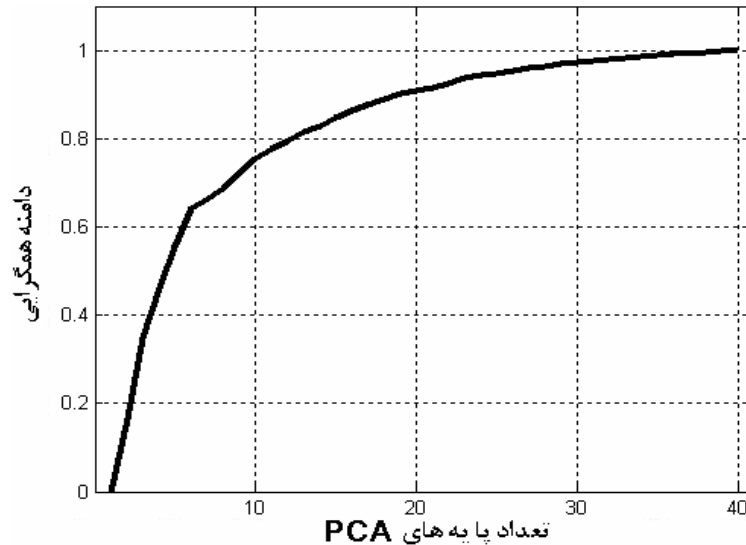
شکل (۲۵-۵) خروجی PCA برای یک لب شاد و یک لب غمگین (نگاشت روی ۴۰ پایه) برای مقایسه



شکل (۵-۲۶) خروجی PCA برای میانگین لب های شاد و لب های غمگین (نگاشت روی ۴۰ پایه) برای مقایسه صحت این مسئله را می توانیم با رسم خروجی PCA با ۴۰ پایه برای لب مورد بررسی قرار دهیم که مطابق شکل (۵-۲۷) می بینیم که PCA روی ۳۰ پایه همگرا شده و تعداد پایه بیشتر الزاما کمک بیشتری نمی کند و میزان اطلاعات بسیار محدودی دارند. نمایش دیگری از همگرا شدن PCA را در شکل (۵-۲۸) برای چشم چپ ملاحظه می کنید.

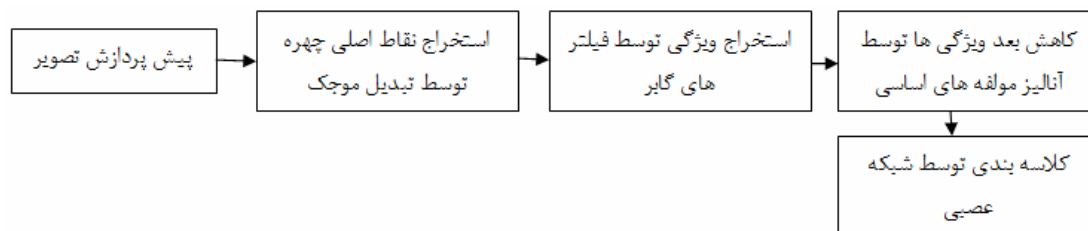


شکل (۵-۲۷) نمایش همگرا شدن PCA با ۳۰ پایه برای لب



شکل (۵-۲۸) نمایش همگرا شدن PCA با ۳۰ پایه برای چشم

دامنه خروجی های گابر به دست آمده از مرحله قبل را روی ۴۰ پایه ی PCA نگاشت کرده و یک ماتریس ۴۰ تایی به دست می آوریم که برای کلاسه بندی وارد شبکه عصبی می شود. شبکه عصبی به کار رفته در این مرحله پرسپترون سه لایه با ۴۰ گره در لایه ورودی، ۵۰ گره در لایه میانی و یک گره در خروجی، پیش خور و پس انتشار است. با استفاده از این سیستم ما کلاسه بندی را در سه بخش انجام می دهیم. در بخش اول سیستم را روی لب برش داده شده پیاده می کنیم و بازده تشخیص ۹۵٪ به دست می آید. در بخش دوم سیستم را فقط روی چشم چپ اعمال می کنیم که نرخ تشخیص ۸۲٪ تحویل می دهد. در بخش سوم سیستم را روی ماتریس چشم و دهان اجرا می کنیم که بازده تشخیص ۹۹٪ می شود که بسیار مطلوب است. در این جا نیز مانند سیستم های قبل چشم ها نسبت به دهان بازده کمتری دارند و در تشخیص احساس شادی و غم نسبت به دهان کمتر تغییر وضعیت داده و به تنهایی برای کلاسه بندی احساسات شادی و غم مناسب نیستند. در عوض در این بخش می بینیم که تلفیق ماتریس های لب و چشم چپ بازده سیستم را به خوبی افزایش داده در حد ۹۹٪ بالا برده است که این رشد قابل ملاحظه در نرخ تشخیص به خاطر امتیازات خاصی است که استفاده از تبدیل گابر به سیستم تشخیص ما ارزانی داشته است. بلوک دیاگرام سیستم مورد نظر را در شکل ۵-۲۹ می بینید.



شکل (۵-۲۹) بلوک دیاگرام سیستم تشخیص احساس با الگوریتم فیلترهای گابر و آنالیز مولفه های اساسی و شبکه عصبی روی اجزای چهره

فیلترهای گابر ساده و سریع و یکی از آسان ترین روش های استخراج ویژگی هستند و ویژگی های استخراج شده از طریق این فیلترها مستقل از اطلاعات فردی و اطلاعات تصویر است. برای مثال در زمینه

ی کار ما ویژگی های استخراج شده از قیافه ی شخص مستقل می شود و وابستگی شدیدی به احساس پیدا می کند. این نکته ی مهمی است چرا که پیش از این وابستگی ویژگی ها به قیافه ی شخص بیش از وابستگی آن ها به احساس شخص بود که مشکل عمده ای محسوب می شد. برای مثال در این جا به کمک فیلتر های گابریل جهت ها را از روی لب با اندازه و فاز استخراج می کنیم و نه جزئیات مربوط به لب که به شخص خاص وابسته شود. استفاده از گابریل سریع و مقرون به صرفه است چرا که برای استخراج ویژگی ها نیاز به محاسبات پیچیده و زمان بر نداریم و سیستم نیاز به هوشمندی خاصی ندارد. مثلا الگوی خاصی را نباید جست و جو و یا ایجاد کرد، یا لازم نیست سیستم با داده های خاصی آموزش ببیند.

سیستم توضیح داده شده را به روش کراس ولیدیشن مدل k-fold آزمایش می کنیم به این ترتیب که از کلاس شاد ۲۷ تصویر و از کلاس غمگین نیز ۲۷ تصویر برای آموزش شبکه عصبی استفاده می شود و با ۳ تصویر شاد و ۳ تصویر غمگین از شخصی که تا به حال داده ای در سیستم نداشته است آزمایش می شود. و این کار ۹۰ بار تکرار می شود و بازده نهایی تشخیص ما ۹۹٪ خواهد بود که مناسب و قابل قبول است. مقایسه نتایج بدست آمده از الگوریتمهای توضیح داده شده در ۱-۳-۵ تا ۳-۳-۵ در جدول ۱-۵ آمده است. با مقایسه ی نتایج حاصل با کارهای انجام شده در این زمینه با سایر روش ها، ملاحظه می کنیم که بازده نهایی این کار به نحو مطلوبی با نتایج دیگر کارها قابل رقابت بوده و تقریبا از ۹۰٪ کارهای قبل، بازده بالاتری دارد.

جدول (۱-۵): مقایسه نتایج بدست آمده از روشهای ۱-۳-۵ تا ۳-۳-۵

نرخ تشخیص (%)	نام پایگاه داده	محل اعمال	نام روش
۸۶	FG-NET	کل چهره	PCA با یک پایه
۸۶	FG-NET	چشم	PCA با یک پایه
۸۹	FG-NET	لب	PCA با یک پایه
۸۹	FG-NET	چشم و لب	PCA با یک پایه
۸۴	FG-NET	کل چهره	PCA با یک پایه و MLP
۸۷	FG-NET	چشم	PCA با یک پایه و MLP
۸۴	FG-NET	لب	PCA با یک پایه و MLP
۸۶	FG-NET	چشم و لب	PCA با یک پایه و MLP
۷۸	JAFEE	کل چهره	PCA با دو پایه
۸۳	JAFEE	کل چهره	PCA با دو پایه و MLP
۷۹	JAFEE	چشم	PCA با دو پایه و MLP
۸۴	JAFEE	لب	PCA با دو پایه و MLP
۸۷	JAFEE	چشم و لب	PCA با دو پایه و MLP
۸۲	JAFEE	چشم	PCA با دو پایه و MLP و گابریل
۹۵	JAFEE	لب	PCA با دو پایه و MLP و گابریل
۹۹	JAFEE	چشم و لب	PCA با دو پایه و MLP و گابریل

۵-۳-۴- روش تشخیص احساسات آماری- شبکه عصبی

در این قسمت برای تشخیص سه حالت شاد، غمگین و عادی از الگوریتمی مبتنی بر اطلاعات آماری در تلفیق با شبکه عصبی سود می‌جوئیم. برای کلاس بندی توسط یک شبکه هوشمند و به کمک اطلاعات آماری جواب این سوالات می‌تواند بسیار تعیین کننده باشد.

الف- آیا یک نوزاد که تا به حال هیچ شخصی را ندیده است تنها با دیدن تصویر یک شخص که از دید عموم شاد است می‌تواند شادی این شخص را تشخیص دهد؟

جواب این سوال مسلماً منفی است چون نوزاد حالات دیگر را ندیده است و نمی‌شناسد پس عملاً نمی‌تواند حالت فرد را تشخیص دهد. این نوزاد تنها با استفاده از تقلید و وانمود کردن می‌تواند قدرت تشخیص خود را بروز دهد. لذا می‌توان نتیجه‌گیری کرد که برای آموزش یک شبکه هوشمند باید از یک مجموعه کامل از تمامی حالات استفاده کرد. بدیهی است که جامع نبودن مجموعه آموزش امکان تشخیص حالات جدید را در گروه‌های مشخص فراهم نمی‌کند.

ب- آیا یک شخص بالغ با دیدن یک تصویر جدید از شخصی که تا به حال ندیده است می‌تواند حالت او را تشخیص دهد؟

جواب این سوال به قابلیت تشخیص و قدرت ذهن فرد بستگی دارد. در واقع اگر شخص فردی را تا به حال ندیده باشد تنها در صورتی می‌تواند حالت چهره او را تشخیص دهد که یک الگوریتم سازمان یافته در ذهن این فرد برای تشخیص حالت یک شخص جدید وجود داشته باشد. می‌توان برای این الگوریتم‌های سازمان یافته این حالت را متصور شد که شخص تمامی حالت‌های یک فرد مشخص و یا گروهی مشخص را دیده و بعد در مقام مقایسه و یا استفاده از یک الگوریتم هوشمند می‌تواند حالات اشخاص جدید را تشخیص دهد. البته این نکته حائز اهمیت است که تمامی افراد در کنار این الگوریتم‌های هوشمند از داده‌های آماری موجود در ذهن خود برای فرار از فکر کردن استفاده می‌کنند.

ج- آیا با دیدن یک جز از چهره یک فرد می‌توان حالت چهره او را تشخیص داد؟

جواب این سوال مسلماً منفی است. حالات کلی ثبت شده در ذهن عموم افراد جامعه این مساله را تایید نمی‌کند. برای مثال ممکن است لب فردی خندان و چشم او گریان باشد. با این وجود وقتی اشخاص به همدیگر نگاه می‌کنند تمرکز آنها بر روی نواحی برجسته صورت است و عملاً هم افراد در حالت واقعی از این نواحی برجسته برای تشخیص حالات چهره استفاده می‌کنند. همچنین بدیهی است کاهش حجم اطلاعات موجود در ذهن به تسهیل در امر تصمیم‌گیری کمک می‌کند لذا استفاده همزمان از این نواحی برجسته به تنهایی کارساز به نظر می‌رسد.

د- چگونه می‌توان یک الگوریتم سازمان یافته برای شناسایی سه حالت شادی، غم و حال عادی با توجه به نکات ارزشمند سوالات فوق به وجود آورد؟

در مقایسه با مغز انسان در شناسایی الگوهای مشخص می‌توان گفت شبکه‌های عصبی یکی از الگوریتم‌های پیشرفته در شناسایی الگوهای عملی هستند. البته قدرت تشخیص این شبکه‌ها به انتخاب مناسب توابع تحریک شبکه، تعداد مناسب سلول‌های عصبی در هر لایه، تعداد لایه‌های پنهان و هم چنین تعداد

ورودی های شبکه وابسته است. در کنار انتخاب یک شبکه هوشمند مناسب، شبکه ای با استفاده از داده های آماری مسلما در تسهیل امر تصمیم گیری بسیار کارساز است. به علاوه به منظور کاهش حجم اطلاعات ورودی و به منظور کاهش حجم محاسبات و آموزش مناسب شبکه استخراج مناسب نقاط برجسته چهره ضروری به نظر می رسد. در این پایان نامه بعد از آزمون های فراوان مشاهده شد که استفاده از یک شبکه عصبی به تنهایی حتی در حالتی که ویژگی های صورت به شکل بسیار بهینه انتخاب شوند- بسیار قابل تمایز و با حجم کم- در شناسایی حالات فرد عملا نتایج خوبی ندارند.

با توجه به تمامی نکات ارائه شده می توان چنین متصور شد که برای پیاده سازی یک الگوریتم کارساز برای تشخیص حالات چهره در ابتدا کاهش تعداد ورودی های چهره امری الزامی است. در این پایان نامه از روش نقاط برجسته موجه به منظور حذف ورودی های ناکارآمد و استخراج چشم ها و لب استفاده می شود. به دنبال آن، این ورودی ها تبدیل به یک بردار می شوند و به دنبال آن یک روند آماری به منظور بررسی شباهت ها و تفاوت های نقاط برجسته صورت افراد مختلف بر روی تعداد زیادی از داده ها انجام می شود. به منظور تشخیص میزان شباهت ها و تفاوت های اجزای صورت افراد مختلف استفاده از توابع وابستگی و انحراف معیار مفید است. انحراف معیار کمتر به معنی شباهت بیشتر و ضریب همبستگی بیشتر به معنی شباهت بیشتر است. این توابع به صورت (4-5) و (5-5) قابل تعریف هستند که رابطه ی 5-5 در رابطه با انحراف معیار است.

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad s = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2} \quad (5-5)$$

بعد از مقایسه چشم چپ و لب در سی تصویر از لحاظ آماری با استفاده از توابع بالا نتایج زیر حاصل شده است که نتایج جمعی از آنها در جدول 5-2 تا 5-6 آمده است.

الف: هنگام مقایسه دو تصویر شاد-شاد مشاهده می شود که لب های آنها شباهت کمتری نسبت به هم در مقایسه با چشم های آنها دارند.

ب- هنگام مقایسه دو تصویر غمگین-غمگین مشاهده می شود که چشم های آنها شباهت کمتری نسبت به هم در مقایسه با لب های آنها دارند.

ج- هنگام مقایسه دو تصویر غمگین- شاد مشاهده می شود که لب های آنها شباهت بسیار کمتری نسبت به هم در مقایسه با چشم های آنها دارند.

د- هنگام مقایسه دو تصویر غمگین-غمگین مشاهده می شود که لبهای آنها شباهت بسیاری به هم دارند.

ه- هنگام مقایسه دو تصویر غمگین- شاد مشاهده می شود که لب های آنها شباهت بسیار کمی دارند.

و- هنگام مقایسه دو فرد شاد-عادی مشاهده می شود که نه لب ها و نه چشم ها شباهتی ندارند و اندازه تشابه نسبتا برابری دارند.

ز- هنگام مقایسه دو فرد غمگین-عادی مشاهده می شود که نه لب ها و نه چشم ها شباهتی به هم ندارند و اندازه تشابه نسبتا برابری دارند.

با محاسبه ی انحراف معیار و ضریب همبستگی می بینیم نتایج بالا در ۸۹٪ از کل تصاویر مشاهده می شود.

جدول (۲-۵) مقایسه چشم ها و لب ها در دو فرد غمگین-غمگین

شماره تصاویر	ضریب همبستگی لب ها	ضریب همبستگی چشمها	انحراف معیار لبها	انحراف معیار چشم ها
۴-۱	۰,۶۶	۰,۳۶	۲۳	۳۲
۲۹-۱	۰,۴۱	۰,۲۲	۲۶	۳۶
۲۹-۴	۰,۷۷	۰,۵	۱۸	۳۲
۲۹-۲۴	۰,۰۳	۰,۴	۳۱	۳۴

جدول (۳-۵) مقایسه چشم ها و لب ها در دو فرد شاد-شاد

شماره تصاویر	ضریب همبستگی لب ها	ضریب همبستگی چشمها	انحراف معیار لبها	انحراف معیار چشم ها
۳۷-۳۳	۰,۱۱	۰,۴۴	۳۳	۳۹
۳۲-۵	۰,۵۲	۰,۶۷	۳۱	۳۱
۳۸-۵	۰,۰۴	۰,۷۲	۴۲	۲۶
۳۸-۳۷	۰,۰۴	۰,۶۴	۳۷	۲۸

جدول (۴-۵) مقایسه چشمها و لبهای دو فرد شاد-غمگین

شماره تصاویر	ضریب همبستگی لب ها	ضریب همبستگی چشمها	انحراف معیار لبها	انحراف معیار چشم ها
۳۳-۱	-۰,۰۳	۰,۴	۳۷	۳۵
۵-۱	۰,۱۶	۰,۵۳	۳۵	۳۲
۳۸-۴	-۰,۴۴	۰,۵۶	۵۰	۲۸
۳۷-۲۴	۰,۲۹	۰,۷۳	۲۷	۲۴

جدول (۵-۵) مقایسه چشمها و لبها در دو فرد شاد-عادی

شماره تصاویر	ضریب همبستگی لب ها	ضریب همبستگی چشمها	انحراف معیار لبها	انحراف معیار چشم ها
۳۳-۳	-۰,۰۴	-۰,۰۱	۵۵	۵۰
۵-۹	۰,۱۷	۰,۱۶	۴۹	۴۶
۳۸-۱۳	-۰,۴۸	-۰,۴۸	۵۶	۵۸
۳۷-۱۵	۰,۱۹	۰,۱۹	۴۳	۲۴

جدول (۶-۵) مقایسه چشمها و لبها در دو فرد غمگین-عادی

شماره تصاویر	ضریب همبستگی لب ها	ضریب همبستگی چشمها	انحراف معیار لبها	انحراف معیار چشم ها
۴-۳	-۰,۰۷	-۰,۰۴	۵۰	۵۱
۲۹-۹	۰,۱۱	۰,۱۳	۴۸	۴۶
۲۹-۱۱	-۰,۳۸	-۰,۴۰	۵۶	۵۰
۲۹-۱۵	۰,۰۹	۰,۱۰	۴۱	۴۰

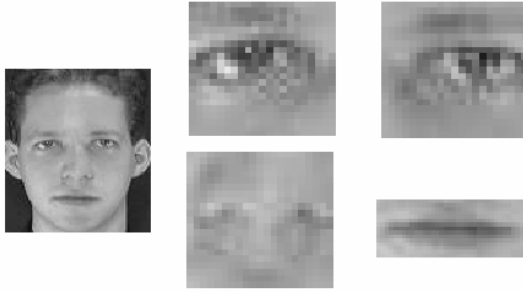
با توجه به نتایج به دست آمده ی بالا نکات زیر را می توان فهمید که لب تنها جزئی است که به عنوان ویژگی قابل تمایز در شبکه عصبی می تواند استفاده شود. در کنار لب و با استفاده از تعمیم نتایج آماری به دست آمده، چشم ها در یک الگوریتم آماری-مقایسه ای می توانند سرعت و دقت الگوریتم ترکیبی

شبکه عصبی-آماری را افزایش دهد. این الگوریتم دو قسمت کلی دارد، یعنی قسمت آماری و قسمت هوشمند که در آن قسمت آماری ضریب (۰,۵×۰,۸۹) و قسمت هوشمند ضریب (۰,۵×۰,۸۹) را دارد و نهایتاً با جمع ضرایب، ضریب تشخیص نهایی به دست می آید. برنامه کامپیوتری اجرایی این الگوریتم به این صورت است که در ابتدا یک تصویر شاد و یک تصویر غمگین انتخاب شده، سپس لب و چشم چپ مطابق شکل ۵-۳۰ و ۵-۳۱ استخراج می شود. پس از استخراج چشم و دهان آن ها را مطابق شکل ۵-۳۲ و ۵-۳۳ به یک بردار تبدیل کرده و به عنوان مبنای مقایسه انتخاب می شوند. به دنبال آن یک شبکه عصبی برای تشخیص حالت شاد و غمگین با چندین نمونه مشخص از لب های شاد و غمگین آموزش می بیند. بعد از بررسی های انجام شده مشاهده می شود که استفاده از یک شبکه عصبی برگشتی بازخورد از جلو با یک لایه پنهان^۱ و یا یک شبکه چند طبقه^۲ برگشتی بازخورد از جلو با دو لایه پنهان مفید است. توابع انتقال، تحریک، آموزش (بایاس/ وزن) و عملکرد به صورت پیش فرض در نرم افزار مطلب^۳ به ترتیب `tansig`، `traingd`، `learngdm` و `mse` انتخاب می شوند. ورودی این شبکه ماتریس دوبعدی لب ۱۵×۳۵ است که تبدیل به یک بردار ستونی می شود. هر کدام از شبکه های عصبی برای ۲۵ تصویر شاد و ۲۵ تصویر غمگین آموزش می بینند. خروجی های مبنای شبکه های عصبی بردارهای قرینه ی پله ای برای دو شبکه شاد و غمگین هستند. بعد از آموزش شبکه اینک به آزمایش الگوریتم برای یک ورودی ناشناس می پردازیم. در این آزمون ابتدا لب و چشم چپ استخراج شده و تبدیل به بردار می شوند. بردار لب به عنوان ورودی دو شبکه عصبی، خروجی هایی را تولید می کند که اگر با خروجی های مبنای یکی از شبکه ها برابر باشد می گوییم شبکه حالت تصویر را شناسایی کرده و ضریب تشخیص (-1) در 0.5×0.89 در الگوریتم کلی لحاظ می شود. سپس الگوریتم مقایسه ای، انحراف معیار و یا ضریب همبستگی لب و چشم چپ جدید را با هر کدام از تصاویر مبنای از قبل تعیین شده محاسبه می کند که اگر هنگام مقایسه با تصویر غمگین شباهت دو لب بیشتر از دو چشم چپ بود ضریب (۰,۵×۰,۸۹) را به عنوان ضریب تشخیص این مرحله قرار می دهد و تصویر جدید را غمگین شناسایی می کند، در غیر این صورت تصویر یا حالت عادی است یا شاد که اگر هنگام مقایسه با تصویر شاد شباهت دو چشم چپ بیشتر از دو لب بود تصویر شاد شناسایی شده و ضریب تشخیص (۰,۵×۰,۸۹) به آن اختصاص می یابد و در غیر این صورت تصویر در قسمت الگوریتم آماری به عنوان عادی با ضریب تشخیص (۰,۵×۰,۸۹) می شود. در صورتی حالت شخص ۱۰۰٪ تشخیص داده می شود که ضریب تشخیص یک باشد و در غیر این صورت حالت شخص به صورت نسبی با مقداری بین صفر و یک تعیین می شود. با پیاده سازی این الگوریتم بر روی تمامی تصاویر پایگاه داده های FG_Net نتایج مفیدی به دست می آید که در جدول (۵-۷) خلاصه می شود.

^۱ feed forward back propagation

^۲ Cascade

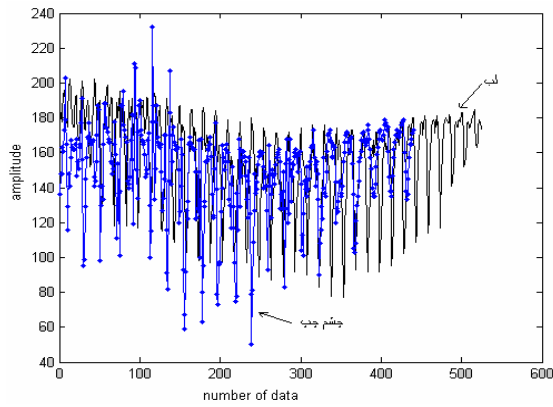
^۳ MATLAB



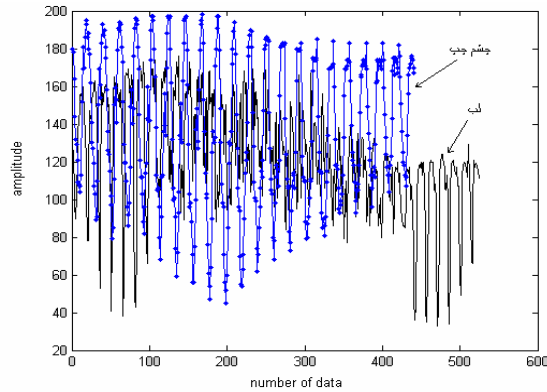
شکل (۵-۳۰): اجزای برجسته استخراج شده یک تصویر غمگین با استفاده از روش نقاط برجسته موجک



شکل (۵-۳۱): اجزای برجسته استخراج شده یک تصویر شاد با استفاده از روش نقاط برجسته موجک



شکل (۵-۳۲): تبدیل ماتریسهای لب و چشم به بردار برای یک فرد غمگین



شکل (۵-۳۳): تبدیل ماتریسهای لب و چشم به بردار برای یک فرد شاد

جدول (۵-۷): نتایج نهایی پیاده سازی این الگوریتم بر روی پایگاه تصاویر

تشخیص حالات چهره				
	صحیح با دقت بالاتر از ۹۰٪	نادرست	ضایع شده	نرخ تشخیص
غمگین	۱۳۰	۴۱	۰	۷۶,۲
شاد	۱۵۹	۱۲	۰	۹۲,۹۸
عادی	۱۵۸	۱۲	۱	۹۲,۳۹

فصل ششم

نتیجه گیری و پیشنهادات

۶-۱- نتیجه گیری

روش های تشخیص احساسات بسیار متنوع و به نسبت پیچیده اند. به علت ضرورت روزافزون به دست آوردن بازده بالا در این سیستم و زمینه های فراوان کاربرد این سیستم، از سی سال قبل تا کنون که دانشمندان کار در این زمینه را شروع کرده اند به طور رو به رشدی روش های تشخیص احساس چهره گسترش یافته، پیچیده شده اند و مدام از محدودیت های ورودی این سیستم کاسته شده است. طبق آخرین گزارشات بهترین نتیجه ها از اتخاذ روش مدل ظهور فعال روی تصاویر دینامیک و آنی برای استخراج ویژگی و کلاسه بندی شبکه عصبی با بازده ۹۸٪ به دست آمده است.

ما در این رساله در فصل اول به معرفی مختصر سیستم تشخیص احساس، بیان اهمیت و کاربرد و شرح خلاصه ی روش های موجود پرداختیم. در فصل دوم توضیح مختصر چندین روش پایه و پرکاربرد در این زمینه و نمونه هایی از هر سه روش ویژگی-پایه، مدل-پایه، و تصویر-پایه را ارائه کردیم. در فصل سوم به معرفی اجمالی تبدیل موجک و فیلترهای گابر پرداخته، در فصل چهارم شبکه های عصبی و تحلیل مولفه های اساسی (PCA) را خلاصه توضیح دادیم. در فصل پنجم به کمک تبدیل موجک و الگوی پیشنهادی از طرف سبه و لویپاس نواحی اصلی چشم و دهان شناسایی و استخراج کردیم. از طریق اعمال تبدیل گابر روی نواحی استخراج شده، ویژگی هایی حاصل شد که توسط تحلیل مولفه های اساسی کاهش حجم یافته، در مرحله ی آخر از طریق شبکه عصبی کلاسه بندی گردید. ما با استفاده از تحلیل مولفه های اساسی و شبکه های عصبی طور جدا و ترکیبی به کلاسه بندی پرداختیم. بطور کلی نتایج حاصله در این رساله به صورت زیر خلاصه می شود.

الف- برای استخراج نقاط اصلی چهره به کمک تبدیل موجک، ابهامات موجود در روش پیشنهادی از طرف سبه و لوپاس برطرف شد.

ب- با اعمال تحلیل مولفه های اساسی روی کل چهره و نواحی استخراج شده موفق به کلاسه بندی احساس شدیم. این روش با آنکه سریع و ساده و نسبت به چرخش سر مقاوم است درصد پاسخ دهی پایینی دارد.

ج- تلفیق تحلیل مولفه های اساسی و شبکه های عصبی و اعمال آن روی کل چهره و نواحی استخراج شده با آنکه الگوریتم را پیچیده کرده، از سرعت آن می کاهد، رشد بازده چندان چشمگیری ندارد.

د- اعمال فیلترهای گابر روی نواحی استخراج شده چشم و دهان برای استخراج ویژگی و سپس اعمال تحلیل مولفه های اساسی و شبکه های عصبی برای کلاسه بندی نسبت به روش های قبل محاسبات بیشتر و سرعت کمتری دارد اما بازده تشخیص آن از همه مطلوب تر است و به ۹۹٪ می رسد.

ه- در استفاده از روش آماری و شبکه عصبی با آنکه محاسبات و پیچیدگی آن کاهش می یابد و بازده خوبی دارد ولی آموزش شبکه و همگرایی آن دشوار است و وابسته به تصاویر افراد و کیفیت آن ها و زاویه ی چهره افراد می باشد.

۲-۶- پیشنهادات

الف- بعد از استخراج نقاط اصلی چهره یعنی همان چشم و دهان، می توانیم گوشه های لب و چشم، لب بالا و پایین و پلک بالا و پایین را به دست آورده از فاکتورهایی مانند پهنای چشم و پهنای دهان، فاصله ی لب بالا و پایین، فاصله ی پلک بالا و پایین، زاویه بین دو پلک، محیط دهان و ویژگی هایی از این قبیل در کنار ویژگی های حاصل از گابر استفاده کرد که به افزایش نرخ تشخیص کمک زیادی می کند.

ب- به کمک فیلترهای گابر به استخراج نواحی اصلی چشم و دهان از کل تصویر چهره بپردازیم و بتوانیم قید های اعمال شده برای روش کنونی را حذف و میزان دقت و تعمیم پذیری سیستم استخراج چشم و دهان را افزایش دهیم.

ج- از روش های کلاس بندی دیگری مانند الگوریتم ژنتیک یا SVM استفاده کنیم. یا شبکه عصبی استفاده شده برای کلاس بندی را تغییر دهیم و از شبکه عصبی RBF یا از انواع دیگر شبکه های MLP کمک بگیریم.

د- در الگوریتم ارائه شده برای استخراج نقاط اصلی چهره به کمک تبدیل موجک می توان به جای تابع مادر هار از توابع مادر دیگری سود جست و نتیجه را در استخراج ارزیابی کرد.

ه- در مراحل بعدی کار به کمک فیلترهای گابری و ابزارهای دیگر بتوانیم این روش تشخیص احساس را برای احساسات بیشتر و در تصاویر رنگی و حالت دینامیکی تعمیم دهیم.

- [1]. A. Mehrabian, "Communication without words" *Psychology Today*, vol 2, no. 4, pp. 53-56, 1968.
- [2]. P. Kakumanu, N. Bourbakis, "A Local-Global Graph Approach for Facial Expression Recognition" 18th IEEE International Conference on Tools with Artificial intelligence (ICTAI), IEEE, 2006.
- [3]. R. Cowie, E. Douglas-Cowie, J. G. Taylor, S. Ioannou, M. Wallace, S. kollias, "An Intelligent System for Facial Emotion Recognition" 0-7803-9332-5/05/\$20.00, IEEE, 2005.
- [4]. Ying-li Tian, Takeo Kanade, Jeffrey Cohn, "Recognizing Lower Face Action Units for Facial Expression Analysis" 0-7695-0580-5/00\$10.00, IEEE, 2000.
- [5]. Zhengyou Zhang, "Feature Based Facial Expression Recognition: Sensitivity Analysis and Experiments with A Multilayer Perceptron" *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 13, No. 6, pp. 893-911, 1999.
- [6]. Jeffrey F. Cohn, Karen Schmidt, Ralph Gross, Paul Ekman, "Individual Differences in Facial Expression: Stability over Time, Relation to Self Reported Emotion, and Ability to Inform Person Identification" ICMI, IEEE, 2002.
- [7]. Irene Kotsia, Ioan Buciu, Ioannis Pitas, "An Analysis of Facial Expression Recognition Under Partial Facial Image occlusion" Elsevier B. V., *Image and Vision Computing*, 2008.
- [8]. Shishir Bashyal, Ganesh K. Venayagamoorthy. "Recognition of facial expressions using Gabor wavelets and learning vector quantization" Elsevier, *Artificial Intelligence*, 12 November 2007.
- [9]. C. Shan, S. Gong, P. W. Mcowan. "Conditional Mutual Information Based Boosting for Facial Expression Recognition" *British Machine Vision Conference*, 2005.
- [10]. I. Cohen, N. Sebe, L. Chen, A. Garg, T. Huang. "Facial Expression Recognition from Video Sequence Temporal and Static Modeling" *Computer Vision and Image Understanding*, 2003.
- [11]. G. Zhao, M. Pietik. "Dynamic Texture Recognition using Volume Local Binary Patterns." *Proc. of European Conference on Computer Vision*, 2006.
- [12]. Peng Yang, Qingshan Liu, Dimitris N. Metaxas, "Boosting Coded Dynamic Features for Facial Action Units and Facial Expression Recognition" 1-4244-1180-7/07/\$25.00, IEEE, 2007.
- [13]. C. L. Huang, Y. M. Huang. "Facial Expression Recognition using Model-Based Feature Extraction and Action Parameters Classification", *Journal of Visual Communication and Image Representation*, vol. 8, pp. 278-290, 1997.
- [14]. G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, "Classifying Facial Actions", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 974-989, 1999.

- [15]. L. Trujillo, G. Olague, R. Hammoud, B. Hernandez, "Automatic feature Localization in Thermal Images for Facial Expression recognition", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 3, pp. 14, 2005.
- [16]. K. Aizawa, H. Harashima, "Model Based Analysis Synthesis Image Coding System For A Person's Face" Elsevier Science Publishers B. V., Signal Processing: Image Communication 1, pp. 139-152, 1989.
- [17]. M. Suwa, N. Sugie, K. Fujimora, "A Preliminary Note on Pattern Recognition of Human Emotional Expression," Proc. 4th Int. Joint Conf. Pattern Recognition, pp. 408-410, 1978.
- [18]. K. Mase, "Recognition of Facial Expression from Optical Flow," IEICE Trans. E&S, vol. 10, pp. 3473-3483, 1991.
- [19]. Y. Yacoob, L. Davis, "Recognition Facial Expression by Spatio-Temporal Analysis," Proc. Int. Conf. Pattern Recognition, Jerusalem, Israel, Vol. 1, pp. 747-749, October 1994.
- [20]. M. Bartlett, P. Viola, T. Sejnowski, L. Larsen, J. Hager, P. Ekman, "Classifying Facial Action," Advances in Neural Information Processing System 8, eds. D. Touretzky, MIT Press, Cambridge, MA, 1996.
- [21]. Irfan A. Essa, Alex P. Pentland, "Coding, Analysis, Interpretation and Recognition of Facial Expression" IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, July 1997.
- [22]. Y. Yacoob, L. Davis, "Recognition Human Facial Expression from Long Image Sequence using Optical Flow," TPAMI, 18(6), 1996.
- [23]. C. Lisetti, d. Rumelhart, "Facial Expression Recognition Using a Neural Network," Proc. FLAIRS, 1998.
- [24]. C. Padgett, G. Cottrell, "Representing Face Images for Emotion Classification," in M. Jordan, T. Petsche(eds.), Advances in NIPS, 9, MIT Press, 1997.
- [25]. M. Resenblum, Y. Yacoob, L. Davis, "Human Expression Recognition from Motion using a Radial Basis Function Network Architecture," IEEE Trans. Neural Networks, 7(5), 1996.
- [26]. W. Fellenz, J. Taylor et al. "Comparing Template Based Feature-Based and Supervised Classification of Facial Expression from Static Image," Proc. CSCC99. 1999.
- [27]. N. Sebe, M. S. Lew, I. Cohn, Y. Sun, T. Gevers, T. S. Huang, "Authentic Facial Expression Analysis," AFGR, 2004.
- [28]. T. Kanade, J. F. Cohn, Y. Tian, "Comprehensive Database for Facial Expression Analysis," AFGR, 2000.
- [29]. M. Bartlett, G. Littlewort, M. Frank, "Recognition Facial Expression: Machine Learning and Application to spontaneous Behavior," CVPR, Vol. 2, 2005.
- [30]. T. Cootes, G. Edwards, C. Taylor, "Active Appearance Models," TPAMI, 23(6), 2001.
- [31]. B. Abboud, F. Davoine, M. Dang, "Facial Expression Recognition and Synthesis based on an Appearance Model," Signal Processing:IC, 19(8), 2004.
- [32]. H. Kuilenburg, M. Wiering, M. Uyl, "A Model Based Method for Automatic Facial Expression Recognition," LNCS 3720, ECML, 2005.

- [33]. S. B. Gokturk, C. Tomasi, b. Girod, "Model-Based Face Tracking for View-Independent Facial Expression Recognition," AFGR, 2002.
- [34]. Y. Tian, T. Kanade, J. F. Cohn, "Recognizing Action Units for Facial Expression Analysis," PAMI, 23(2), 2001.
- [35]. M. Valstar, I. Patras, M. Pantic, "Facial Action Unit Recognition using Temporal Templates," IEEE Workshop on Human Robot Interaction, 2004.
- [36]. M. Pantic, I. Patras, "Dynamic of Facial Expression: Recognition of Facial Actions and Their Temporal Segments From Face Profile Image Sequences," SMC (B), 36(2), 2006.
- [37]. M. Turk, A. Pentland, "Eigenfaces for Recognition," J. Cogn. Neurosci. 3, 71-86, 1991.
- [38]. C. Padgett, G. Cottrell, "Identifying Emotion in Static Images," Proc. 2nd Joint Symp. Neural Computation, vol. 5, La Jolla, CA, 1997.
- [39]. G. Cottrell, J. Metcalf, "Face, Gender and Emotion Recognition using Holons," Advances in Neural Information Processing Systems 3, ed.D. Touretzky, Morgan and Kaufman, San Mateo, 1991.
- [40]. A. Rahardja, A. Sowmya, W. Wilson, "A Neural Network Approach to Component versus Holistic Recognition of Facial Expression in Images," Intelligent Robots and Computer Vision X: Algorithms and Techniques, vol. 1607, SPIE Proc. 1991.
- [41]. A. Lanitis, C. Taylor, T. Cootes, "Automatic Interpretation and Coding of Face Images using Flexible Models," IEEE Trans. Patt. And Mach. Intell. 19, 7, pp. 743-756, 1997.
- [42]. Zhengyou Zhang, Michael Lyons, Michael Schuster, Shigeru Akamatsu, "Comparison Between Geometry-Based and Gabor-Wavelet-Based Facial Expression Recognition Using Multi-Layer Perceptron."
- [43]. M. Lyons, S. Akamatsu, M. Kaamachi, J. Gyoba, "Coding Facial Expression with Gabor Wavelets," Proc. 3rd IEEE Int. Conf. Automatic face and Gesture Recognition, Nara, Japan, pp. 200-205, April 1998.
- [44]. E. Loupiau, N. Sebe, "Wavelet Based Salient Points for Image Retrieval," RR 99.11, Laboratoire Reconnaissance de Formes et Vision, INSA Lyon, March 1999.
- [45]. N. Sebe, Q. Tian, E. Loupiau, M. Lew, T. Huang, "Evaluation of Salient Point Techniques" Springer-Verlag Berlin Heidelberg 2002, LNCS 2383, pp. 367-377, 2002.
- [46]. Ying-li Tian, Takeo Kanade, Jeffrey Cohn, "Recognizing Upper Face Action Units for Facial Expression Analysis" 1063-6919/00\$10.00, IEEE, 2000
- [47]. Y. Yacoob, I. S. Davis, "Recognition Human Facial Expression from Long Image Sequences using Optical Flow," IEEE Trans. On Pattern Analysis and Machine Intelligence, 18(6):636-642, June 1996.
- [48]. M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. "Measuring facial expressions by computer image analysis." Psychophysiology, 36:253-264, 1999.
- [49]. J. Lien, T. Kanade, J. F. Chon, and C. C. Li. "Detection, tracking, and classification of action units in facial expression." Journal of Robotics and Autonomous System, in press.

- [50]. Y. Kwon and N. Lobo. "Age Classification from Facial Images." In Proc. IEEE Conf. Computer Vision and Pattern Recognition, pages 762–767, 1994. [8] J.J.
- [51]. Muid Mufti, Assia Khanam, "Fuzzy Rule Based Facial Expression Recognition" CIMCA-IAWTIC, IEEE, 2006.
- [52]. Muid Mufti, George Vachtsevanos, "Automated Fault Detection and Identification Using a Fuzzy Wavelet Analysis Technique." IEEE AUTOTESTCON '95, Atlanta, August 95, USA.
- [53]. Haihong Zhang, Yan Guo, "Facial Expression Recognition Using Continuous Dynamic Programming" 0-7695-1074-4/01\$10.00, IEEE, 2001.
- [54]. I. Matthews, S. Baker, "Active Appearance Models revisited" International Journal of Computer Vision, Vol 60, No. 2, pp. 135-164, 2004.
- [55]. R. Gross, I. Matthews, S. Baker, "Constructing and Fitting Active Appearance Models with Occlusion" Proceedings of IEEE Workshop on Face Processing in video, 2004.
- [56]. Hyun-Chul Choi, Se-Young Oh, "Real Time Facial Expression Recognition using Active Appearance Model and Multilayer Perceptron" SICE-ICASE International Joint Conference, Oct 18-21, 2006 in Bexco, Busan, Korea, 2006.
- [57] Ezio Malis, "Improving vision-based control using efficient second-order minimization techniques," Proceedings of IEEE International Conference of Robotics and Automation, 2004.
- [58] Selim Benhimane and Ezio Malis, "Real-time image-based tracking of planes using Efficient Second-order Minimization," Proceedings of IEEE IRSJ International Conference on Intelligent Robotics and Systems, 2004, pp. 943-948.
- [59]. Tangqi Tang, Benzai Deng, "Facial Expression Recognition using AAM and Local Facial Features" 3th International Conference on Natural Computation (ICNC), IEEE, 2007.
- [60] Aziz Umit Batur and Monson H. Hayes, "Adaptive Active Appearance Models", IEEE Transactions on Image Processing, Vol. 14, No. 11, pp: 1707-1721, November 2005.
- [61]. Cheng-Chang Lien, yang-Kai Chang, Chin-Chiang Tien, "A Fast Facial Expression Recognition Method at Low Resolution Images" International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2006.
- [62]. A. Utsumi and N. Tetsutani, "Adaptation of appearance model for human tracking using geometrical pixel value distributions", Proceedings of Sixth Asian Conference on Computer Vision, 2004, vol. 2, pp. 794-799.
- [63]. Ying-li Tian, Takeo Kanade, Jeffrey F. Cohn, "Evaluation of Gabor-Wavelet-Based Facial Action Unit Recognition in Image Sequences of Increasing Complexity" fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002.
- [64]. Zhan Yong-Zhao, Ye Jing-Fu, Niu De-Jiao, Cao Peng, "Facial Expression Recognition Based on Gabor Wavelet Transformation and Elastic Templates Matching" Third International Conference on Image and Graphics, IEEE, 2004.
- [65]. O. Bertrand, J. Bohorquez, J. Pernier, "Time-Frequency Digital Filtering Based on an Invertible Wavelet Transform: An Application to Evoked Potentials", IEEE Trans. On Biomedical Eng. Vol. BME -41, No.1, PP. 77-88, Jan. 1994.

- [66]. I. Daubechies, "The Wavelet Transform, Time-Frequency Localization and Signal Analysis", IEEE Trans. On Info. Theory, Vol. IT-36, No. 5, PP. 961-1005, Sep. 1990.
- [67]. J. S. Lim, A. V. Oppenheim, "Advanced Topics in Signal Processing", Prentice Hall, 1977.
- [68]. S. Mallat. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", IEEE Trans. On Pattern Analysis and Machine Intell. Vol. 11, No. 7, PP. 674-693, July 1989.
- [69]. S. Mallat, "Multi Frequency Channel Decomposition of Image and Wavelet Models", IEEE Trans. Acoust., Speech, Signal Proc., Vol. 37, No. 12, PP. 2091-2110, December 1989.
- [70]. O. Rioul, M. Vetterli, "Wavelet and Signal Processing", in IEEE Signal Processing Magazine, PP. 14-38, October 1991.
- [71]. M. B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, L. Raphael, "Wavelet and Their Application, 1992.
- [72]. A. K. Soman, P. P. Vaidyanathan, "On Orthonormal Wavelet and Paraunitary Filter-Banks", IEEE Trans. On Signal Processing, Vol. 41, March 1993.
- [73]. P. P. Vaidyanathan, "Multi Rate System and Filter-Banks", Prentice Hall, 1993.
- [74]. Cohen, A., I. Daubechies, B. Jawerth, P. Vial (1993), "Multiresolution analysis, wavelets and fast wavelet transform on an interval," CRAS Paris, Ser. A, t. 316, pp. 417-421.
- [75]. Coifman, R.R., Y. Meyer, M.V. Wickerhauser (1992), "Wavelet analysis and signal processing in Wavelets and their applications," M.B. Ruskai et al. (Eds.), pp. 153-178, Jones and Bartlett.
- [76]. Boul Alen and peter J. Black" An Effective Real -time Implementation of winger - will Distribution.
- [77]. Osama A. Ahmad and mostafa M. fahmy "critically sampled Gabor transform with localized biorthoGonal function" IEEE, 1998.
- [78]. Ya Chang, Changbo Hu, Matthew Turk, "Manifold of Facial Expression," Computer Science Department, University of California.
- [79]. Zhengyou Zhang, "Feature Based Facial Expression Recognition: Experiments with A Multilayer Perceptron" INRIA Sophia Antipolis, France, February 1998.
- [80]. Yu Song, Kun He, Jiliu Zhou, Zhiming Liu, Kui Li, "Multi Resolution Feature Extraction in Human Face, IEEE, International Conference on Information Acquisition, 2004
- [81]. Vibha S. Vyas, Priti Rege, "Automated Texture Analysis with Gabor Filter" GVIP Journal, Volume 6, Issue 1, July 2006.
- [82]. Hui Zhao, Zhiliang Wang, Jihui Men, "Facial Complex Expression Recognition Based on Fuzzy Kernel Clustering and Support Vector Machines" Third International Conference on Natural Computation, ICNC 2007.
- [83]. P. Ekman, T. Huang, T. Sejnowski, J. Hager, eds., "Final Report to NSF of the Planning Workshop on Facial Expression Understanding" Technical Report, Nat'l Science Foundation, Human Interaction Lab., Univ. of California, San Francsco, 1993.

- [84]. M. Pantic, L. J. M. Rothkrantz, "An Expert System for Multiple Emotional Classification for Facial Expressions" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v.22 n.12, p.1424-1445, December 2000.
- [85]. Spiros V. Ioannou, Amaryllis t. Raouzaoui, Vasilis A. Tzouvaras, Theofilos P. Mailis, Kostas C. Karpouzis, Stefanos D. Kollias, "Emotion Recognition through Facial Expression Analysis Based on A Neurofuzzy Network" *Elsevier, Neural Networks* 18, pp. 423-435, 2005.
- [86]. Duan-Duan Yang, Lian-Wen Jin, Jun-Xun Yin, Li-Xin Zhen, Jian-Cheng Huang, "Facial Expression Recognition with Pyramid Gabor Features and Complete Kernel Fisher Linear Discriminate Analysis" *International Journal of Information Technology*, Vol. 11, No. 9. 2005.
- [87]. Simon Haykin, "Neural Networks A Comprehensive Foundation," second edition, Prentice Hall, 1999.
- [88]. Javad Haddadnia, Majid Ahmadi, "N-feature neural network human face recognition," *Image and Vision Computing*, Vol. 22, pp. 1071-1082, 2004.
- [89]. MATLAB help
- ۹۰ - آر. بیل و تی. جکسون، ترجمه دکتر محمود البرزی، " آشنایی با شبکه های عصبی." موسسه ی انتشارات علمی دانشگاه صنعتی شریف، چاپ اول ۱۳۸۴.
- [91]. Andrew J. Calder, A. Mike Burton, Paul Miller, Andrew W. Young, "A Principal Component Analysis of Facial Expressions" *Elsevier, Vision Research* 41, 2001
- [92]. B. Scholkopf, A. Smola, K. R. Muller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem" *Neural Computation*, Vol. 10, No. 5, pp. 1299-1319, 1998.
- [93]. A. J. Calder, A. M. Burton, P. Miller, A. W. Young, S. Akamatsu, "A Principle Component Analysis of Facial Expression," *Vision Research*. Vol. 41, pp. 1179-1208, 2001.
- [94]. S. Dubussion, F. Devoine, M. Masson, "A Solution for Facial Expression Representation and Recognition" *Signal Processing: Image Communication*. Vol. 17, pp. 657-673, 2002.
- [95]. M. Matsugu, k. Mori, Y. Mitari, Y. Kaneda, "Subject Independent Facial Expression Recognition with Robust Face Detection using A Convolutional Neural Network." *Neural Networks*, Vol. 16, pp. 555-559, 2003.
- [96]. M. Mudrova, A. Prochazka, "Principal Component Analysis in Image Processing," *Institute of Chemical Technology, Prague, Department of Computing and Control Engineering*.
- [97]. Jonathon Shlens, "A Tutorial on Principal Component analysis," *Institute for Nonlinear science, University of California*, Version 2, 2005.
- [98]. Lindsay I Smith, "A Tutorial on Principal Component analysis," February 26, 2002.
- [99]. S. Vaseghi, H. Jetelova, "Principal and Independent Component Analysis in Image Processing," *Brunel University, London*.
- [] P. Ekman, W. V. Friesen, "Unmasking the Face." New Jersey: Prentice Hall, 1975. ۱۰۰
- [] P. Ekman, "Emotion In The Human Face," Cambridge University press, 1982. ۱۰۱
- [102]. D.A. Karras, S.A. Karkanis, B.G. Mertzios, "Image Compression Using the Wavelet Transform on Textural Regions of Interest"; *Univ. of Athens*.

- [103]. M. Mudrov¹, A. Proch²azka, “Principal Component Analysis in Image Processing”, Institute of Chemical Technology, Prague, Department of Computing and Control Engineering.
- [104]. N. Sebe¹, Q. Tian², E. Loup³ias, M. Lew¹, and T. Huang², “Evaluation of Salient Point Techniques”, CIVR 2002, LNCS 2383, pp. 367–377, 2002
- [105]. Praseeda Lekshmi, Dr.M.Sasikumar, “A Neural Network Based Facial Expression Analysis using Gabor Wavelets”, PROCEEDINGS OF WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY, VOLUME 32, AUGUST 2008, ISSN 2070-3740
- [106]. DAW-TUNG LIN, “Facial Expression Classification Using PCA and Hierarchical Radial Basis Function Network” *Department of Computer Science and Information Engineering, National Taipei University, Sanshia, 237 Taiwan*, JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 22, 1033-1046 (2006)

پ-۱-آنالیز فوریه

آنالیزکننده های سیگنال ابزارهای ریاضی قوی ای در دست دارند. شاید از مشهورترین آنها آنالیز فوریه باشد که یک سیگنال را به المانهای سینوسی با فرکانسهای مختلف تجزیه می کند. در واقع آنالیز فوریه یک تکنیک ریاضی است که سیگنال را مطابق شکل پ-۱-۱ از حوزه زمان به حوزه فرکانس می برد [66].



شکل (پ-۱-۱) آنالیز فوریه

می دانیم که سه نوع تکنیک تبدیل فوریه مختلف اما مربوط به هم وجود دارد: انتگرال فوریه، بسط سری فوریه و تبدیل فوریه گسسته^۱.

انتگرال فوریه دو تابع پیوسته را به هم مربوط می سازد: سیگنال و طیف آن. این تبدیل و معکوس آن در حالت یک بعدی به صورت زیر تعریف می شود:

$$F(x) = \int_{-\infty}^{+\infty} f(s) e^{j 2\pi(xs)} ds \quad (\text{پ-۱-۱})$$

بسط سری فوریه ی یک تابع متناوب یا یک تابع گذرا را که بتوان آن را یک سیکل از یک تابع متناوب در نظر گرفت، به صورت رشته ای از ضرایب فوریه با دوره ی محدود یا نامحدود نمایش می دهیم. سری فوریه و معکوسش با تعریف یک متغیر گسسته $S = n\Delta s$ به دست می آیند که L پرپود و $\Delta s = \frac{L}{L}$ است.

$$f(x) = \Delta s \sum_{n=0}^{\infty} F_n e^{j 2\pi(n\Delta s x)} \quad (\text{پ-۱-۲})$$

$$F_n = f(n\Delta s) = \int_0^L f(x) e^{-j 2\pi(n\Delta s x)} dx \quad (\text{پ-۱-۳})$$

تبدیل تبدیل فوریه گسسته یک تابع نمونه برداری شده را با یک طیف نمونه برداری شده نمایش می دهد، به طوری که تعداد نمونه های مستقل در هر دو حوزه یکسان است. این تبدیل با تعریف متغیر گسسته $X = i\Delta x$ اجرا می شود. اگر دارای باند محدود باشد و مطابق تئوری نمونه برداری، نمونه برداری شده باشد داریم $g_i = g(i\Delta x)$ و

¹ Discret Fourier Transform (DFT)

$$g_i = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} G_k e^{j 2\pi i k / N} \quad (\text{پ-۱-۴})$$

$$G_k = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} g_i e^{-j 2\pi k i / N} \quad (\text{پ-۱-۵})$$

در تمام روش‌های سه‌گانه بالا، توابع سینوس و کسینوس با فرکانس‌های مختلف یک دسته از توابع پایه ارتونرمال را شکل می‌دهند. همچنین هر ضریب تبدیل با انجام ضرب داخلی تابع تبدیل شده با یکی از توابع پایه به دست می‌آید. برای تبدیل فوریه گسسته از ضرب داخلی گسسته و توابع پایه گسسته استفاده می‌شود، درحالی‌که ضرب داخلی انتگرالی و توابع پایه پیوسته برای انجام دو تبدیل انتگرال فوریه و بسط سری فوریه استفاده می‌گردند. در هر سه حالت تبدیل معکوس عبارت است از جمع توابع پایه که در ضرایب تبدیل ضرب شده‌اند. این حاصل جمع برای تبدیل فوریه پیوسته به صورت انتگرال در می‌آید [67].

برای بسیاری از سیگنال‌ها، آنالیز فوریه بسیار مفید است زیرا محتوای فرکانسی سیگنال از اهمیت زیادی برخوردار است. اما آنالیز فوریه مشکلاتی به همراه دارد. در انتقال به محدوده فرکانسی، اطلاعات زمانی سیگنال گم می‌شود. وقتی به تبدیل فوریه یک سیگنال نگاه می‌کنیم، این غیر ممکن است که بگوییم که چه موقع یک حادثه ویژه در سیگنال رخ داده است. اگر سیگنال ساکن باشد این مشکلی جدی نیست. البته، بسیاری از سیگنال‌های مورد علاقه و بررسی دانشمندان مشخصه‌ای غیر ساکن و گذرا دارند مانند رانش، تغییرات ناگهانی در ابتدا و انتهای سیگنال. این مشخصات اغلب از مهمترین قسمت‌های سیگنال هستند که تبدیل فوریه برای آشکارسازی آنها مناسب نمی‌باشد. برای رفع این مشکل تبدیل‌های زمان فرکانس و زمان مقیاس از قبیل تبدیل فوریه زمان کوتاه پیشنهاد شده است.

پ-۲- تبدیل فوریه زمان کوتاه

پ-۲-۱- آنالیز سیگنال‌های غیر ایستاد

هدف از آنالیز سیگنال، استخراج اطلاعات و ویژگی‌های مربوط به آن با استفاده از تبدیل‌های آن است. بسیاری از روش‌های تجزیه و تحلیل سیگنال شرط‌های اولیه‌ای را برای آن در نظر می‌گیرند که اگر این شرایط برقرار نباشد ممکن است نتایج کاملاً اشتباهی را ارائه دهند. اکثر تبدیل‌های کلاسیک شرط ایستاد بودن را برای سیگنال در نظر می‌گیرند. متداول‌ترین تبدیلی که به خوبی شناخته شده است تبدیل فوریه است که ویژگی‌های سیگنال را در حوزه‌های فرکانس بیان می‌کند. اما از آنجا که هر تغییر زمانی ناگهانی در سیگنال غیر ایستاد $x(t)$ در کل محور فرکانس گسترده است لذا آنالیز فوریه متداول برای سیگنال‌های غیر ایستاد کاربرد مفیدی نخواهد داشت. به این دلیل یک تجزیه تحلیل مناسب برای سیگنال‌های غیر ایستاد که در موارد عملی بسیار با آنها روبرو هستیم مورد نیاز است [68].

یک روش معمول آنالیز چنین سیگنال‌هایی ایجاد وابستگی به زمان در آنالیز فوریه است که این روش ایده‌ی فرکانس لحظه‌ای را در مورد سیگنال‌های غیر ایستاد مطرح می‌کند. روش دیگر، تصحیح توابع پایه سینوسی استفاده شده در تبدیل فوریه به توابع پایه‌ای است که تمرکز بیشتری را در زمان داشته

باشند. این ایده ها یک نمایش زمان - فرکانس را برای سیگنال ارائه داده و ما را به آنالیز تبدیل فوریه زمان کوتاه و سپس به تبدیل موجک سوق می دهد.

پ-۲-۲- تبدیل فوریه زمان کوتاه زمان پیوسته

سیگنال غیر ایستادن $x(t)$ در زمانهای مختلف از پنجره ی زمانی $g(t)$ دیده می شود و مشخصات سیگنال خروجی استخراج می شود. اگر سیگنال $x(t)$ از طریق پنجره ی $g(t)$ دیده شود در این صورت تبدیل فوریه سیگنال پنجره شده ی $x(t)g(t-\tau)$ تبدیل فوریه ی زمان کوتاه را نتیجه می دهد.

$$X_{STFT}(\tau, f) = \int_{-\infty}^{+\infty} x(t)g(t-\tau) \exp(-j2\pi ft) dt \quad (۶-۱-۶)$$

$X_{STFT}(\tau, f)$ بیانگر محتوای فرکانس سیگنال $x(t)$ در لحظه ی τ است. نتیجه ی چنین آنالیزی به شدت به انتخاب پنجره ی $g(t)$ بستگی دارد. اولین بار دنیس گابر از پنجره ی گوسی $g(t) = g(0)e^{-at^2}$ به این منظور استفاده کرد.

پ-۲-۳- تبدیل فوریه زمان کوتاه زمان گسسته

در این حالت مشابه حالت زمان پیوسته سیگنال $x(n)$ در پنجره $g(n-m)$ ضرب شده و تبدیل فوریه حاصل ضرب $x(n)g(n-m)$ برای شیفت های متفاوت پنجره محاسبه می شود. به عبارت دیگر $X_{STFT}(e^{j\omega}, m)$ به ازای مقادیر مختلف m از رابطه ی زیر به دست می آید.

$$X_{STFT}(e^{j\omega}, m) = \sum_{-\infty}^{+\infty} x(n)g(n-m)e^{-j\omega n} \quad (۷-۱-۷)$$

پ-۳-رزولوشن فرکانسی، دقت زمانی، اصل عدم قطعیت

با توجه به نمودار فرکانس زمان اگر بخواهیم بهترین تصمیم گیری را در مورد ماهیت سیگنال و فرکانس های موجود در آن در زمان مشخص t_1 داشته باشیم باید پنجره زمانی به مرکز t_1 و دارای حداقل تداوم زمانی ممکن باشد. به عبارت دیگر دقت زمانی تبدیل، به طول زمانی پنجره بستگی دارد و هر چه تداوم زمانی پنجره کمتر باشد دقت زمانی بیشتری خواهد داشت. به طور مشابه اگر به دنبال مناسب ترین تعبیر سیگنال در یک فرکانس معین مثل f_1 باشیم لازم است که پهنای باند فیلتر میان گذر به مرکز f_1 حداقل مقدار ممکن باشد که این موضوع به رزولوشن فرکانسی تعبیر می شود. یعنی هرچه پهنای باند فیلتر میان گذر کمتر باشد رزولوشن فرکانس بالاتری خواهیم داشت [69].

این دو مفهوم یعنی دقتهای فرکانسی و زمانی بیانگر یکی از مشکلات مهم در آنالیز تبدیل فوریه زمان کوتاه است که همانا انتخاب پنجره با طول مناسب است. افزایش دقت زمانی کاهش رزولوشن فرکانسی را به دنبال دارد و بر عکس کاهش دقت زمانی افزایش دقت فرکانسی را باعث می شود.

از طرفی هدف اصلی استفاده از پنجره ی تبدیل فوریه زمان کوتاه بالابردن دقت های زمانی فرکانسی است که با توجه به مطالب فوق، رسیدن به این دو هدف به طور همزمان دارای محدودیت هایی است.

دقت زمانی تبدیل فوریه زمان کوتاه به وسیله ی Δt مشخص می شود:

$$\Delta t^2 = \frac{\int t^2 |g(t)|^2 dt}{\int |g(t)|^2 dt} \quad (\text{پ-۱-۸})$$

رزولوشن فرکانسی در تبدیل فوریه زمان کوتاه به وسیله ی Δf مشخص می شود:

$$f^2 = \frac{\int f^2 |G(f)|^2 df}{\int |G(f)|^2 df} \quad (\text{پ-۱-۹})$$

باید توجه کنیم که رزولوشن فرکانسی و دقت زمانی را نمی توان به طور همزمان و به اندازه ی دلخواه کوچک کرد. زیرا حاصل ضرب آن دارای حد پایینی است.

$$\Delta t \Delta f \geq \frac{1}{4\pi} \quad (\text{پ-۱-۹})$$

این مطلب به اصل عدم قطعیت هایزنبرگ^۱ معروف است و بدین معنی است که فقط می توان یک مصالحه بین دقت زمانی و رزولوشن فرکانسی در نظر گرفت [76].

پ-۴- تبدیل موجک

هر چند که تبدیل فوریه زمان کوتاه ابزار مفیدی برای آنالیز سیگنال های غیر ایستان است اما این تبدیل خالی از عیب نیست. مهمترین اشکال آنالیز تبدیل فوریه زمان کوتاه این است که برای یک پنجره ی دلخواه دقت های زمانی و فرکانسی به دست آمده، بر روی تمام بازه ی زمان-فرکانس مقدار ثابتی است. اگر $x(t)$ سیگنال فرکانس بالا باشد تعداد زیادی از پریودهای آن توسط پنجره ثابت در تبدیل فوریه زمان کوتاه در بر گرفته می شود و اگر $x(t)$ سیگنال فرکانس پایین باشد پنجره ثابت، تعداد کمی از سیکل ها را در بر می گیرد. بنابراین انتخاب پنجره با طول زیاد باعث دقت زمانی کم در فرکانس بالا می شود و بر عکس انتخاب یک پنجره باریک، رزولوشن فرکانسی کمی را در فرکانس های پایین به همراه خواهد داشت. لذا سیگنالی متشکل از فرکانس های مختلف را نمی توان با تبدیل فوریه زمان کوتاه به نحو بهینه تجزیه و تحلیل کرد [70].

برای رفع مشکل فوق یک روش این است که به Δt و Δf در صفحه ی زمان-فرکانس اجازه ی تغییرات داده شود تا بتوان تجزیه و تحلیل مولتی رزولوشن را به دست آورد. به عبارت دیگر برای استخراج ویژگی های فرکانس بالای سیگنال باید تابع پنجره ای با تداوم زمانی کم و برای استخراج ویژگی های فرکانس های پایین سیگنال باید پنجره ای با طول زیاد انتخاب کرد. به عبارت دیگر در تحلیل فیلتر بانکی، رزولوشن زمانی باید با فرکانس مرکزی فیلتر بانک های آنالیز افزایش یابد. این مطلب پایه و اساس تبدیل موجک است. در بخش بعد به معرفی اجمالی انواع تبدیل های موجک می پردازیم.

پ-۴-۱- انواع تبدیل موجک

¹ Heisenberg

همانند تبدیل فوریه تبدیل موجک دارای سه نوع می‌باشد: تبدیل موجک پیوسته^۱، بسط سری موجک و تبدیل موجک گسسته^۲.

باید توجه کنید توابع پایه موجک حتی اگر ارتونرمال نیز نباشند می‌توانند جهت انجام تبدیل مورد استفاده قرار گیرند. این بدان معناست که به عنوان مثال بسط سری موجک ممکن است یک تابع با باند محدود را با تعداد نامحدودی از ضرایب نمایش دهد. چنانچه تعداد محدودی از این رشته‌ی نامحدود ضرایب انتخاب شوند، در صورت انجام تبدیل معکوس تنها تقریبی از تابع اصلی به دست خواهد آمد. به طور مشابه تبدیل موجک گسسته ممکن است به تعداد نمونه‌ی بیشتری نسبت به تابع اصلی نیاز داشته باشد تا بتواند آن را به طور کامل بازسازی کند یا حتی تقریبی قابل قبول از آن به دست دهد.

دسته توابعی که با آنها کار خواهیم کرد عبارت است از توابعی که انتگرال مربع آنها در طول محور حقیقی محدود باشد. این نوع توابع را به صورت $L^2(R)$ نمایش می‌دهند. بنابراین $f(x) \in L^2(R)$ فیدین معناست که:

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx < \infty \quad (\text{پ-۱-۱۰})$$

در تحلیل موجک دسته‌ای از توابع پایه را با تغییر مقیاس و انتقال یک تابع اولیه $(\psi(x))$ که آن را موجک اساسی می‌نامند به وجود می‌آورند. این تابع دارای نوسانات محدودی است و معمولاً در یک مکان متمرکز شده‌است و اگر $|x| \rightarrow \infty$ مقدار آن به سرعت به سمت صفر میل می‌کند. بنابراین داریم:

$$\psi(x) \in L^2(R) \quad (\text{پ-۱-۱۱})$$

پ-۱-۴-۱-۱- تبدیل موجک پیوسته

تبدیل موجک پیوسته که انتگرال موجک نیز نامیده می‌شود توسط گراسمن و مورلت پایه‌گذاری شده‌است. به تحلیل ساده ریاضی آن توجه کنید.

تعریف: اگر $\psi(x)$ یک تابع با مقادیر حقیقی باشد و طیف آن $(\Psi(s))$ معیار قبولی بیان شده در رابطه ی زیر را برقرار نماید در آن صورت $\psi(x)$ یک موجک اساسی خواهد بود.

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\Psi(s)|^2}{|s|} ds < \infty \quad (\text{پ-۱-۱۲})$$

البته با توجه به این که S در مخرج انتگرال قرار دارد، باید داشته باشیم:

$$\Psi(0) = 0 \Rightarrow \int_{-\infty}^{+\infty} \psi(x) dx = 0 \quad (\text{پ-۱-۱۳})$$

همچنین از آنجایی که داریم $\Psi(\infty) = 0$ می‌توانیم مشاهده کنیم که دامنه طیف یک موجک قابل قبول شبیه به تابع تبدیل یک فیلتر میان‌گذر باشد. در واقع، هر پاسخ ضربه فیلتر میان‌گذر با متوسط صفر که با افزایش فرکانس به سرعت میرا می‌شود می‌تواند به عنوان یک موجک اساسی برای تبدیل موجک استفاده شود.

¹ Continuous Wavelet Transform (CWT)
² Discret Wavelet Transform (DWT)

دسته‌ای از توابع پایه موجک که با علامت $\psi_{a,b}(x)$ نمایش داده می‌شوند را می‌توان با انتقال و تغییر مقیاس موجک اساسی به دست آورد:

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \quad (\text{پ-۱-۱۴})$$

که در این رابطه $a > 0$ است و اعداد a و b حقیقی هستند. متغیر a معرف مقیاس (پهنای) یک تابع پایه به خصوص است و متغیر b میزان انتقال را در طول محور x (اعداد حقیقی) نشان می‌دهد. معمولاً $\psi(x)$ در مرکز مختصات قرار دارد و بنابراین $\psi_{a,b}(x)$ در $x=b$ تمرکز یافته است. انتگرال موجک مربوط به $f(x)$ با استفاده از $\psi(x)$ به صورت زیر خواهد بود:

$$W_f(a,b) = \langle f, \psi_{a,b} \rangle = \int_{-\infty}^{+\infty} f(x) \psi_{a,b}(x) dx \quad (\text{پ-۱-۱۵})$$

یعنی تبدیل پیوسته موجک نیز به صورت ضرب داخلی تابع ورودی و توابع پایه تعریف می‌شود. گراسمن و مورلت نشان داده اند که تبدیل موجک پیوسته معکوس به صورت زیر محاسبه می‌شود:

$$f(x) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} W_f(a,b) \psi_{a,b}(x) db \frac{da}{a^2} \quad (\text{پ-۱-۱۶})$$

ضریب مقیاس $\frac{1}{\sqrt{a}}$ که در سمت راست رابطه پ-۱-۱۴ ظاهر می‌شود جهت نرمالیزاسیون توابع پایه می‌باشد، زیرا داریم:

$$\left\| f\left(\frac{x-b}{a}\right) \right\| = \sqrt{\int_{-\infty}^{+\infty} \left| f\left(\frac{x-b}{a}\right) \right|^2 dx} = \sqrt{a} \|f(x)\| \quad (\text{پ-۱-۱۷})$$

از آنجایی که موجک اساسی دارای متوسط صفر است تمام توابعی که از تغییر مقیاس و انتقال آن به دست می‌آیند، نیز دارای متوسط صفر خواهند بود. تبدیل موجک پیوسته می‌تواند در ابعاد مختلف مثلاً در دو بعد بسط داده شود [72,73].

پ-۱-۱-۴-۱- تبدیل موجک پیوسته دوبعدی

تبدیل موجک پیوسته $W(a,b)$ از یک تابع یک بعدی $f(x)$ به صورت تابعی دو متغیره می‌باشد. یعنی یک متغیر بیشتر از $f(x)$ دارد و به همین جهت تبدیل موجک پیوسته تابع اولیه را با حجم بیشتری از اطلاعات نمایش می‌دهد. تعداد متغیرهای این تبدیل به ازای اضافه شدن هر متغیر تابع ورودی، یکی اضافه می‌شود. اگر $f(x,y)$ یک تابع دو بعدی باشد، تبدیل موجک آن به صورت زیر خواهد بود:

$$W_f(a,b_x,b_y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) \psi_{a,b_x,b_y}(x,y) dx dy \quad (\text{پ-۱-۱۸})$$

که b_x و b_y معرف انتقال در دو بعد می‌باشند. تبدیل موجک معکوس دوبعدی گسسته عبارت است از:

$$f(x,y) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_f(a,b_x,b_y) \psi_{a,b_x,b_y}(x,y) db_x db_y \frac{da}{a^3} \quad (\text{پ-۱-۱۹})$$

که در رابطه فوق $\psi(x,y)$ موجک اساسی دوبعدی بوده و داریم:

$$\psi_{a,b_x,b_y}(x,y) = \frac{1}{|a|} \psi\left(\frac{x-b_x}{a}, \frac{y-b_y}{a}\right) \quad (\text{پ-۱-۲۰})$$

می‌توان با استفاده از مفهوم بانک فیلتری تبدیل موجک پیوسته را به صورت رابطه کانولوشن نمایش داد. در این جا نیز یک افزونگی وجود دارد، یعنی می‌توان فیلترهای بی‌شماری را به این منظور استفاده نمود. در واقع هر $\Psi(u,v)$ تابع تبدیل متناظر با $(\psi(x,y))$ که در هر نقطه روی محور فرکانس بجز مبدأ غیر صفر باشد، می‌تواند به عنوان یک فیلتر مورد قبول در تبدیل موجک در نظر گرفته شود و تصویر اصلی از نظر تئوری با استفاده از خروجی این فیلترها قابل بازسازی است. بنابراین نتیجه کار عبارتست از تبدیل موجک پیوسته که سیگنال یا تصویر اولیه را بصورت فشرده‌ای نمایش نمی‌دهد، اما تجزیه و تحلیل سیگنال یا تصویر با استفاده از تبدیل بدست آمده می‌تواند به صورتی مؤثرتر انجام شود.

پ-۱-۱-۲- بیان تبدیل موجک پیوسته بر اساس تئوری بانک فیلتری

می‌توان با استفاده از مفهوم بانک فیلتری تبدیل موجک پیوسته را به صورت رابطه کانولوشن نمایش داد. ابتدا تابع موجک پایه در مقیاس a را به صورت زیر تعریف می‌کنیم:

$$\psi_a(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x}{a}\right) \quad (\text{پ-۱-۲۱})$$

این تابع یک موجک پایه با مقیاس a و نرمالیزه شده با $a^{-1/2}$ می‌باشد و یک دسته از توابعی را که با افزایش a عریض تر می‌شوند تعریف می‌کند. همچنین با تعریف $\tilde{\psi}_a(x)$ به صورت زیر:

$$\tilde{\psi}_a(x) = \psi_a^*(-x) = \frac{1}{\sqrt{a}} \psi^*\left(-\frac{x}{a}\right) \quad (\text{پ-۱-۲۲})$$

که در آن $\tilde{\psi}_a(x)$ مزدوج مختلط منعکس شده‌ی $\psi_a(x)$ می‌باشد. اگر $\psi(x)$ حقیقی و زوج باشد که همواره با این حالت سروکار داریم عمل انعکاس و مزدوج‌گیری بی‌تأثیر خواهد بود. اکنون می‌توان تبدیل موجک پیوسته را به صورت زیر نوشت:

$$W_f(a,b) = \int_{-\infty}^{\infty} f(x) \tilde{\psi}_a(b-x) dx = f * \tilde{\psi}_a \quad (\text{پ-۱-۲۳})$$

هر مقدار a ، فیلتر میان‌گذر متفاوتی را تعریف می‌کند. از رابطه (پ-۱-۱۶) خواهیم داشت:

$$f(x) = \frac{1}{C_\psi} \int_0^\infty \int_{-\infty}^\infty [f * \tilde{\psi}_a](b) \psi_a(b-x) db \frac{da}{a^2} = \frac{1}{C_\psi} \int_0^\infty [f * \tilde{\psi}_a * \psi_a](x) \frac{da}{a^2} \quad (\text{پ-۱-۲۴})$$

رابطه اخیر دلالت بر این دارد که ترکیب خروجی‌های فیلترها، $f(x)$ را به دست می‌دهد. این موضوع بیست سال قبل از گراسمن و مورلت، توسط کالدرون^۱ مطرح شده‌است. لازم به یادآوری است که با توجه

به خاصیت $\mathfrak{T}\{f(ax)\} = \frac{1}{|a|} F\left(\frac{s}{a}\right)$ خواهیم داشت:

$$\psi_a(s) = \mathfrak{T}\{\psi_a(x)\} = \sqrt{a} \Psi(as) \quad (\text{پ-۱-۲۵})$$

^۱ Calderon

یعنی با افزایش a فرکانس مرکزی مربوط به فیلترهای میان‌گذر کاهش یافته و توابع تبدیل در حوزه فرکانس باریک‌تر می‌شوند.

پ-۴-۱-۲- بسط سری موجک

نوع دوم از تبدیل موجک قدری از نوع اول محدودتر است. در این‌جا نیز یک موجک، تغییر مقیاس یافته و انتقال داده می‌شود تا یک دسته از توابع پایه را شکل دهد. البته در این‌جا میزان تغییر مقیاس و انتقال به جای اعداد حقیقی، توسط اعداد صحیح مشخص می‌شوند.

پ-۴-۱-۲-۱- موجک دیادیک

در بسط سری موجک توابع پایه با تغییر مقیاس باینری یعنی انقباض با ضریب دو، موجک اساسی $(\psi(x))$ و انتقال دیادیک آن را به وجود می‌آورد. انتقال دیادیک عبارتست از انتقالی به اندازه $k/2^j$.

می‌دانیم تابع $\psi(x)$ در صورتی یک موجک ارتوگونال است که خانواده‌ی $\{\psi_{j,k}(x)\}$ توابع پایه ارتونرمال در $L^2(\mathbb{R})$ را شکل دهد که در آن $\{\psi_{j,k}(x)\}$ مجموعه‌ای از توابع زیر می‌باشد:

$$\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k) \quad (\text{پ-۱-۲۶})$$

اعداد صحیح $j < \infty$ و $-\infty < k$ به ترتیب میزان n تغییر مقیاس و میزان انتقال را نشان می‌دهند. خانواده موجک فوق در صورتی ارتونرمال است که اولاً داشته باشیم:

$$\langle \psi_{j,k}, \psi_{l,m} \rangle = \delta_{j,l} \delta_{k,m} \quad (\text{پ-۱-۲۷})$$

که در این رابطه اعداد l و m صحیح هستند و $\delta_{j,k}$ تابع دلتای کرونکر است و $\langle \cdot, \cdot \rangle$ نشان‌دهنده ضرب داخلی می‌باشد. ثانیاً باید بتوان هر تابع $f(x) \in L^2(\mathbb{R})$ را بصورت زیر نوشت:

$$f(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_{j,k} \psi_{j,k}(x) \quad (\text{پ-۱-۲۸})$$

که ضرایب تبدیل نیز با ضرب داخلی و بصورت زیر بدست می‌آیند:

$$c_{j,k} = \langle f(x), \psi_{j,k}(x) \rangle = 2^{j/2} \int_{-\infty}^{\infty} f(x) \psi(2^j x - k) dx \quad (\text{پ-۱-۲۹})$$

روابط بالا بسط سری موجک را برای $f(x)$ برحسب $\psi(x)$ نشان می‌دهند. در این‌جا نیز می‌بینیم که یک تابع پیوسته با استفاده از یک رشته نامحدود دوطرفه نمایش داده می‌شود و در حالت کلی تبدیل مذکور فوق کامل خواهد بود. از آنجایی که توابع پایه از دو طرف نامحدود هستند، بازسازی کامل در صورت استفاده از تمام جمله‌ها امکان‌پذیر خواهد بود. البته چنانچه $\psi(x)$ بطور مناسب انتخاب شود می‌توان سری را با تعداد جملات محدود تقریب زد به طوری که خطای حاصله ناچیز باشد. اگر $f(x)$ دارای دوره محدود بوده و موجک اساسی نیز در خارج از محدوده معین با سرعت میرا گردد، بسیاری از ضرایب با

$|k|$ بزرگ قابل صرف نظر خواهند بود. به طور مشابه، ضرایب با $|j|$ بزرگ معمولاً به اندازه کافی کوچک می‌شوند.

پ-۴-۱-۲-۲- موجک دیادیک فشرده

اگر $f(x)$ و $\psi(x)$ در خارج از محدوده $[0,1]$ صفر باشند، خانواده توابع پایه ارتونرمال می‌تواند فقط با یک اندیس n بطور کامل مشخص شود:

$$\psi_n(x) = 2^{j/2} \psi(2^j x - k) \quad (\text{پ-۱-۳۰})$$

که j و k عملاً توابعی از n هستند، بدین صورت که برای $k=0,1,\dots,2^j-1$ و $j=0,1,\dots$ داریم:

$$n=2^j+k \quad (\text{پ-۱-۳۱})$$

تبدیل معکوس نیز بصورت زیر می‌باشد:

$$f(x) = \sum_{n=0}^{\infty} c_n \psi_n(x) \quad (\text{پ-۱-۳۲})$$

که البته فرض می‌شود $\psi_0(x) = 1$ و ضرایب تبدیل با استفاده از رابطه زیر بدست می‌آیند:

$$C_n = \langle f(x), \psi_n(x) \rangle = 2^{j/2} \int_{-\infty}^{\infty} f(x) \psi(2^j x - k) dx \quad (\text{پ-۱-۳۳})$$

در این جا یک تابع پیوسته توسط یک رشته نامحدود (همانند بسط سری فوریه) نمایش داده می‌شود. افزونگی موجود در انتگرال موجک در این جا صادق نمی‌باشد زیرا اگر یک یا چند $\psi_n(x)$ مشابه با $f(x)$ باشد، می‌توان سری را با تقریب خوبی محدود در نظر گرفت و از ضرایب کوچک صرف نظر نمود. در اینجا می‌توان اصول تبدیل موجک گسسته را بیان کرد. اگر $f(i \Delta t)$ یک تابع گسسته با N نقطه باشد، و $\psi(x)$ یک موجک دیادیک فشرده باشد، می‌توان با استفاده از صورت گسسته روابط (پ-۱-۳۲) و (پ-۱-۳۳) تبدیل موجک گسسته را محاسبه کرد. هر دوی این روابط بصورت حاصل جمع های N نقطه‌ای در می‌آیند.

پ-۴-۱-۲-۳- موجک هار^۱

تبدیل موجک هار یکی از اولین مثال‌ها برای تبدیل موجک از نوع ارتونرمال، دیادیک و فشرده می‌باشد. توابع پایه‌اش با انتقال و تغییر مقیاس تابع هار (بعنوان موجک اساسی) تعریف می‌شوند. تابع هار یک جفت پالس مستطیلی فرد و ساده‌ترین و قدیمی‌ترین موجک اساسی با دوره محدود می‌باشد. تابع هار در هر تغییر مقیاس به اندازه ضریب $1/2$ باریک می‌شود و کاهش مقیاس می‌یابد. سپس هر موجک که باریک‌تر شده، متناسب با عرضش انتقال می‌یابد و بنابراین دسته کاملی از موجک‌ها تمام فاصله را می‌پوشانند. از آنجایی که مقیاس موجک تقسیم بر دو می‌شود، برای آنکه خاصیت ارتونرمال بودن باقی بماند، دامنه‌اش در $\sqrt{2}$ ضرب می‌شود. نتیجه‌ی تمام این کارها یک دسته از توابع پایه ارتونرمال است.

¹ Haar Wavelet

پ-۴-۱-۳- تبدیل موجک گسسته^۱

تبدیل موجک گسسته بسیار شبیه به تبدیلات همسانی^۲ است. این تبدیل برای فشرده‌سازی، پردازش و آنالیز تصویر بطور وسیعی استفاده می‌شود. با استفاده از یک دسته از توابع پایه ارتونرمال، می‌توان تبدیل موجک گسسته را محاسبه کرد. روش‌های گوناگونی برای معرفی الگوریتم تبدیل موجک گسسته مورد استفاده اند که به تحلیل مولتی‌رزولوشن یا زمان - مقیاس و الگوریتم‌های هرمی اشاره می‌کنیم.

الف- تحلیل مولتی رزولوشن^۳

تئوری بانک فیلتری یک مفهوم مناسب برای نمایش سیگنال‌های مرکب از مؤلفه‌های نوسانی (مانند نوت‌های موسیقی و تپش‌ها) ارائه می‌دهد. این مؤلفه‌ها شامل چندین سیکل از نوسان در دامنه‌شان می‌شوند. اما در تحلیل تصویر، مؤلفه‌های مهم و متمرکز در یک نقطه بطور کامل نوسانی نیستند و بعنوان مثال تنها یک سیکل یا حتی یک قسمت از یک سیکل را مانند خط‌ها، لبه‌ها و نقطه‌ها شامل می‌شوند. اشیاء در یک تصویر اندازه‌های متفاوت دارند. به عنوان مثال یک لبه می‌تواند یک تغییر تیز از مشکی به سفید و یا یک تغییر تدریجی در یک فاصله زمانی قابل ملاحظه باشد. برای درک این مطلب می‌توان نقشه‌برداری را مثال زد که نقشه‌ها در مقیاس‌های متفاوتی رسم می‌شوند. مقیاس یک نقشه عبارتست از نسبت اندازه حقیقی یک منطقه به اندازه نشان داده شده‌اش در نقشه. در مقیاس‌های بزرگ (مثل کره زمین) خصوصیات عمده مثل قاره‌ها و دریاها قابل رؤیت می‌باشند و جزئیاتی مثل خیابان‌های شهر قابل مشاهده نخواهند بود. در مقیاس‌های کوچک‌تر، جزئیات قابل مشاهده هستند و کلیات حذف می‌شوند. بنابراین برای دسترسی به یک نقطه در یک مکان دور نیاز به یک دسته از نقشه‌ها در مقیاس‌های متفاوت داریم. تبدیل موجک بر مبنای این تحلیل مولتی‌رزولوشن به وجود آمده است. همانند تحلیل زمان-فرکانس یک سیگنال که یک سیگنال در یک فضای دوبعدی نمایش داده می‌شود، در این‌جا محور عمودی بجای فرکانس "مقیاس" است. تغییر مقیاس در این‌جا با انبساط و انقباض موجک اساسی برای ایجاد توابع پایه تبدیل مورد استفاده صورت می‌پذیرد. موجک اساسی $(\psi(x))$ بصورت $\psi(x/a)$ تغییر مقیاس می‌یابد یعنی با $a > 1$ عریض و با $a < 1$ باریک می‌شود. با مقادیر بزرگ a ، توابع پایه منبسط شده برای یافتن ویژگی‌های بزرگ به کار برده می‌شوند و با مقادیر کوچک a ، جزئیات تصویر مد نظر قرار می‌گیرند.

ب- الگوریتم‌های هرمی

فرض کنید که از یک تصویر دیجیتالی با ۱۰۲۴ در ۱۰۲۴ پیکسل، ۱۰ تصویر دیگر با نمونه‌برداری یک در میان یعنی حذف سطر و ستون‌های زوج تصویر در ۱۰ مرحله متوالی، به دست آوریم. در نتیجه تصاویر

¹ Discrete wavelet transform (DWT)

² Unitary transform

³ Multiresolution analysis

۵۱۲×۵۱۲، ۲۵۶×۲۵۶، ... و نهایتاً ۱×۱ خواهیم داشت. حال اگر بعنوان مثال آشکارسازی لبه را روی هر یک از تصاویر فوق اجرا کنیم و از یکی از اپراتورهای آشکارساز لبه ۳×۳ استفاده گردد، لبه‌های تیز در تصویر اصلی آشکارسازی شده، درحالی‌که تغییرات آهسته روشنایی در تصاویر ۵۱۲×۵۱۲ و ۲۵۶×۲۵۶ پیکسلی و لبه‌های خیلی نرم در تصاویر ۱۶×۱۶ پیکسلی و تصویرهای کوچک‌تر ظاهر می‌گردد. تبدیل هار نمونه‌ای از روش فوق است که در حدود یک قرن پیش معرفی گردید. با ملاحظه تصاویر پایه تبدیل هار می‌توان مفهوم جستجوی لبه‌ها در تصویر توسط اپراتورها با اندازه‌های متفاوت را مشاهده نمود. ممکن است ادعا شود که تمام لبه‌ها (چه کوچک و چه بزرگ) در تصویر اصلی ۱۰۲۴×۱۰۲۴ پیکسلی ملاحظه می‌گردد و هیچ تغییری در رزولوشن برای آشکارسازی آن‌ها نیاز نیست. ولی مسئله در این جاست که لبه‌های بزرگ (تغییرات تدریجی سطوح خاکستری) برای آشکارسازی توسط اپراتورهای پنجره‌ای متداول با مشکل مواجه می‌شوند. البته می‌توان مقیاس اپراتور را زیاد کرد (پنجره‌های بزرگ‌تر) تا لبه‌های بزرگ آشکار شوند اما مؤثرتر این است که مقیاس تصویر کوچک شود زیرا استفاده از یک اپراتور بزرگ برای جاروب تصویر با رزولوشن بالا برای یافتن لبه‌های بزرگ از نظر محاسباتی مقرون به صرفه نمی‌باشد. انواع مختلف تحلیل مولتی‌رزولوشن تحت نام‌های مختلف طی سال‌ها مطالعه شده‌اند. اما تنها در سال‌های اخیر تشابه اساسی بین مولتی‌رزولوشن و بانک فیلتری تشخیص داده شده‌است که با عنوان "تبدیلات موجک" مطرح گردیده است.

پس از توضیح مختصر این روش‌های معرفی الگوریتم تبدیل موجک گسسته را بیان می‌کنیم. برای تعریف تبدیل موجک زمان گسسته سعی می‌کنیم مسئله را از از روش حوزه فرکانس حل کنیم. با فرض $a=2$ و بدون در نظر گرفتن ضریب $\sqrt{a^k}$ که مربوط به نرمالیزاسیون انرژی است معادل رابطه‌ی $H_k(f) = \sqrt{a^k} H(a^k f)$ در تعریف تبدیل موجک زمان پیوسته، برای فیلترهای دیجیتال داریم:

$$H_k(e^{j\omega}) = H(e^{j2k\omega}) \quad (\text{پ-۱-۳۴})$$

$$f_k(n) = h_k^*(-n) \quad (\text{پ-۱-۳۵})$$

این رابطه معادل $H_k(z) = H(z^{2^k})$ در حوزه Z است. فیلترهای $H_k(e^{j\omega})$ باند های اضافی نامطلوبی را شامل می‌شود که به منظور حذف آن‌ها، فیلتر پایین‌گذر $G(z)$ را با آن‌ها سری می‌کنیم.

$$H(z), G(z)H(z^2), G(z)G(z^2)H(z^4), \dots \quad (\text{پ-۱-۳۶})$$

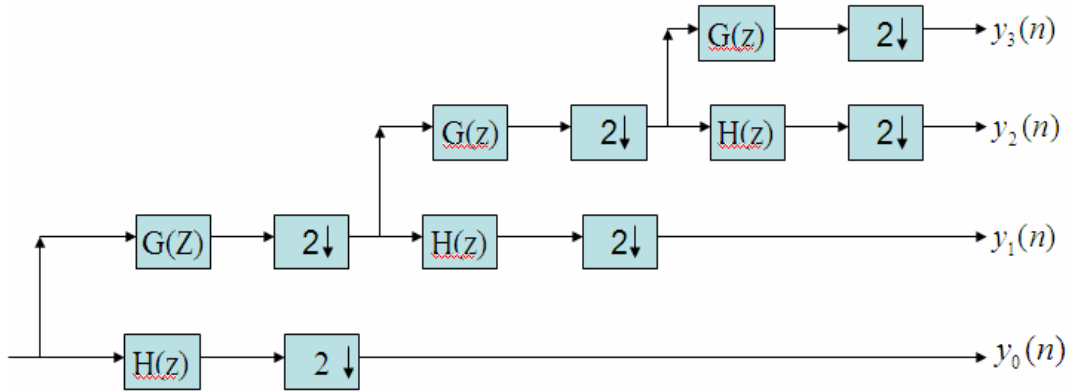
با این تفاوت که فقط نرمالیزاسیون انرژی منظور نشده‌است که این کار نیز با انتخاب مناسب دامنه‌ی فیلترهای $H(z)$ و $G(z)$ امکان‌پذیر است. فیلترهای مشخص شده در رابطه‌ی (پ-۱-۳۶) را می‌توان با ساختار درختی و یا معادل غیردرختی آن، مطابق شکل (پ-۱-۲) نمایش داد، که در حالت سه سطحی رسم شده‌است. با توجه به این شکل می‌توان نوشت:

$$X_k(z) = H_k(z)X(z) \Leftrightarrow X_k(n) = \sum_m x(m)h_k(n-m) \quad (\text{پ-۱-۳۷})$$

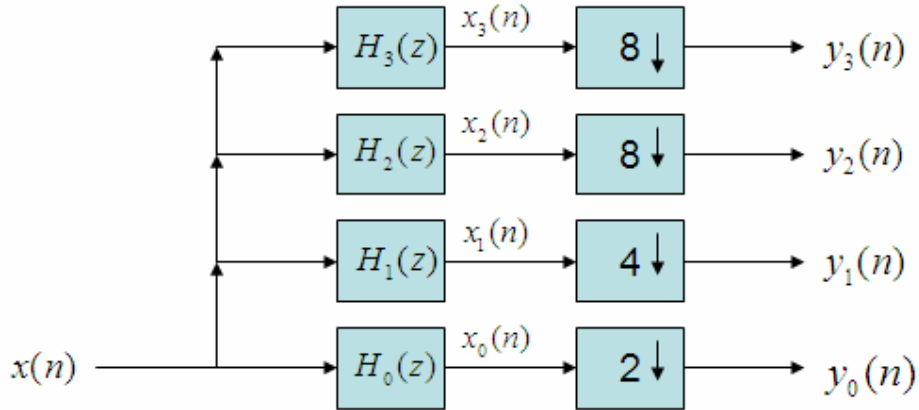
و از آنجایی‌که $y_k(n)$ ها ضرایب تبدیل موجک هستند، تعریف زمانی تبدیل موجک گسسته را می‌توان بصورت زیر نوشت:

$$\begin{cases} X_{DTWT}(k, m) = \sum_{n=-\infty}^{+\infty} x(n)h_k(2^{k+1}m - n) \\ X_{DTWT}(M-1, m) = \sum_{n=-\infty}^{+\infty} x(n)h_{M-1}(2^{M-1}m - n) \end{cases} \quad \begin{matrix} (پ-۱-۳۸) \\ 0 \leq k \leq M-2 \end{matrix}$$

(a)



(b)



شکل (پ-۱-۲): (a) تبدیل موجک با استفاده از سیستم QMF دو دویی با ساختار درختی سه سطحی. (b) سیستم چهار کاناله معادل.

پ-۴-۱-۴- عکس تبدیل موجک زمان گسسته

برای بدست آوردن تبدیل عکس، فرض می‌کنیم که بانک سنتز شکل (پ-۱-۳) را به گونه‌ای طرح کرده‌ایم که به همراه بانک آنالیز شکل (پ-۱-۲) تشکیل سیستمی PR دهند. با توجه به این‌که:

$$F_2(z) = G_s(z) G_s(z^2) H_s(z^4), F_1(z) = G, F_0(z) = H_s(z) \quad (پ-۱-۳۹)$$

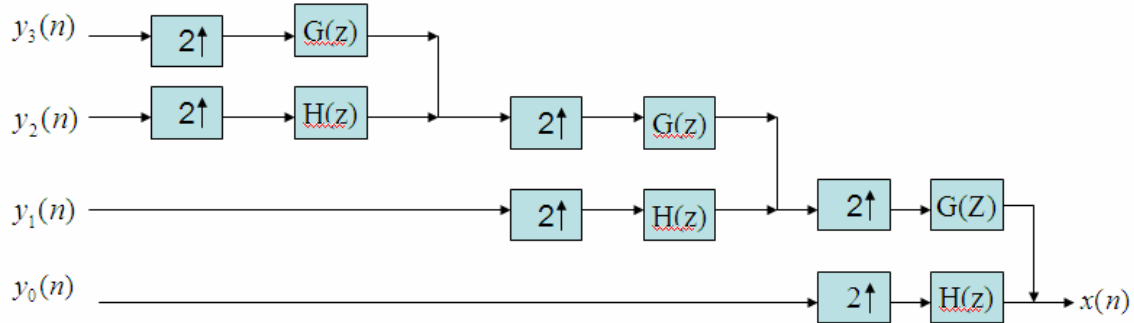
و با توجه به PR بدون سیستم:

$$X_0(z) = F_0(z)y_0(z^2) + F_1(z)y_1(z^4) + \dots + F_{M-2}(z)y_{M-2}(z^{2^{M-1}}) + F_{M-1}(z)y_{M-1}(z^{2^{M-1}}) \quad (\text{پ-۱-۴۰})$$

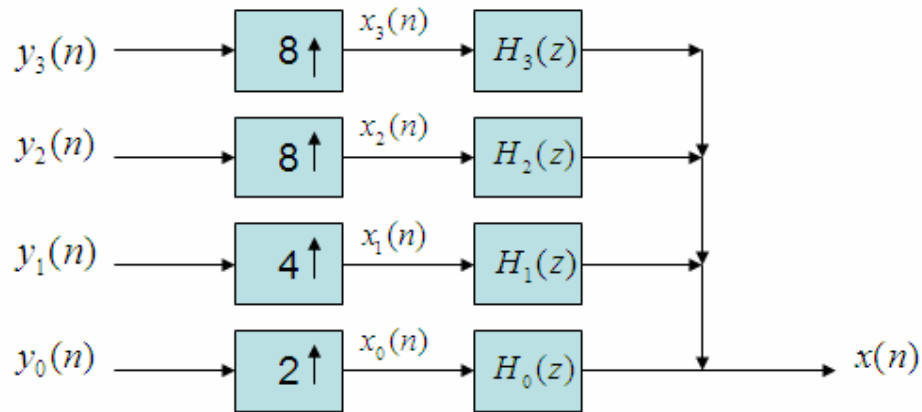
که معادل این رابطه در حوزه‌ی زمان بصورت زیر است:

$$X(n) = \sum_{k=0}^{M-2} \sum_{m=-\infty}^{+\infty} y_k(m) f_k(n - 2^{k+1}m) + \sum_{m=-\infty}^{+\infty} y_{M-1}(m) f_{M-1}(n - 2^{M-1}m) \quad (\text{پ-۱-۴۱})$$

(الف)



(ب)



شکل (پ-۱-۳) الف - بانک سنتز ب - معادل غیر درختی آن.

که با تعریف توابع پایه $\eta_{km}(n)$ بصورت زیر:

$$\eta_{km}(n) = f_k(n - 2^{k+1}m) \quad 0 \leq k \leq M-2 \quad (\text{پ-۱-۴۱})$$

$$\eta_{M-1,m}(n) = f_{M-1}(n - 2^{M-1}m)$$

معکوس تبدیل موجک زمان گسسته بصورت زیر خواهد شد:

$$x(n) = \sum_{k=0}^{M-1} \sum_{m=-\infty}^{+\infty} X_{DTWT}(k, m) \eta_{km}(n) \quad (\text{پ-۱-۴۲})$$

همچنان که قبلاً اشاره شد، ما علاقه‌مند به پایه‌های متعامد یک‌هستیم، که با فرض متعامد یک‌ه بودن $\{\eta_{km}(n)\}$ ، از رابطه‌ی (پ-۱-۴۲) نتیجه می‌شود که:

$$X_{DTWT}(k, m) = \sum_{n=-\infty}^{+\infty} x(n) - \eta_{km}^*(n) \quad (\text{پ-۱-۴۳})$$

با مقایسه‌ی (پ-۱-۴۳) با (پ-۱-۴۰) و با توجه به رابطه‌ی (پ-۱-۴۲)، براحتی نتیجه می‌شود که:

$$f_k(n) = h_k^*(-n) \quad (\text{پ-۱-۴۴})$$

با این انتخاب، نرمالیزاسیون انرژی فیلترهای بانک آنالیز و سنتز انجام خواهد شد. در واقع این کار با انتخاب مناسب دامنه‌ی فیلترهای $H(z)$ و $G(z)$ امکان‌پذیر است.

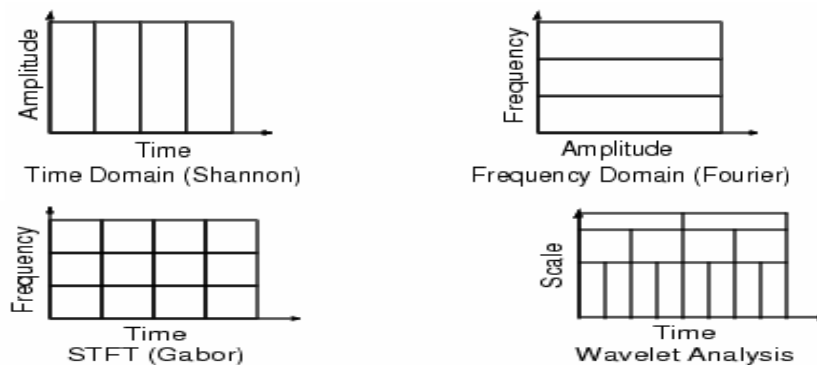
پ-۵- مقایسه رزولوشن زمانی و فرکانسی تبدیل موجک

جا دارد که مشابه تبدیل فوریه زمان کوتاه، رزولوشن‌های زمانی و فرکانسی را در تبدیل موجک مورد بررسی قرار دهیم. با بررسی تبدیل موجک سیگنالی متشکل از دو ضربه در صفحه‌ی زمان - مقیاس ملاحظه می‌شود که در مقیاس‌های پایین (فرکانس‌های بالا)، رزولوشن زمانی بسیار زیاد است به گونه‌ای که پالس‌ها، با پهنایی تقریباً برابر با مقدار واقعی آن‌ها در حوزه‌ی زمان و کاملاً تفکیک‌شده نمایش داده شده‌اند. از طرفی هرچه به مقیاس‌های بالاتر (فرکانس‌های پایین‌تر) نزدیک می‌شویم، دقت زمانی تبدیل کاهش یافته و پالس‌ها با گستردگی زیادی در صفحه‌ی زمان - مقیاس نمایش داده می‌شوند به گونه‌ای که برای مقیاس‌های خیلی بالا، با هم تداخل نیز پیدا می‌کنند. شکل پ-۱-۴ بیانگر تغییرات دقت‌های زمانی و فرکانسی، در حوزه‌ی زمان فرکانس و مقایسه تبدیل موجک با تبدیل‌های دیگر است.

(الف)



(ب)



شکل (پ-۱-۴) (الف) نمایش تغییرات زمانی و فرکانسی در تبدیل موجک (ب) مقایسه تبدیل‌های متفاوت روی سیگنال

پ-۲-۱-مثال: فرض می‌کنیم الگوهای ما تشکیل شده از تعداد سه مداد و تعداد سه میخ. و در مورد هر

یک از این الگوها، طول و وزن آن را می‌دانیم $I_r = \begin{bmatrix} \text{طول} & \text{الگو} \\ \text{وزن} & \text{الگو} \end{bmatrix}$

$$I_1 = \begin{bmatrix} 10 \\ 80 \end{bmatrix} \quad a_1 \rightarrow \text{میخ}$$

$$I_2 = \begin{bmatrix} 25 \\ 25 \end{bmatrix} \quad a_2 \rightarrow \text{مداد}$$

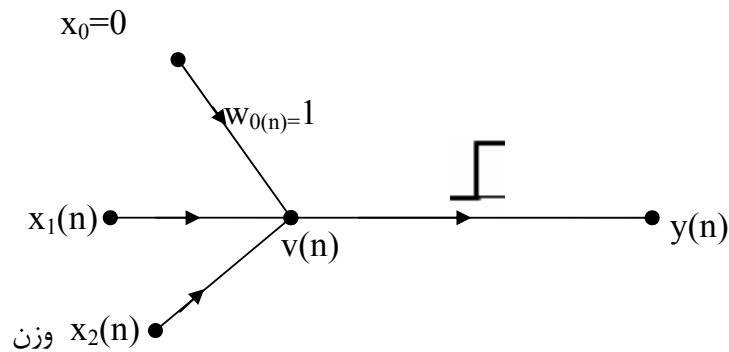
$$I_3 = \begin{bmatrix} 7 \\ 70 \end{bmatrix} \quad a_3 \rightarrow \text{میخ}$$

$$I_4 = \begin{bmatrix} 22 \\ 20 \end{bmatrix} \quad a_4 \rightarrow \text{مداد}$$

$$I_5 = \begin{bmatrix} 5 \\ 75 \end{bmatrix} \quad a_5 \rightarrow \text{میخ}$$

$$I_6 = \begin{bmatrix} 30 \\ 22 \end{bmatrix} \quad a_6 \rightarrow \text{مداد}$$

مدل نرون زیر را در نظر می‌گیریم.



شکل (پ-۲-۱) مدل نرون به کار رفته در مثال

از آنجا که در این روش یک سری عملیات را تکرار می‌کنیم، زمان را نیز وارد مدل کرده ایم به این شکل که برای مثال $y(n)$ بیانگر خروجی در تکرار n ام می‌باشد. فرض می‌کنیم

$$y(n) = \begin{cases} 1 & v(n) \geq 0 \\ 0 & v(n) < 0 \end{cases} \quad (\text{پ-۲-۱})$$

در اینجا برای آموزش شبکه باید تابع خطایی را تعریف کنیم تا به کمک آن بتوانیم اختلاف خروجی شبکه با خروجی مطلوب را اندازه گیری کنیم و ضرائب وزنی را در جهت کاهش این خطا تغییر دهیم. تابع خطا را به صورت زیر تعریف می کنیم.

$$e(n) = d(n) - y(n) \quad (\text{پ-۲-۲})$$

که در آن $d(n)$ خروجی مطلوب است، در نظر می گیریم. و $d(n)$ را به صورت ۱ برای میخ و ۰ برای مداد در نظر می گیریم. تأثیر خطا در تکرار n ام برای اصلاح ضرائب وزنی در تکرار $(n+1)$ را به این شکل تعیین می کنیم:

$$w_i(n+1) = w_i(n) + \Delta_i(n) \quad (\text{پ-۲-۳})$$

که در آن $\Delta_i(n) = -\eta e(n) x_i(n)$ می باشد و η ضریب ثابت مثبت است. به این ترتیب اگر $e(n) = 0$ باشد، تغییری در پارامترهای آزاد تکرار بعدی ایجاد نمی شود. و شرط اجرای ۵ دوره (هر بار که همه ورودی ها را به شبکه اعمال کنیم می گوئیم یک دوره کامل شد است) و یا رسیدن به خطای صفر ($e(n) = 0$) در طول یک دوره کامل را برای پایان عملیات، تعیین می کنیم. حال از $\eta = 0.1$ در نظر می گیریم و به طور اتفاقی پارامترهای آزاد را تعیین می کنیم:

$$w_1 = -1, w_2 = 1$$

تکرار اول

$$\begin{aligned} x_1(1) &= 10 & w_1(1) &= 1 & d(1) &= 1 \\ x_2(1) &= 80 & w_2(1) &= -1 \\ V(1) &= 1 \times (10) - 1 \times (80) < 0 \Rightarrow y(1) = 0 & e(1) &= 1 - 0 = +1 \\ \Delta_1(1) &= 1 \times (0.1) \times 10 = 1 & \Delta_2(1) &= 1 \times (0.1) \times 80 = 8 \end{aligned}$$

تکرار دوم

$$\begin{aligned} x_1(2) &= 25 & w_1(2) &= 1 + 1 = 2 & d(2) &= 0 \\ x_2(2) &= 25 & w_2(2) &= -1 + 8 = 7 \\ V(2) &= 2(25) + 7(25) > 0 \Rightarrow y(2) = 1 & e(2) &= 0 - 1 = -1 \\ \Delta_1(2) &= (-1) \times (0.1) \times (25) = -2/5 & \Delta_2(2) &= (-1)(0.1)(25) = -2.5 \end{aligned}$$

تکرار سوم

$$\begin{aligned} x_1(3) &= 7 & w_1(3) &= 2 - 2.5 = -0.5 & d(3) &= 1 \\ x_2(3) &= 70 & w_2(3) &= 7 - 2.5 = 4.5 \\ V(3) &= (-0.5)(7) + (4.5)(70) > 0 \Rightarrow y(3) = 1 & e(3) &= 0 \\ \Delta_1(3) &= (0) \times (0.1) \times (7) = 0 & \Delta_2(3) &= (0)(0.1)(70) = 0 \end{aligned}$$

تکرار چهارم

$$\begin{aligned} x_1(4) &= 22 & w_1(4) &= -0.5 & d(4) &= 0 \\ x_2(4) &= 20 & w_2(4) &= 4.5 \end{aligned}$$

$$V(4) = (-0.5)(22) + 20(4.5) > 0 \Rightarrow y(4) = 1 \quad e(4) = -1$$

$$\Delta_1(4) = (-1) \times (0.1) \times (22) = -2.2 \quad \Delta_2(4) = (-1)(0.1)(20) = -2$$

تکرار پنجم

$$x_1(5) = 5 \quad w_1(5) = -2.7 \quad d(5) = 1$$

$$x_2(5) = 75 \quad w_2(5) = 2.5$$

$$V(5) = (-2.7)(5) + (75)(2.5) > 0 \Rightarrow y(5) = 1 \quad e(5) = 0$$

$$\Delta_1(5) = \Delta_2(5) = 0$$

تکرار ششم

$$x_1(6) = 30 \quad w_1(6) = -2.7 \quad d(6) = 0$$

$$x_2(6) = 22 \quad w_2(6) = 2.5$$

$$V(6) = -2.7 \times 30 + 2.5 \times 22 < 0 \Rightarrow y(6) = 0 \Rightarrow e(6) = 0$$

$$\Delta_1(6) = \Delta_2(6) = 0$$

در اینجا دوره اول به پایان رسید. حال دوره دوم را آغاز می‌کنیم.

تکرار هفتم

$$x_1(7) = 10 \quad w_1(7) = -2.7 \quad d(7) = 1$$

$$x_2(7) = 80 \quad w_2(7) = 2.5$$

$$V(7) = -2.7 \times 10 + 2.5 \times 80 > 0 \Rightarrow y(7) = 1 \Rightarrow e(7) = 0$$

$$\Delta_1(7) = \Delta_2(7) = 0$$

به همین ترتیب تا پایان دوره دوم یعنی تا پایان تکرار دوازدهم ادامه می‌دهیم و در صورت انجام این کار مشاهده می‌شود خطاهای مربوط به دوره دوم برابر صفر است پس شبکه همگرا شده است. بنابراین

$$W_1 = -2/7 \quad \text{و} \quad W_2 = 2/5$$

حال الگوی جدید را به شبکه ارائه می‌دهیم و فرض می‌کنیم این الگو یا میخ است و یا مداد. بردار

مشخصات این الگو را به صورت $\begin{bmatrix} 8 \\ 75 \end{bmatrix}$ در نظر می‌گیریم. حدس می‌زنیم این الگو یک میخ باشد (با توجه

به بردار مشخصات آن). می‌خواهیم ببینیم آیا شبکه هم، همین حدس را می‌زند.

$$x_1 = 8 \quad w_1 = -2.7$$

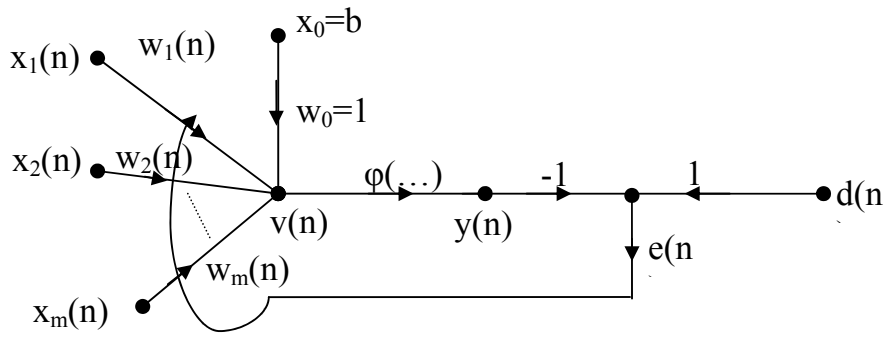
$$x_2 = 75 \quad w_2 = 2.5$$

$$V = 8 \times (-2.7) + 75 \times 2.5 > 0 \Rightarrow y = 1 \Rightarrow$$

الگوی یک میخ است.

روش فوق مبتنی بر آموزش از طریق تصحیح خطاست. زیرا در هر تکرار سعی کردیم پارامترهای آزاد را در جهتی تغییر دهیم که خطا کاهش پیدا کند. مثلاً در مثال فوق، در جایی که خروجی مطلوب یک بود و خروجی شبکه صفر، ضرائب را افزایش دادیم. که این افزایش ضرائب وزنی در راستای زیاد شدن $V(i)$ و نهایتاً یک شدن جواب خروجی است.

روش مطرح شده را که مبتنی بر آموزش از طریق تصحیح خطاست می‌توان به کمک شکل پ-۲-۲ نمایش داد.



شکل (پ-۲-۲) نمایش روش مطرح شده در مثال

در مثال قبل فضای مشخصات را به دو محدوده تفکیک کردیم. حال می‌خواهیم ببینیم، تابع ممیز به چه شکل است در این مثال شبکه به این شکل عمل می‌کند که $v = w_1x_1 + w_2x_2$ را در نظر می‌گیرد و اگر $v < 0$ در خروجی یک محدوده و اگر $v \geq 0$ ، در خروجی محدوده دیگر را مشخص می‌کند. بنابراین $v=0$ را به عنوان معیار انتخاب می‌کنیم

$$v = w_1x_1 + w_2x_2 = 0 \quad -2.7x_1 + 2.5x_2 = 0$$

$$\Rightarrow x_2 = \frac{2.7}{2.5} x_1$$

که نشان دهنده یک خط راست است. که سمت چپ این خط محدوده مدادها و سمت راست (و روی آن) محدوده میخ هاست. هر الگوی جدیدی را هم که به ورودی شبکه ارائه دهیم، شبکه بردار مشخصات آن را با خط فوق می‌سنجد و اگر تصویر الگو در فضای مشخصات در سمت چپ این خط قرار گیرد، شبکه آن را به عنوان مداد و اگر در سمت راست یا روی آن قرار گیرد به عنوان میخ، معرفی می‌کند کارایی روشی که در این مثال به کار بردیم محدود است.

$$x_2 = \frac{w_1}{w_2} x_1 \quad (\text{پ-۲-۴})$$

اگر از هر الگو دو مشخصه در دست داشتیم (X_1 و X_2) و در شبکه هم دو پارامتر آزاد داشتیم یعنی تعداد مشخصه‌ها برابر با تعداد پارامترهای آزاد بود مشاهده کردیم که شبکه تنها یک خط راست را می‌تواند بر عنوان تفکیک کننده فضا در نظر بگیرد. پرسپترون می‌تواند طبقاتی را که در دو سوی خط قرار دارند را تفکیک کند لیکن حالت های فراوانی وجود دارد که جدایی طبقات بسیار پیچیده تر است. مثلاً مورد XOR که یک مسئله ساده به نظر می‌رسد ولی از آنجا که هیچ خط راستی پیدا نمی‌شود که بتواند دو گروه خروجی را از هم تفکیک کند، پرسپترون قادر به حل این مسئله نیست.

پیوست ۳-

پ-۳-۱- چند یادآوری، قرارداد و تعریف

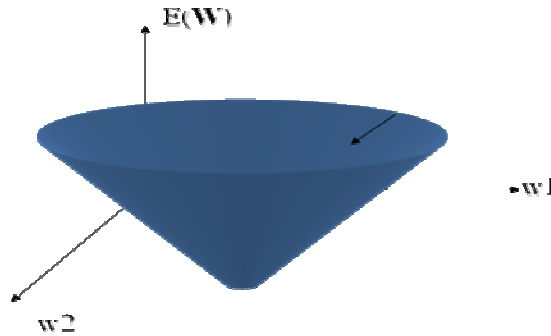
الف) بردارها را با حروف بزرگ و پررنگ و یا با علامت \rightarrow نشان می‌دهیم.

ب) گرادیان:

فرض می‌کنیم $y=f(X)$ است و گرادیان $f(X)$ را به صورت زیر تعریف می‌کنیم:

$$\overline{\nabla f(x)} = [\partial f / \partial x_1 \quad \partial f / \partial x_2 \quad \dots \quad \partial f / \partial x_m]^T \quad (\text{پ-۳-۱})$$

از ریاضیات می‌دانیم $\overline{\nabla f(X)}$ در نقطه X ، مطابق شکل پ-۳-۱ دارای جهتی است که اگر از نقطه X در آن جهت حرکت کنیم، بیشترین افزایش در $f(X)$ را خواهیم داشت. و $-\overline{\nabla f(X)}$ در نقطه X ، دارای جهتی است که اگر از نقطه X در آن جهت حرکت کنیم، بیشترین کاهش را در $f(X)$ را خواهیم داشت.



شکل (پ-۳-۱) نمایش $-\overline{\nabla f(X)}$

ج) تابع $f(t) = \frac{1}{1+e^{-t+\theta}}$ را تابع سیگموئیدی می‌نامیم و

$$f'(t) = \frac{e^{-t+\theta}}{(1+e^{-t+\theta})^2} = \frac{1}{1+e^{-t+\theta}} \left(1 - \frac{1}{1+e^{-t+\theta}} \right) \Rightarrow$$

$$f'(t) = f(t)(1-f(t)) \quad (\text{پ-۳-۲})$$

د) اگر $y = f(u_1, u_2, \dots, u_n)$ که در آن U_i تابعی از t است آنگاه

$$\frac{\partial y}{\partial t} = \frac{\partial f}{\partial u_1} \frac{\partial u_1}{\partial t} + \frac{\partial f}{\partial u_2} \frac{\partial u_2}{\partial t} + \dots + \frac{\partial f}{\partial u_n} \frac{\partial u_n}{\partial t} \quad (\text{پ-۳-۳})$$

هـ) از اینجا به بعد منظور از گره در لایه پنهان و خروجی، نرون و در لایه ورودی همان گره است.

و) از اینجا به بعد همواره در نظر می‌گیریم:

۱- مجموعه C بیانگر مجموعه تمام گره‌های لایه L ام است و اندیس j هم مختص لایه L ام است

۲- مجموعه g بیانگر مجموعه تمام گره های لایه $(L-1)$ ام است (در صورت وجود چنین لایه ای) و اندیس i هم مختص این لایه است.

۳- مجموعه h بیانگر مجموعه تمام گره های لایه $(L+1)$ ام است (در صورت وجود چنین لایه ای) و اندیس k هم مختص این لایه است.

(z) w_{ji} ضریب وزنی از گره i به گره j است.

($x_0=b=0$) ج از اینجا به بعد در هر نرون مقدار بایاس را برابر صفر در نظر می گیریم

حال به یافتن مناسب ترین اصلاحات برای پارامترهای آزاد می پردازیم.

اصلاح را به شکل زیر در نظر می گیریم و سعی می کنیم $\Delta_{ji}(n)$ مناسب را بیابیم

$$w_{ij}(n+1) = w_{ij}(n) + \Delta_{ji}(n) \quad (\text{پ-۳-۴})$$

مجموعه تمام ضرائب وزنی شبکه در تکرار n ام را به صورت بردار $\vec{W}_{ji}(n)$ نمایش می دهیم و می دانیم

خطا تابعی از این بردار است: $\varepsilon(n) = f(\vec{W}_{ji}(n))$. مجموعه تمام $\Delta_{ji}(n)$ ها را به صورت بردار $\vec{\Delta}_{ji}(n)$

نمایش می دهیم. حال به دنبال بهترین جهت برای این بردار می گردیم با توجه به مفهوم گرادیان مشخص

می شود بهترین جهت برای این بردار در جهت بردار $-\vec{\nabla} \varepsilon(n) = -\vec{\nabla} f(\vec{W}_{ji}(n))$ است. بنابراین

$$\vec{\Delta}_{ji}(n) = -\eta \vec{\nabla} f(\vec{W}_{ji}(n)) \quad (\text{پ-۳-۵})$$

که در آن η عدد ثابت و مثبت است که به آن نرخ یادگیری می گوئیم. پس

$$\Delta_{ji}(n) = -\eta \frac{\partial \varepsilon(n)}{\partial W_{ji}(n)} \quad (\text{پ-۳-۶})$$

پ-۳-۲-روش انتشار به عقب

در این قسمت هدف ما یافتن $\Delta_{ji}(n)$ ها با توجه به رابطه پ-۳-۶ است. تابع خطا در تکرار n ام را به

شکل زیر تعریف می کنیم.

$$\varepsilon(n) = \frac{1}{2} \sum_{b \in q} (d_b(n) - y_b(n))^2 \quad (\text{پ-۳-۷})$$

که در آن مجموعه q بیانگر مجموعه تمام گره های لایه آخر است. $y_b(n)$ خروجی واقعی گره b که یک

گره خروجی است در تکرار n ام و $d_b(n)$ خروجی مطلوب گره b در تکرار n ام است. مسئله را به دو

بخش تفکیک می کنیم:

۱- L لایه خروجی باشد.

۲- L لایه پنهان باشد.

(۱) L لایه خروجی باشد:

در این حالت

$$\varepsilon(n) = \frac{1}{2} \sum_{j \in c} (d_j(n) - y_j(n))^2 \quad (\text{پ-۳-۸})$$

$$\frac{\partial \varepsilon}{\partial W_{ji}(n)} = \frac{\partial \varepsilon(n)}{\partial V_j(n)} \frac{\partial V_j(n)}{\partial W_{ji}(n)} \quad (\text{پ-۳-۹})$$

و می دانیم:

$$V_j(n) = \sum_{i \in g} W_{ji}(n) y_i(n) \quad (\text{پ-۳-۱۰})$$

پس:

$$\frac{\partial V_j(n)}{\partial W_{ji}(n)} = y_i(n) \quad (\text{پ-۳-۱۱})$$

تعریف می کنیم:

$$\delta_j(n) = - \frac{\partial \varepsilon(n)}{\partial V_j(n)} \quad (\text{پ-۳-۱۲})$$

پس

$$\delta_j(n) = - \frac{\partial \varepsilon(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial V_j(n)} \quad (\text{پ-۳-۱۳})$$

می دانیم $y_j(n) = \varphi(V_j(n))$ پس

$$\frac{\partial y_j(n)}{\partial V_j(n)} = \varphi'(V_j(n)) \quad (\text{پ-۳-۱۴})$$

و

$$\frac{\partial \varepsilon(n)}{\partial y_j(n)} = -(d_j(n) - y_j(n)) \quad (\text{پ-۳-۱۵})$$

پس

$$\delta_j(n) = -(d_j(n) - y_j(n)) \varphi'(V_j(n)) \quad (\text{پ-۳-۱۶})$$

و اگر تابع φ را یک تابع سیگموئیدی در نظر بگیریم داریم:

$$\delta_j(n) = (d_j(n) - y_j(n)) y_j(n) (1 - y_j(n)) \quad (\text{پ-۳-۱۷})$$

پس داریم:

$$\Delta_{ji}(n) = +\eta \delta_j(n) y_i(n) \quad (\text{پ-۳-۱۸})$$

$$\Delta_{ji}(n) = +\eta y_i(n) y_j(n) (d_j(n) - y_j(n)) (1 - y_j(n)) \quad (\text{پ-۳-۱۹})$$

(۲) L لایه پنهان باشد.

$$\frac{\partial \varepsilon(n)}{\partial W_{ji}(n)} = \frac{\partial \varepsilon(n)}{\partial V_j(n)} \frac{\partial V_j(n)}{\partial W_{ji}(n)}$$

$$\Rightarrow -\frac{\partial \varepsilon(\mathbf{n})}{\partial V_j(\mathbf{n})} = \delta_j(\mathbf{n}), \frac{\partial V_j(\mathbf{n})}{\partial W_{ji}(\mathbf{n})} = y_i(\mathbf{n}) \quad (\text{پ-۳-۲۰})$$

پس:

$$\Delta_{ji}(\mathbf{n}) = \eta \delta_j(\mathbf{n}) y_i(\mathbf{n}) \quad (\text{پ-۳-۲۱})$$

$$\frac{\partial \varepsilon(\mathbf{n})}{\partial V_j(\mathbf{n})} = \frac{\partial \varepsilon(\mathbf{n})}{\partial y_j(\mathbf{n})} \cdot \frac{\partial y_j(\mathbf{n})}{\partial V_j(\mathbf{n})} \Rightarrow \delta_j(\mathbf{n}) = -\frac{\partial \varepsilon(\mathbf{n})}{\partial y_j(\mathbf{n})} \cdot \phi'(V_j(\mathbf{n})) \quad (\text{پ-۳-۲۲})$$

بافرض اینکه ϕ تابع سیگموئیدی است داریم:

$$\delta_j(\mathbf{n}) = -\frac{\partial \varepsilon(\mathbf{n})}{\partial y_j(\mathbf{n})} y_j(\mathbf{n})(1 - y_j(\mathbf{n})) \quad (\text{پ-۳-۲۳})$$

می دانیم $\varepsilon(x)$ تابعی از تک تک $V_k(\mathbf{n})$ هاست و هر $V_k(\mathbf{n})$ تابعی از $y_j(\mathbf{n})$ است. پس:

$$\frac{\partial \varepsilon(\mathbf{n})}{\partial y_j(\mathbf{n})} = \sum_{k \in h} \frac{\partial \varepsilon(\mathbf{n})}{\partial v_k(\mathbf{n})} \frac{\partial v_k(\mathbf{n})}{\partial y_j(\mathbf{n})} \quad (\text{پ-۳-۲۴})$$

و

$$\frac{\partial v_k(\mathbf{n})}{\partial y_j(\mathbf{n})} = \frac{\partial}{\partial y_j(\mathbf{n})} \left[\sum_{J \in C} W_{kj}(\mathbf{n}) y_j(\mathbf{n}) \right] = W_{kj}(\mathbf{n}) \quad (\text{پ-۳-۲۵})$$

$$\Rightarrow \frac{\partial \varepsilon(\mathbf{n})}{\partial y_j(\mathbf{n})} = \sum_{k \in h} \frac{\partial \varepsilon(\mathbf{n})}{\partial v_k(\mathbf{n})} W_{kj}(\mathbf{n}) \quad (\text{پ-۳-۲۶})$$

تعریف:

$$\delta_k(\mathbf{n}) = -\frac{\partial \varepsilon(\mathbf{n})}{\partial v_k(\mathbf{n})} \quad (\text{پ-۳-۲۷})$$

$$\Rightarrow \frac{\partial \varepsilon(\mathbf{n})}{\partial y_j(\mathbf{n})} = \sum_{k \in h} -\delta_k(\mathbf{n}) W_{kj}(\mathbf{n}) \quad (\text{پ-۳-۲۸})$$

با توجه به (پ-۳-۲۳) می توانیم بنویسیم:

$$\Rightarrow \delta_j(\mathbf{n}) = \left(\sum_{k \in h} \delta_k(\mathbf{n}) W_{kj}(\mathbf{n}) \right) y_j(\mathbf{n})(1 - y_j(\mathbf{n})) \quad (\text{پ-۳-۲۹})$$

L لایه پنهان است و طبق رابطه (پ-۳-۲۳) و (پ-۳-۲۹) داریم:

$$\Rightarrow \Delta_{ji} = \eta y_i(\mathbf{n}) y_j(\mathbf{n}) \left(\sum_{k \in h} \delta_k(\mathbf{n}) W_{kj}(\mathbf{n}) \right) (1 - y_i(\mathbf{n})) \quad (\text{پ-۳-۳۰})$$

دو حالت فوق را به این شکل خلاصه می کنیم

$$\Delta_{ji}(\mathbf{n}) = \eta \delta_j(\mathbf{n}) y_i(\mathbf{n}) \quad (\text{پ-۳-۳۱})$$

که در آن L لایه خروجی است.

$$\delta_j(\mathbf{n}) = (d_j(\mathbf{n}) - y_j(\mathbf{n})) y_j(\mathbf{n})(1 - y_j(\mathbf{n})) \quad (\text{پ-۳-۳۲})$$

پس اگر L لایه پنهان باشد:
(پ-۳-۳۳)

$$\delta_j(n) = \left(\sum_{k \in h} \delta_k(n) W_{kj}(n) \right) y_j(n) (1 - y_j(n))$$



Shahrood University of Technology
Robotic and Electrical Engineering Faculty

Electronic Engineering Master Period Thesis
Sadness and Happiness Expression Recognition Using
Two Dimensional Grey Scale Facial Images

Nooshin Nabizadeh Ardakani

Supervisor Professor
Dr. Ali Soleimani Eivari
Consultor Professor
Dr. Alireza Ahmadifard

June 2009