
بسم الله الرحمن الرحيم

دانشکده فنی و مهندسی

ردیابی لب و لب‌خوانی با استفاده از پردازش تصاویر ویدیوئی

استاد راهنما : دکتر علیرضا بهراد

نگارش : وحید عزتی چهار قلعه

تابستان 88

تقدیم به:

- پدر و مادر مهربانم که بی هیچ چشمداشتی، عاشقانه به پایمان می سوزند و لبخند می زنند و در هر شرایطی کنارمان هستند و فراموشمان نمی‌کنند.
- برادر و خواهرهای عزیزم که همیشه مشوق و کمک حال من در این عرصه بوده اند.

تشکر و قدردانی

برخود لازم می دانم که از استاد بزرگوارم جناب آقای دکتر علیرضا بهراد که علاوه بر راهنمایی این پایان نامه، نکات اخلاقی و علمی فراوان به من آموختند تشکر و قدردانی نمایم.

همچنین از سایر اساتید گرانقدر بویژه آقایان دکتر محمد میکائیلی ، دکتر غزنوی قوشچی و که افتخار شاگردی آنها را در طول دوران تحصیل داشتم، تشکر می نمایم.

از دوستان عزیزم در دانشگاه شاهد بویژه نیما حاتمی، علیرضا بساق زاده، پیمان کاوه، احمد نجومی، هادی حسنی، بهزاد داغستانی وسایر دوستان که مرا در امر جمع آوری داده همراهی و همیاری کردن صمیمانه تشکر می کنم.

از عموی عزیز و خانواده صمیمیشان که در طول این سالها مرا به عنوان عضوی از خانواده شان پذیرا شدند و زحمات زیادی را طی این سالها برای من کشیدند نهایت سپاس را دارم.

چکیده

لب‌خوانی از سالیان پیش یکی از موضوعات و ابزارهای مهم برای افراد کم شنوا و ناشنوا بوده تا این افراد درک مناسبی نسبت به گفته‌های شخصی که در حال صحبت کردن است داشته باشند.

اخیر اً لب‌خوانی با استفاده از تصاویر ویدیویی (تصاویر متوالی) یکی از موضوعات مورد علاقه محققان بوده که طی چند دهه اخیر تحقیقات گسترده‌ای راجع به این مساله انجام داده و مقاله‌های متعددی در این باره چاپ نموده‌اند، چرا که استفاده از تصاویر ویدیویی از حرکات لب و دهان و اطلاعات حاصل از آن در شناسایی و تشخیص گفتار تحت شرایط صوتی نابهنجار و نویزی کمک موثری به شخص می‌کند. درحالت کلی گرچه نرخ شناسایی و تشخیص گفتار، با سیستم های لب‌خوانی پایین است ولی در چنین محیطهایی استفاده از اطلاعات تصویری به مراتب بهتر از اطلاعات صوتی می‌باشد. و بدین دلیل تلاشهای فراوانی برای بهبود عملکرد چنین سیستمهایی صورت گرفته است. هدف این پایان نامه معرفی یک سیستم لب‌خوانی میتنی بر پردازش تصویر برای کلمات فارسی می‌باشد.

مراحل اصلی یک سیستم لب‌خوانی بصورت زیر می‌باشد:

1- بدست آوردن ناحیه لب از هر فریم ویدیویی

2- استخراج ویژگی‌های مهم از ناحیه لب

3- شناسایی کلمات بیان شده توسط هر گوینده با استفاده از پردازش زمانی ویژگی ها

از آنجا که مهمترین بخش یک سیستم لب‌خوانی بدست آوردن ویژگی‌های مناسب برای تشخیص گفتار است و این امر جز با استخراج مناسب لب از ناحیه چهره فرد میسر نخواهد شد بنابراین ما در این تحقیق از یک روش جدیدی برای جداسازی ناحیه لب از ناحیه پوست صورت شخص استفاده کرده‌ایم. در روش ارایه شده برای لب‌خوانی در این تحقیق ابتدا روشی برای جداسازی بهتر ناحیه لب از ناحیه پوست ارایه می‌شود سپس تصویر

بدست آمده به عنوان بردار ویژگی به الگوریتم فازی جهت خوشه‌بندی صورت به دو ناحیه لب و پوست داده می‌شود. سپس با استفاده از یک آستانه‌گیری تطبیقی ناحیه لب را جدا ساخته و برای بدست آوردن پیرامون لب، مرز ناحیه لب را بدست می‌آوریم. اما از آنجا که مرز بدست آمده بدرستی بر روی مرز لب تصویر اصلی تطبیق نمی‌شود از مدل پیرامون فعال جهت حل این مشکل استفاده می کنیم. در مرحله بعد، از چندین ویژگی مانند ویژگی‌های هندسی لب به عنوان بردار ویژگی استفاده می‌کنیم و به عنوان ورودی به یکی از روشهای طبقه‌بندی از جمله شبکه عصبی داده و در نهایت شناسایی لازم انجام خواهد گرفت.

فهرست مطالب

فهرست شکلها

لیست جدول ها

فصل **1**

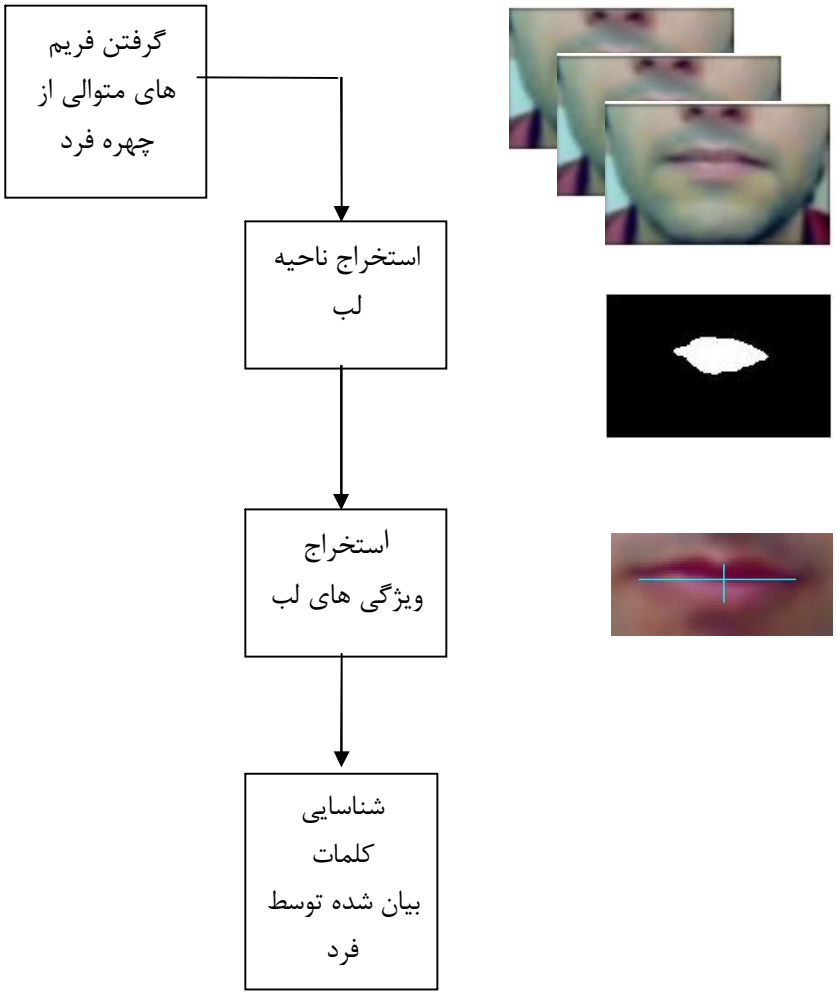
مقدمه

اخیرا سیستمهای اتوماتیک شناسایی بصری کلمات بیان شده توسط یک فرد یا همان سیستم لب‌خوانی، مورد تحقیق و بررسی فراوانی قرار گرفته‌اند. این سیستم‌ها نقش به سزایی در سیستم‌های چندرسانه‌ای[1] از جمله سیستم شناسایی گفتار- تصویری[2]، سیستم‌های موبایل، سیستم‌های ارتباطی بین انسان و کامپیوتر(HCI[3]) دارند. همچنین از یک سیستم لب‌خوانی می‌توان برای بهبود کاربردهای تشخیص افراد، کنترل ماشین و همچنین در بازی‌های پویانمایی استفاده کرد.

مستند است که در قرن 17 (1648م) گزارشی مبنی بر دارا بودن اطلاعات مفید گفتاری در حرکات صورت گوینده وجود دارد. اما پایه‌ای‌ترین تحقیق در این زمینه در اوایل دهه 80 (1976) صورت گرفته است[1]. از همین سالها به بعد تلاشهای بسیار زیادی برای ارایه روشهای دقیق و سریع چه بصورت سخت‌افزاری(بسیار کم)[2]و[3]، و چه بصورت نرم‌افزاری صورت پذیرفته است. بیشتر این تحقیقات روی لب‌خوانی از طریق دید از جلو (دوربین کاملا در جلو شخص قرار داشته و چهره شخص کاملا رو به دوربین است) صورت گرفته است. اما اخیر اً تحقیقاتی روی لب‌خوانی با استفاده از دید از نیم رخ صورت پذیرفته است که تا حدودی نرخ تشخیص کلمات را بهبود بخشیده است[4]،[5]و[6].

هر سیستم لب خوانی دارای مراحلی از جمله ناحیه‌بندی لب[4] جهت استخراج ناحیه لب از زمینه صورت، استخراج ویژگی‌های لب از ناحیه لب از تصاویر متوالی ویدیوئی و در نهایت شناسایی کلمات از طریق ویژگی‌های استخراج شده است. در شکل(1-1) شمایی از یک سیستم لب‌خوانی استاندارد نمایش داده شده است .

[1] Multimedia
[2] Visual speech
[3] Human Computer Interaction
[4] Lip segmentation



شکل1- 1: مراحل یک سیستم لب‌خوانی ساده

برای ناحیه بندی لب روشهای متفاوتی استفاده شده است. الگوریتم‌های متفاوتی از جمله مدل پیرامون فعال[1] و نیز مدلهای بهبود یافته آن طی چند سال اخیر مورد استفاده قرار گرفته‌اند. عیبهای این روش این است که در کمینه[2]های محلی گیر می‌کنند و همچنین سرعت همگرایی آنها به سمت مرز شیء مورد نظر

[1] Active contour model
[2] Minimum

پایین است. همچنین مدل دیگری بر اساس شکل لب معروف به مدل شکل فعال[1] ارایه شده است که این مدل از چندین نقطه روی لب جهت تقسیم‌بندی شکل لب و همچنین اطلاعات آماری جهت تشکیل یک مدل از لب استفاده می کند. مشکل اصلی این روش نیاز آن به داده‌های آموزشی بسیار زیاد می‌باشد. روشهای مبتنی بر اطلاعات رنگی و سطوح خاکستری جهت جداسازی ناحیه لب از صورت نیز مورد بررسی فراوانی قرار گرفته‌اند. از آنجا که فضای اولیه هر تصویر رنگی که با چشم انسان قابل تطبیق است فضای RGB می‌باشد در ابتدا از آن برای جداسازی لب از پوست در صورت استفاده می‌شد ولی نتایج نشان داده است که این فضا نسبت به تغییرات شدت روشنایی حساس بوده و برای همین به فضاهای یکنواخت[2] تبدیل می‌شود. از جمله این فضاها می‌توان به فضاهای رنگی، HSV,CIELAB,CIELUV,... اشاره کرد که در متن تحقیق توضیح داده خواهند شد.

بعد از بدست آوردن ناحیه لب، مرحله بعدی استخراج ویژگی‌های مناسب از آن است. در سیستم‌های لب‌خوانی چهار نوع ویژگی در نظر گرفته می‌شود: ویژگی‌های مبتنی بر تصویر که همه پیکسلهای لب در نظر گرفته می‌شوند و بنابراین هیچگونه اطلاعاتی از بین نمی‌رود ولی با این حال حجم محاسباتی زیادی داشته و کمتر مورد استفاده قرار می‌گیرد. ویژگی دیگر ویژگی‌های مربوط به شکل هندسی لب می‌باشند که از جمله می توان به ارتفاع لب و پهنای آن اشاره کرد. سومین ویژگی، معروف به ویژگی‌های ناحیه‌ای هستند که از پر کاربردترین آنها می‌توان به تحلیل مولفه‌های ویژه[3] (PCA) و همچنین تحلیل مولفه‌های مستقل[4] اشاره کرد(ICA). از این روشها برای کاهش بعد با استفاده از بردارها و مقادیر ویژه ماتریس کوواریانس مربوط به شکل مورد نظر(لب) استفاده می‌کنند. چهارمین ویژگی که از پر کاربردترین ویژگی‌ها در سیستم‌های

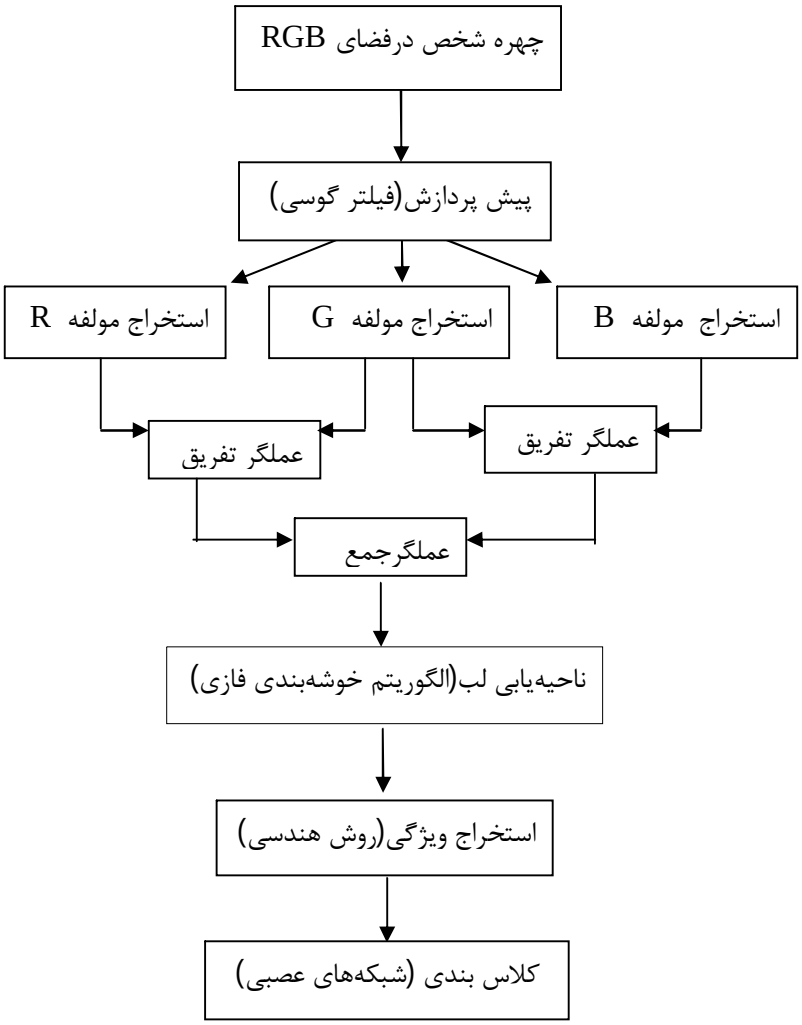
[1] Active shape model
[2] Uniform color spaces
[3] Principle component analysis
[4] Independent componenet analysis

لب‌خوانی است ویژگی‌های مبتنی بر مدل هستند. از جمله این مدلها می توان به مدلهای پیرامون فعال اشاره کرد.

بعد از اینکه ویژگی‌های مناسبی از طریق ناحیه لب استخراج شد، شناسایی کلمات بیان شده با استفاده از ویژگی‌های بدست آمده، انجام می‌شود. برای این کار طبقه‌بندهای[1] مختلفی ارایه شده‌اند که از جمله می توان به شبکه های عصبی[2]، SVM [3]، مدلهای مخفی مارکوف (HMM)[4] اشاره کرد. اما از بین این طبقه‌بندها تحقیقات نشان داده است که مدلهای مخفی مارکوف در بیشتر موارد مورد استفاده قرارگرفته است.

همانطور که بیان شد یکی از اصلی‌ترین مراحل یک سیتم لب‌خوانی استخراج ناحیه لب و بدست آوردن ویژگی‌های مناسب می‌باشد. در بیشتر تحقیقات که از اطلاعات رنگی استفاده شده است، بیشتر از فضاهای رنگی غیر از فضای رنگی RGB استفاده شده است که نیاز به تبدیلات رنگی و محاسبات اضافی دارند. که این بدلیل حساس بودن این فضا به تغییرات شدت روشنایی می‌باشد. اما در این تحقیق سعی شده‌است از فضای رنگی RGB و یک روش جدید(الگوریتم خوشه‌بندی فازی) برای استخراج لب استفاده شود و مشکل حساسیت به تغییرات شدت روشنایی تاحدودی از بین رود. و نیز از چند ویژگی جهت مقایسه روشها برای یافتن مناسب‌ترین ویژگی، استفاده خواهد شد. و در نهایت از طبقه‌بند شبکه عصبی جهت شناسایی کلمات بیان شده استفاده می‌شود.در شکل 1-2 بلوک دیاگرام این سیستم آورده شده است.

[1] classifier
[2] Neural Networks
[3] Support Vector Machin
[4] Hidden Marcov Model



شکل1- 2 : بلوک دیاگرام سیستم پیشنهادی

ساختار این پایان‌نامه به شرح زیر است. به دنبال این فصل در فصل دوم مروری بر روشهای ارایه شده توسط گروههای مختلف محققین خواهیم داشت. فصل سوم شامل الگوریتم پیشنهادی برای ناحیه‌بندی لب و استخراج ویژگی از آن می‌باشد. در فصل چهارم نتایج حاصل از الگوریتم پیشنهادی و شناسایی کلمات بر اساس چند روش طبقه‌بندی مورد بررسی قرار خواهد گرفت. فصل پنجم نیز جمع‌بندی و نتیجه‌گیری لازم و نیز پیشنهادات لازم را در بر خواهد گرفت.

فصل **2**

مروری بر روشهای لب‌خوانی

1-2 مقدمه

همانطور که قبلا گفته شد شناسایی گفتار از طریق حرکات لب در دهه‌های اخیر، محققین زیادی را به تلاش انداخته تا سیستم‌های بهتر با نرخ شناسایی بالاتر توسعه دهند [7] برای بدست آوردن نرخ شناسایی بالا بدست آوردن ویژگی‌های مناسب از ناحیه لب مورد نیاز است. بنابراین برای رسیدن به این هدف می‌بایست ناحیه لب را به درستی استخراج کرد. در واقع اگر بتوان ناحیه لب را بدرستی از ناحیه پوست صورت فرد تشخیص داد می‌توان ویژگی‌های بهتر و دقیق‌تری بدست آورد.

بعد از اینکه ناحیه لب به درستی از ناحیه پوست صورت تشخیص داده شد می‌توان آن را با یک سری پردازش‌ها از زمینه صورت جدا کرد. ناحیه لب بدست آمده دارای اطلاعات گفتار-تصویری خوبی است که دارای بردار ویژگی با بعد کمتر نیز می‌باشد. از این اطلاعات می‌توان در سیستم های دیگر بینایی ماشین برای تشخیص رفتار و حرکات انسان نیز استفاده کرد. برای مثال در تشخیص حالات چهره فرد و نیز در کنترل سیستمهای مربوط به اتومبیل(رادیو و ضبط) و یا در پویانمایی و کاربردهای کدینگ نیز مورد استفاده قرار می‌گیرند. از این ناحیه بدست آمده ویژگی‌های مناسب می‌بایست استخراج شوند که از جمله می‌توان به ویژگی‌های هندسی مانند ارتفاع و پهنای لب و یا مرز های لب و غیره اشاره کرد.

مرحله بعد طبقه‌بندی بردارهای ویژگی برای هر کلمه بیان شده است.

در این فصل سعی می‌کنیم روشهای مختلف ناحیه‌بندی لب، استخراج ویژگی‌های مناسب و طبقه‌بندی آنها را مورد بررسی قرار دهیم.