



## استخراج کانتور لب با استفاده از تصاویر رنگی

رضا شالباف

کارشناسی ارشد مهندسی پزشکی

دانشگاه صنعتی امیرکبیر

[rshbme@yahoo.com](mailto:rshbme@yahoo.com)

منصور وفادوست

عضو هیئت علمی دانشکده مهندسی

پزشکی دانشگاه صنعتی امیرکبیر

[vmansur@cic.aut.ac.ir](mailto:vmansur@cic.aut.ac.ir)

احمد شالباف

کارشناسی ارشد مهندسی پزشکی

دانشگاه صنعتی امیرکبیر

[a\\_shbme@yahoo.com](mailto:a_shbme@yahoo.com)

**چکیده:** بهره‌گیری از اطلاعات تصویری مانند شکلهای لب و حرکات لب می‌تواند دقت و مقاومت سیستم‌های تشخیص اتوماتیک گفتار را ارتقاء دهد. در این مقاله، یک الگوریتم جدید برای استخراج کانتور لب بر اساس ناحیه بندی تصاویر ارائه شده است. در این الگوریتم، با بهره‌گیری از یک مدل 16 نقطه‌ای برای توصیف کانتور لب، از شایستگی‌های روشهای مبتنی بر مدل‌های نقطه‌ای و مدل‌های پارامتری استفاده شده است. در ابتدا بوسیله روشهای خوشه بندی فازی، یک نگاشت احتمال از تصاویر رنگی بدست می‌آوریم سپس با استفاده از این نگاشت، یک تابع انرژی بر اساس ناحیه بندی تصاویر که احتمال اتصال ناحیه لب و غیر لب را ماکزیمم می‌کند، تعریف می‌شود و با فرآیند بهینه سازی نقطه ای تکرارپذیر، مدل لب را بر روی نگاشت احتمال تصویر تحت بررسی منطبق می‌کنیم. در این فرآیند با هر بار تکرار، تنظیم 16 نقطه لب بوسیله 2 منحنی درجه دوم و چهارم کنترل می‌شود. آزمایشات نشان می‌دهد که الگوریتم پیشنهاد شده دارای نتایج قابل قبولی بر روی تصاویر ساده لب در 20 فرد مختلف می‌باشد.

**واژه‌های کلیدی:** استخراج کانتور لب، مدل لب، استخراج و ویژگیهای گفتاری، تصاویر رنگی.

### 1) مقدمه

در سالهای اخیر علاقه زیادی به اطلاعات تصویری در سیستمهای بازشناسی گفتار ایجاد شده است و اطلاعات تصویری مانند حرکات لب می‌تواند به طور قابل ملاحظه‌ای دقت سیستم‌های تشخیص گفتار را خصوصاً در محیطهای نویزی بهبود دهد [1]. بنابراین یک الگوریتم قابل اطمینان و دقیق برای استخراج ویژگیهای تصویری مربوط به اطلاعات گفتار موردنیاز می‌باشد. به هر حال وجود تغییرات زیاد در تصاویر ناشی از گویندگان مختلف، لحن متنوع ادای کلمات توسط آنها و شرایط نوری، این امر را مشکل ساخته است. بعلاوه کنتراست کم شدت روشنایی و رنگ بین لب و پوست صورت، مشکل بزرگی برای آشکارسازی کانتور لب می‌باشد.

محققان زیادی روشهای مختلفی برای استخراج اطلاعات گفتار از دنباله تصاویر لب پیشنهاد کرده‌اند که این رویکردها به دو گروه اصلی دسته بندی می‌شوند: رویکردهای مبتنی بر

پیکسل [2] و رویکردهای مبتنی بر مدل [3-5]. در رویکردهای مبتنی بر پیکسل، ویژگیها به طور مستقیم از تصاویر داده شده استخراج می شوند [2]. هیچ اطلاعاتی از دست نمی رود و فرآیندهای ساده‌ای برای استخراج ویژگیهای لب وجود دارد؛ اما بردار ویژگی چند بعدی دارای اطلاعات تکراری می باشد و ضمناً چرخش، ابعاد و جابجایی لب در تصاویر در حین گفتار، بر عهده مرحله طبقه بندی گذاشته می شود.

در رویکردهای مبتنی بر مدل، یک مدل از لب ساخته می شود که این مدل با استفاده از مجموعه کوچکی از پارامترها توصیف می شود. امتیاز این نوع روش این است که ویژگیهای مهم در یک فضای با ابعاد کم ارائه شده است و نسبت به جابجایی، چرخش، ابعاد و شدت روشنایی تغییرناپذیر می باشند. دو مشکل اساسی در روشهای مبتنی بر مدل وجود دارد:

1- به چه طریق مدلی ساخته شود که شکل واقعی لب را توصیف کند.

2- به چه طریق تحت شرایط نوری متفاوت و در گویندگان مختلف به صورت دقیق و قابل اطمینان کانتور لب استخراج شود.

تعدادی از روشهای مبتنی بر مدل که برای مدلسازی کانتور لب و استخراج ویژگیهای مرتبط با گفتار پیشنهاد شده اند می توان به: روش مدل‌های مرز فعال (Active Contour Model) [3]، مدل‌های شکل فعال (Active Shape Model (ASM)) [4] و الگوهای انعطاف پذیر (Deformable Template) [5] اشاره کرد. مدل‌های مرز فعال، با یک مجموعه منحنی‌های به هم متصل ساخته می شوند که این منحنی‌ها تحت نیروهای خارجی و داخلی با مرز شکل مطابقت پیدا می کنند. مشکل این روشها، مقدار دهی اولیه مدل و سرعت پردازش پایین آن می باشد. به علاوه از آنجائیکه در این روش نمی توان محدودیتهایی در شکل اعمال کرد، ممکن است منجر به شکل های غیر عملی شود.

در مدل‌های شکل فعال از نقاط برای توصیف جزئیات شکل لب استفاده می شود و این نقاط توسط چند مدول تغییرات شکل بدست آمده از مجموعه داده آموزشی کنترل می شوند. مهمترین

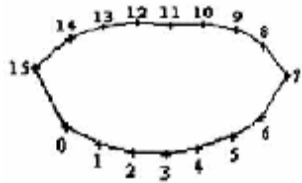
برتری استفاده از ASM این است که هیچ فرض جدیدی در مورد شکلهای واقعی صورت نمی گیرد؛ اما ساختن چنین مدلی نیاز به مجموعه داده آموزشی وسیعی دارد و فرآیند آموزش نسبتاً وقت گیر و خسته کننده می باشد و اگر تعداد نقاط مدل تغییر پیدا کند فرآیند آموزشی، یکبار دیگر باید انجام شود. در ضمن این روش، شکلهای متفاوت از مجموعه داده آموزشی را نمی تواند به خوبی تخمین بزند.

الگوهای انعطاف پذیر از مدل‌های پارامتری برای توصیف شکل فیزیکی تصویر استفاده می کنند و با محدودیتهای ابتکاری اعمال شده، تغییرات شکل را محدود می کنند و شکلهای استخراج شده مطمئناً طبیعی می باشند. به علاوه این روش، شکلهای واقعی را با استفاده از تعداد کمی از پارامترها توصیف می کند.

موضوع مهم در روشهای مبتنی بر مدل، فرمول سازی درست تابع انرژی می باشد که مدل لب را بر روی لب واقعی در تصویر منطبق کند. در بعضی از تابعهای انرژی به عنوان ویژگی اولیه از اطلاعات لبه شامل لبه ها، دره ها و پیک استفاده می شود اما اگر تفاوت قابل ملاحظه و کافی بین ناحیه لب و غیرلب نباشد، این رویکردها در پیدا کردن کانتور لب با مشکل مواجه می شوند. در این مقاله به منظور اینکه استخراج کانتور لب به عنوان مسئله قسمت بندی ناحیه به جای مسئله آشکارسازی لبه مطرح شود، از تابع انرژی بر اساس ناحیه بندی تصویر استفاده کرده ایم. با تابع انرژی مبتنی بر ناحیه، الگوریتم، برای تصاویر لب با کنتراست رنگی کم، مطمئن تر می باشد و نتایج نهایی، نسبت به تکنیکهای بر اساس لبه، حساسیت کمتری به مقداردهی اولیه پارامترهای مدل دارد.

در این مقاله، یک روش مبتنی بر مدل برای استخراج کانتور لب از تصاویر رنگی، پیشنهاد کرده ایم. دقت، قابلیت اطمینان و زمان پردازش، نگرانیهای اولیه در ارائه الگوریتم بوده است. مدل لب پیشنهادی، یک مدل مبتنی بر نقطه با محدودیتهای هندسی مشخص می باشد. امتیازات چنین مدلی در ادامه ذکر شده است:

1- از آنجائیکه ناحیه لب بوسیله یک مدل خطی تکه‌ای تعریف می شود محاسبات تابع انرژی نسبت به مدل‌های



شکل 1: مدل 16 نقطه ای لب

بوسیله  $I_p = \{X_{pi}, Y_{pi}\}, i=0,1,\dots,15$  می‌توانند توصیف شوند. این نقاط به 2 گروه تقسیم شده اند:  $P_0 - P_7$  و  $P_8 - P_{15}$  لب پایینی را توصیف می‌کنند و لب بالایی را توصیف می‌کنند.  $P_7$  و  $P_{15}$  گوشه‌های لب و  $P_{11}$  نقطه فرورفتگی می‌باشد. ضمناً به علت تغییرات زیاد ناشی از جابجایی، ابعاد و چرخش لب در تصاویر، فرمول بندی شکل‌های واقعی لب با استفاده از محور مختصات اصلی، مشکل و محاسبات ریاضی نسبتاً پیچیده ای نیاز دارد که برای ساده سازی این مسائل و غیر متأثر بودن مدل از چرخش، اندازه و جابجایی لب، از فرآیند نرمالیزاسیون توصیف شده در بخش (3-1) استفاده شده است.

### 1-3) نرمال سازی

برای کاهش تعداد پارامتر مدلهایی که بر روی نقاط کانتور لب تطبیق داده می‌شوند و غیر متأثر بودن مدل از چرخش، اندازه و جابجایی لب، از یک فرآیند نرمالیزاسیون استفاده شده است. ضمناً با این فرایند، نقاط گوشه لب ( $P_7$  و  $P_{15}$ )، روی محور  $X$  و نقطه  $P_{11}$ ، روی محور  $Y$  انتقال داده شده است. در ابتدا مرکز اصلی لب به طریقه زیر در وسط بین دو گوشه لب ( $P_7$  و  $P_{15}$ ) تنظیم می‌شود:

$$X_c = \frac{X_{p7} + X_{p15}}{2}, Y_c = \frac{Y_{p7} + Y_{p15}}{2} \quad (1)$$

شعاع نرمال سازی و زاویه چرخش بوسیله معادلات زیر بدست می‌آیند:

$$r = \sqrt{(X_{p7} - X_c)^2 + (Y_{p7} - Y_c)^2} \quad (2)$$

$$q = \tan^{-1} \left( \frac{Y_{p15} - Y_{p7}}{X_{p15} - X_{p7}} \right)$$

مکان همه نقاط به طریقه زیر نرمال سازی می‌شود:

$$Y_{pin} = [(Y_{pi} - Y_c) \cos q - (X_{pi} - X_c) \sin q] / r \quad (3)$$

نتایج نرمال سازی در شکل (2) نشان داده شده است.

پارامتری ساده‌تر می‌باشد.

2- بهینه سازی نقاط معمولاً قابل اطمینان تر و موثرتر از بهینه سازی پارامترها می‌باشد.

3- مدل هندسی لب که از 2 منحنی تشکیل شده است به عنوان یک محدودیت شکلی معتبر برای لب می‌باشد و بنابراین دیگر نیازی به آموزش داده‌ها برای ساختن لب که برای ASM مورد نیاز است، نمی‌باشد.

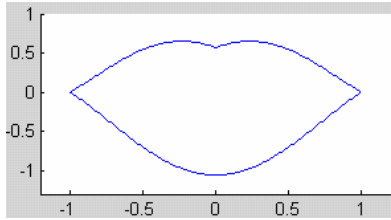
نتایج عملی نشان می‌دهد که الگوریتم ما می‌تواند شکل درست لب را در حالت‌های مختلف بدست آورد. مجموعه دادگان جمع آوری شده در قسمت 2 توصیف شده است. جزئیات مدل لب و محدودیتهای هندسی در قسمت 3 توصیف شده‌اند. در قسمت 4 فرآیند الگوریتم استخراج کانتور لب ارائه شده است. قسمت 5 نتایج عملی که بوسیله الگوریتم ما بدست آمده است را نشان می‌دهد.

### 2) مجموعه دادگان جمع آوری شده

مجموعه دادگان جمع آوری شده، شامل 20 نفر گوینده از گروه سنی 20 تا 50 سال، شامل 16 مرد و 4 زن می‌باشد که دارای شکل لب و رنگ پوست متفاوتی بوده و 5 نفر از گویندگان دارای ریش و سبیل نیز می‌باشند. ضمناً گویندگان به هیچ وجه آرایش نکرده‌اند و هیچ علامت مشخصی بر روی لب آنها قرار داده نشده است. هر یک از گویندگان هر کدام از شش کلمه فارسی یک، دو، سه، چهار، پنج و شش را 10 مرتبه ادا کرده‌اند. دادگان جمع آوری شده از لحاظ شرایط نوری حین تصویربرداری از تنوع مناسبی برخوردار می‌باشد. ضمناً دقت تصاویر گرد آوری شده  $100 \times 120$  می‌باشد و صورت گویندگان مختلف در حین ادای کلمات دارای چرخش و جابجایی لب می‌باشد.

### 3) مدل لب

در این الگوریتم، از یک مدل 16 نقطه ای برای توصیف کانتور لب استفاده شده و نقاط مرز لب  $P_0$  تا  $P_{15}$ ، در جهت پادساعتگرد در شکل (1) برچسب زده شده است. بنابراین، مجموعه پارامترهای مدل لب



شکل 3: مدل هندسی لب

### (3-3) فرایند تصدیق مدل لب

فرآیند بدست آوردن شکل واقعی لب به طریقه زیر می‌باشد:

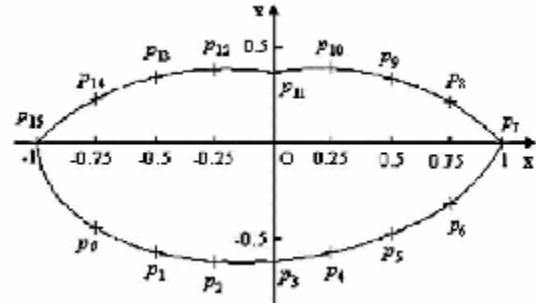
1- همه نقاط لب نرمال سازی شوند و 2 منحنی درجه دوم و چهارم به نقاط نرمال سازی شده مربوطه (برای مثال منحنی فرمول بندی شده بوسیله معادله (4) با نقاط پایینی لب) با روش حداقل مربعات خطا تطبیق پیدا کند و بنابراین مجموعه پارامترهای  $I_g$  بدست می‌آید.

2- 16 نقطه لب با استفاده از پارامترهای هندسی منحنی‌ها بدست می‌آیند. مختصات  $x$  نرمال سازی شده این نقاط بر روی محور  $x$  تعیین شده است برای مثال، مختصات  $x$  نقاط کانتور پایینی لب  $x_{p0} - x_{p6}$  برابر با  $\left\{-\frac{3}{4}, -\frac{1}{2}, -\frac{1}{4}, 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}\right\}$  و نقاط کانتور بالایی لب  $x_{p7} - x_{p15}$  برابر با  $\left\{1, \frac{3}{4}, \frac{1}{2}, \frac{1}{4}, 0, -\frac{1}{4}, -\frac{1}{2}, -\frac{3}{4}, -1\right\}$  می‌باشد و مختصات  $y$  نرمال سازی شده نقاط لب به آسانی بر طبق پارامترهای منحنی‌ها می‌تواند محاسبه شود. در نهایت محورهای مختصات نرمال سازی شده با معکوس کردن فرآیند نرمال سازی می‌تواند به محورهای مختصات اصلی تبدیل شوند.

### (4) پیاده سازی استخراج کانتور لب

#### (1-4) جداسازی لب و تولید نگاهت احتمال

به عنوان قسمت ابتدایی بسیاری از الگوریتمهای استخراج کانتور لب، جداسازی لب می‌باشد که با استفاده از جداسازی لب، ویژگیهای مهم لب بدست می‌آید و با استفاده از این ویژگیها در تابع انرژی، مکان کانتور لب را مشخص می‌کنیم. این ویژگیها با استفاده از رویکردهای مختلف که بر روی تصاویر سطوح خاکستری یا تصاویر رنگی عمل می‌کنند،



شکل 2: مدل 16 نقطه ای نرمالیزه شده لب

### (2-3) توصیف محدودیتهای هندسی لب

به منظور اطمینان از معنی دار بودن کانتور لب، از یک مدل هندسی بهره گرفته شده که این مدل شامل 2 منحنی درجه دوم و چهارم به هم اتصال یافته می‌باشد و معادلات این منحنی‌ها به طریقه زیر است:

برای کانتور لب پایینی که از گوشه چپ  $P_{15}(-1,0)$  شروع می‌شود و به گوشه راست  $P_7(1,0)$  ختم می‌شود:

$$y_1(x) = h_1 \left(1 - \frac{x^2}{1}\right) + 4 * q_1 (x^4 - x^2) \quad (4)$$

برای کانتور لب بالایی که از گوشه چپ  $P_{15}(-1,0)$  شروع می‌شود و به نقطه  $P_7(1,0)$  ختم می‌شود:

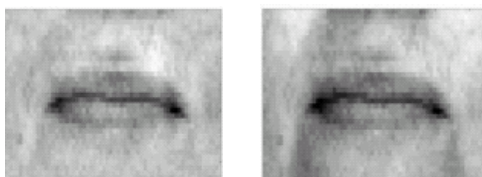
$$y_2(x) = -h_2 \left(1 - \frac{(x-a)^2}{(1-a)^2}\right) + 4 * q_2 \left(\frac{(x-a)^4}{(1-a)^4} - \frac{(x-a)^2}{(1-a)^2}\right) \quad (5)$$

پارامتر  $a$ ، مبین میزان فرورفتگی کانتور لب بالایی در نقطه وسط و پارامترهای  $q_1$  و  $q_2$  میزان انحنا و پارامترهای  $h_1$  و  $h_2$  میزان ارتفاع دو کانتور را تعیین می‌کنند. در شکل (3) مدل هندسی لب نشان داده شده است. در نهایت، محدودیتهای هندسی شکل واقعی لب با استفاده از مجموعه پارامترهای  $I_g = \{h_1, q_1, h_2, q_2, a\}$  می‌تواند توصیف شود. مدل هندسی برای توصیف شکلهای مختلف لب در گویندگان مختلف از این جهت انعطاف پذیر است که هیچ فرض محدودکننده‌ای در مورد ارتباط بین لب بالایی و لب پایینی بجز دو نقطه گوشه وجود ندارد. به هر حال بدست آوردن این پارامترها به طور مستقیم از تصویر نسبتاً پیچیده می‌باشد و بنابراین مدل هندسی تنها به عنوان یک محدودکننده برای شکل‌های واقعی لب عمل می‌کند.

سازی به طریقه زیر انجام می‌شود:

$$L'(i, j) = L(i, j) + \frac{L(j) - u(j)}{N - 1} (i - 1) + m_{Lu} - L(j) \quad (6)$$

که  $L(j)$  و  $u(j)$  به ترتیب مین تغییرات شدت روشنایی مرز بالا و پایین هستند.  $m_{Lu}$  مقدار میانگین  $L(j)$  و  $u(j)$  و  $N$  تعداد ردیفهای تصویر می‌باشد. تصویر اولیه و نتایج نرمال سازی شدت روشنایی در شکل (4) نشان داده شده است.



الف: تصویر شدت روشنایی اولیه. ب: تصویر شدت روشنایی نرمالیزه شده

شکل 4: الف) تصویر شدت روشنایی اولیه. ب) تصویر شدت روشنایی نرمالیزه شده

فرآیند پیش پردازش مهم دیگر در الگوریتم خوشه بندی، ماسک کردن پیکسلهای مربوط به دندانها می‌باشد. در FCMS، پیکسلهای دندان و مناطق شدت روشنایی کم که دارای اطلاعات رنگی نویزی و غیرقابل اعتماد باشند، دقت خوشه بندی دو کلاس (ناحیه لب و غیر لب) را کاهش می‌دهند. بنابراین این پیکسلها باید قبل از خوشه بندی ماسک شوند. هر پیکسل با  $a^* \leq t_a$  به عنوان پیکسل دندان در نظر گرفته می‌شود و آنها را در فرآیند خوشه بندی به حساب نمی‌آوریم [10] که  $t_a$  یک آستانه برای مولفه های رنگی  $a$  می‌باشد (ضمیمه مشاهده شود). ضمناً پیکسلهایی با  $L^* \leq 35\%$  رفرنس سفید، ماسک می‌شوند.

بعد از خوشه بندی، مقدار احتمال مناسب برای پیکسلهای ماسک شده را تعیین می‌کنیم. از آنجائیکه پیکسلهای دندان، همیشه داخل ناحیه دهان می‌باشند و به عنوان لب بالایی یا پایینی در نظر گرفته نشوند، به آنها مقدار احتمال بالایی تخصیص می‌دهیم. برای پیکسلهای با شدت روشنایی خیلی کم، احتمال  $0/5$  در نظر گرفته می‌شود زیرا احتمال اینکه پیکسل لب یا غیر لب باشند مساوی است. در نهایت نگاشت احتمال تصویر اولیه بوسیله یک فیلتر پایین گذر  $3 \times 3$  هموار می‌شود. یک مثال از فرآیند جداسازی در شکل (5)

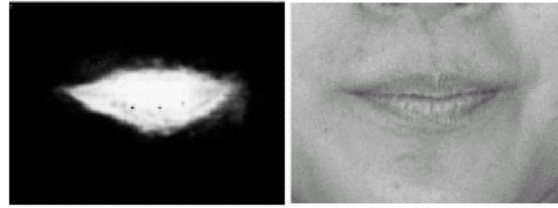
بدست می‌آیند. پردازش تصاویر رنگی در سالهای اخیر مورد توجه بسیاری از محققین قرار گرفته است. زیرا رنگ، اطلاعات اضافی که در تصاویر سطوح خاکستری وجود ندارد را فراهم می‌کند. در رفرنس [6]، یک مدل گوسین مخلوط که اطلاعات رنگ یک گوینده مشخص را مدل می‌کند، برای تفاوت قائل شدن پیکسلهای لب از زمینه آموزش داده می‌شود و در رفرنس [7]، یک تبدیل رنگی برای بزرگ جلوه دادن تفاوت بین لب و پوست پیشنهاد شده است.

در این مقاله از الگوریتم خوشه بندی فازی با تابع شکلی (Fuzzy C-Means (FCMS)) در رفرنس [8]، برای جداسازی لب بهره گرفته شده است. در این روش خوشه بندی، از هر دو اطلاعات رنگ و فاصله فضایی (Spatial) استفاده شده است و توابع شکلی در معیار تفاوت تابع هدف جاسازی شده است تا بین پیکسلهایی که اطلاعات رنگ شبیه هم، اما در ناحیه های متفاوت قرار گرفته‌اند، تفاوت قائل شویم. بر اساس مینیمم سازی تابع هدف، الگوریتم یک مقدار احتمال مناسب برای هر پیکسل در تصویر اختصاص می‌دهد. این یک روش یادگیری بدون سرپرستی می‌باشد و نه فرض اولیه ای در مورد توزیع ویژگیها در نظر می‌گیرد و نه آموزش مورد نیاز می‌باشد. بنابراین نتایج جداسازی بین ناحیه لب و غیر لب با وجود تغییرات رنگ لب قابل اطمینان تر است.

در آزمایشات ما تصاویر بوسیله یک دوربین با فرمت RGB گرفته شده است. در FCMS، از یک معیار تفاوت اقلیدسی (Euclidean) برای توصیف تفاوت رنگ قابل ادراک استفاده می‌شود. اما فضای رنگ RGB یکنواخت نمی‌باشد [9]. به طور مثال، فاصله بین دو نقطه در فضای رنگ RGB متناسب با تفاوت رنگ آنها نمی‌باشد. در نتیجه یک تبدیل رنگی مناسب مورد نیاز می‌باشد. فضای رنگ یکنواخت LAB برای سهولت پردازشهای بعدی به کار برده شده است [9] و بردار رنگ  $(L^*, a^*, b^*)$  برای بیان اطلاعات رنگ هر پیکسل استفاده شده است.

شدت روشنایی غیر یکنواخت، نتایج خوشه بندی را با دادن اطلاعات غیردقیق (مولفه  $L$ ) خراب می‌کند. در نتیجه، نرمال سازی تغییرات شدت روشنایی مورد نیاز می‌باشد. فرآیند نرمال

نشان داده شده است.



الف: تصویر لب اولیه. ب: نگاهت احتمال (الف) که بوسیله

شکل 5: الف) تصویر لب اولیه. ب) نگاهت احتمال (الف) که بوسیله الگوریتم FCMS ایجاد شده است

است، گسسته می‌باشد،  $Prob_L(x, y)$  و  $Prob_{nL}(x, y)$  در معادله (8) در مکانهای گسسته تعریف شده‌اند. اما به علت پیوسته بودن کانتور لب ارائه شده، استفاده تنها از شبکه اعداد صحیح برای ارزیابی تطابق مدل لب کافی نمی‌باشد. بنابراین از درون یابی دو سویه (bilinear interpolation) بهره گرفته می‌شود تا تابع هزینه را برای دامنه پیوسته بسط دهیم و معادله (8) برای مینیمم سازی  $F_B$  به طریقه زیر می‌تواند فرمول بندی شود:

$$F_B = - \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} \int_0^{g_i(u)} f(u, v) dv du \quad (9)$$

که  $f(u, v) = \log(Prob_L(u, v)) - \log(Prob_{nL}(u, v))$  تفاوت احتمال لگاریتمی بین ناحیه لب و غیر لب می‌باشد و  $g_i(u)$  مرز لب بین نقطه  $i^{th}$  و  $(i+1)^{th}$  با استفاده از درون یابی درجه اول می‌باشد بنابراین:

$$g_i(u) = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} (u - x_i) + y_i \quad (10)$$

که  $(x_i, y_i)$  مختصات نقطه  $i^{th}$  ام،  $x_N = x_0$  و  $y_N = y_0, (N=16)$  می‌باشد.

### 3-4 مقداردهی اولیه مدل

مدل ابتدایی لب با بهره گیری از نگاهت احتمال ایجاد شده بوسیله الگوریتم FCMS، به سادگی استخراج می‌شود. با بررسی کردن ستونها، مرزهای راست و چپ لب بدست می‌آیند و مرزهای بالا و پایین، از محور مرکزی بین نقاط گوشه راست و چپ بدست می‌آیند. سپس از معادلات دو منحنی تعریف شده برای تطبیق با نقاط لب پایینی و بالایی به طور جداگانه استفاده می‌شود و پارامترهای هندسی اولیه بدست می‌آیند. در نهایت مدل نقطه‌ای اولیه با استفاده از این پارامترها و روش توصیف شده در قسمت (3-3) بدست می‌آید. از آنجائیکه تابع انرژی بر اساس ناحیه می‌باشد، نتایج نهایی استخراج در مقایسه با تکنیکهای مبتنی بر لبه، حساسیت کمتری به تنظیم اولیه پارامترها دارد. در آزمایشات ما مقداردهی اولیه مدل با چنین تخمین ساده‌ای منجر به پیدا کردن مناسب کانتور لب شده است.

### 2-4 فرمول سازی تابع انرژی

برای بدست آوردن یک مدل دقیق، یک تابع انرژی برای تعیین کردن پارامترهای بهینه مدل به طریقه‌ای که پیکسل‌های داخل ناحیه لب دارای احتمال بالای بودن پیکسل لب، در حالی که پیکسل‌های خارجی دارای احتمال پایینی باشند، تعیین می‌شود. فرض می‌کنیم که ناحیه لب و غیر لب با هم، هم پوشانی ندارند. زمانی که تابع انرژی در معادله (7) ماکزیمم شود، جداسازی به خوبی صورت گرفته است:

$$\max \left\{ C(I_p) = \prod_{(x,y) \in R_L(I_p)} Prob_L(x, y) * \prod_{(x,y) \in R_{nL}(I_p)} prob_{nL}(x, y) \right\} \quad (7)$$

که  $I_p$  پارامترهای مدل 16 نقطه‌ای می‌باشد و  $R_L$  ناحیه احاطه شده بوسیله مدل نقطه‌ای و  $R_{nL}$  ناحیه خارج از مدل نقطه‌ای می‌باشد.  $Prob_L(x, y)$  احتمال اینکه پیکسل در مکان  $(x, y)$  یک پیکسل لب باشد و  $prob_{nL}(x, y)$  احتمال اینکه پیکسل خارج از لب باشد را بیان می‌کند. از معادله (7) لگاریتم گرفته می‌شود و از آنجائیکه  $\sum_{(x,y) \in R} \log(Prob_{nL}(x, y))$  در فرآیند بهینه سازی ثابت می‌ماند ( $R$  ناحیه کامل تصویر می‌باشد)، این عبارت می‌تواند از فرمول بندی تابع هزینه حذف شود و معادله (8) بدست می‌آید:

$$\max \left\{ \begin{aligned} \log(C(I_p)) &= \sum_{(x,y) \in R_L(I_p)} \left[ \log(Prob_L(x, y)) - \log(Prob_{nL}(x, y)) \right] \\ &= \sum_{(x,y) \in R_L(I_p)} f(x, y) \end{aligned} \right\} \quad (8)$$

از آنجائیکه نگاهت احتمال که بوسیله جداسازی لب ایجاد شده

#### 4-4 فرایند بهینه سازی

با اصلاحاتی بر روی روش بهینه سازی بر اساس نقاط در رفرنس [11] از آن استفاده می‌شود. فرآیند بهینه سازی، یک فرآیند تکراری است و در هر بار تکرار، مکان نقاط لب به منظور کاهش تابع انرژی  $F_B$  تنظیم می‌شوند. بنابراین به خوبی سیر تکامل و تغییر شکل کانتور لب کنترل می‌شود و در نتیجه به سرعت همگرایی بدست می‌آید. بردار جابجایی  $\Delta I_p$  و مدل جدید لب به طریقه زیر می‌توانند به روز شوند:

$$\Delta I_p = \{dx_i, dy_i\} = \left\{ -\frac{\partial F_B}{\partial x_i}, -\frac{\partial F_B}{\partial y_i} \right\}$$

$$i = 0, 1, \dots, 15 \quad (11)$$

$$I_{p,new} = I_{p,old} + w\Delta I_p \quad (12)$$

که  $w$  اندازه جابجایی (Step Size) در هر مرحله و مقدار آن از طریق تجربه 0.05 می‌باشد.

بعد از نرمال سازی، 2 منحنی درجه دوم و چهارم با 16 نقطه لب تطبیق داده و پارامترهای  $I_g$  بدست می‌آیند سپس مدل 16 نقطه‌ای محدود شده با استفاده از این پارامترها بدست می‌آید. روش استخراج کانتور لب به طریقه زیر می‌باشد:

1- ایجاد نگاشت احتمال با استفاده از الگوریتمهای FCMS

و بدست آوردن پارامترهای هندسی اولیه  $I_g$  و پارامترهای

مدل 16 نقطه‌ای  $I_p$

2- محاسبه بردار جابجایی  $\Delta I_p$  و کانتور جدید

لب  $I_{p,new}$  با استفاده از روش بهینه سازی توصیف شده

در رفرنس [11].

3- بدست آوردن مدل 16 نقطه‌ای محدود شده بوسیله

فرآیند تصدیق شکل لب.

4- محاسبه تابع انرژی  $F_B$  و تکرار مرحله 2 تا تغییرات

تابع هزینه کمتر از آستانه از پیش تعریف شده باشد.

#### 5 نتایج آزمایشگاهی

##### 1-5 نتایج استخراج لب

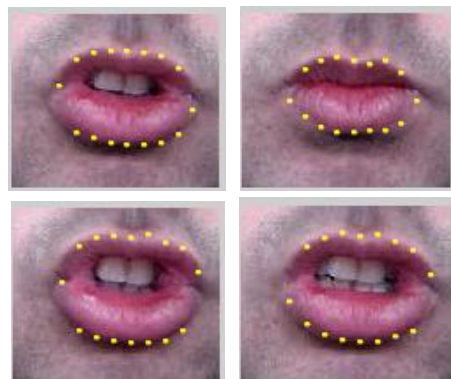
1200 تصویر با اندازه  $100 \times 120$  از 20 فرد مختلف با

استفاده از الگوریتم پیشنهاد شده، آنالیز شده است. مشاهده

شده که در بیشتر موارد کانتور لب به طور دقیق بدست آمده است. همچنین ما موارد نامناسب را بررسی کردیم (کمتر از 3 درصد از مجموعه تست) و متوجه شدیم که علت اصلی نادرست بودن کانتور لب، نگاشت احتمال ضعیف تصویر اولیه که به علت کنتراست خیلی ضعیف بین لب و زمینه ایجاد شده است، می‌باشد. با بهره گیری از پارامترهای مدل بدست آمده از فریم قبلی به عنوان پارامترهای اولیه برای فریم بعدی، تعداد تکرارها به طور قابل توجهی کاهش پیدا کرده است (متوسط تکرار برای یک تصویر 15 می‌باشد). نتایج آزمایشگاهی نشان می‌دهد که سیستم زمان واقعی ما می‌تواند کانتور لب را به صورت همزمان استخراج کند. شکل (6) و (7) نتایج ردیابی لب را نشان می‌دهد.



شکل 6: راست به چپ، بالا به پایین: نتایج انطباق مدل 16 نقطه‌ای به کانتور لب در فریمهای 1، 5، 12، 17 از دنباله تصاویر فردی در حالت ادای کلمه یک



شکل 7: راست به چپ، بالا به پایین: نتایج انطباق مدل 16 نقطه‌ای به کانتور لب در فریمهای 1، 5، 10، 13 از دنباله تصاویر فردی در حالت ادای کلمه چهار

مشاهده شده است که ناحیه دندان نسبت به بقیه تصویر لب مقدار  $a^*$  پایبندی دارد. نتایج آزمایشگاهی رضایت بخشی بر روی تصاویر مختلفی از لب با آستانه دندان  $t_a$  که به طریقه زیر تنظیم شده، بدست آمده است:

$$t_a = \begin{cases} m_a - s_a & \text{if } (m_a - s_a) \pi > 9 \\ 9 & \text{otherwise} \end{cases} \quad (13)$$

$$t_a = \begin{cases} m_a - s_a & \text{if } (m_a - s_a) \pi > 9 \\ 9 & \text{otherwise} \end{cases}$$

### مراجع

- [1] Zhang. Y; Levinson. S; Huang. T, "Speaker independent audio-visual speech recognition", Proceedings of IEEE Int. Conf. on Multimedia and Expo, Vol. 2, New York, USA, July 2000, pp. 1073-1076
- [2] Yu. K; Jiang. X; Bunke. H, "Lipreading Using Signal Analysis Over Time", Signal\_Processing, Vol. 77, (1999), pp. 195-208
- [3] Chiou. G.I; Hwang. J.N, "Lipreading from color video", IEEE Trans. Image Processing, Vol. 6, (1997), pp. 1192-1195.
- [4] Luetttin. J; Thacker. N.A; Beet. S.W, "Locating and Tracking Facial Speech Feature", 13<sup>th</sup>. Int. Conf. on Pattern Recognition, Vienna, Austria, Vol. 1, (1996),
- [5] Rao. R.R; Mersereau. R.M, "Lip modeling for visual speech recognition", 28 Asilomar Conf. on Signal Systems and Computer, Vol. 1, (1995), pp. 587-590.
- [6] Tian. Y; Kanade. T; Cohn J.F, "Robust lip tracking by combination shape, color and motion", Proceeding of Fourth Asian Conf. on Computer Vision, Taipei, January 2000, pp. 1040-1045.
- [7] Eveno. N; Caplier. A; Coulon. P.Y, "New color transformation for lips segmentation", Proceeding of IEEE fourth Workshop on Multimedia signal Processing, Cannes, France, October 2001, pp. 3-8
- [8] Wang. S.L; Leung. S.H; Lau. W.H, "lip segmentation by fuzzy clustering incorporating with shape function", Proceedings of IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Vol. 1, Orlando, May 2002, pp. 1077-1080
- [9] CIE, Colorimetry, CIE Pub. No. 15.2, Bureau Central de la CIE, Vienna, Austria, 1986
- [10] Liew. A.W.C; Leung. W.H, "Segmentation of color lip images by spatial fuzzy clustering", IEEE Trans. Fuzzy syst, Vol. 11, 2003, pp. 542-549
- [11] Sum. K.L; Lau. W.H; Leung. S.H; Liew. A.W.C; Tse. K.W, "A New optimization procedure for extracting the point-based lip contour using active shape model", Proceeding of IEEE Int. conf. on Acoustic, Speech and signal processing, Vol. 3, Salt Lake City, USA, May 2001, pp. 1485-1488.

اطلاعات تصویری از شکلهای لب و حرکات آن، دقت و اطمینان سیستمهای تشخیص گفتار را خصوصاً در محیطهای نویزی بهبود می دهد. بنابراین توسعه یک الگوریتم دقیق و قابل اطمینان برای استخراج ویژگیهای تصویری مربوط به گفتار حیاتی می باشد؛ اما به هر حال وجود تغییرات زیاد در تصاویر مختلف ناشی از گویندگان مختلف، لحن متفاوت ادای کلمات توسط آنها و شرایط نوری، این امر را مشکل ساخته است. به علاوه کنتراست کم شدت روشنایی و رنگ بین لب و پوست صورت، مشکل بزرگ دیگر برای آشکارسازی کانتور لب می باشد.

در این مقاله، یک الگوریتم جدید برای استخراج کانتور لب مبتنی بر مدل که شایستگیهای مدلهای نقطه ای و مدلهای پارامتری را دارد برای غلبه بر کاستیهای موجود استفاده شده است. دقت، قابلیت اطمینان و زمان پردازش نگرانیهای اولیه ما در طراحی الگوریتم بوده است. از یک مدل 16 نقطه ای برای بیان کانتور لب بهره گرفته شده است. بوسیله روشهای خوشه بندی فازی که توابع شکلی را در خود جای داده، یک نگاشت احتمال از تصاویر رنگی بدست می آوریم. سپس با استفاده از این نگاشت، یک تابع انرژی بر اساس ناحیه بندی تصاویر که احتمال اتصال ناحیه لب و غیر لب را ماکزیم می کند، تعریف می شود و با فرآیند بهینه سازی نقطه ای تکرارپذیر، مدل لب را بر روی نگاشت احتمال تصویر تحت بررسی منطبق می کنیم. در این فرآیند با هر بار تکرار، تنظیم 16 نقطه لب بوسیله 2 منحنی درجه دوم و چهارم کنترل می شود که این دو منحنی با محدود کردن این نقاط، شکل فیزیکی مناسبی برای لب ایجاد می کنند.

الگوریتم پیشنهاد شده بر روی 1200 تصویر لب از 20 فرد مختلف تست شده است. نتایج آزمایشگاهی نشان می دهد که الگوریتم ما در بیشتر موارد کانتور لب را به طور دقیق مشخص کرده است و در نهایت یک سیستم استخراج کانتور لب زمان واقعی با استفاده از الگوریتم ارائه شده، پیاده سازی شده است.