

بسم الله الرحمن الرحيم



دانشکده مهندسی کامپیوتر و فناوری اطلاعات

پایان نامه کارشناسی ارشد مهندسی هوش مصنوعی

مدل سازی چهره با استفاده از واژه نامه انسداد زاویه ای

نگارنده: سینا اخلاقی

استاد راهنما

آقای دکتر حمید حسن پور

استاد مشاور

آقای دکتر وحید ابوالقاسمی

بهمن ۱۳۹۵



باسمه تعالی

شماره:

تاریخ:

ویرایش:

مدیریت تحصیلات تکمیلی

دانشکده :

گروه :

پایان نامه کارشناسی ارشد آقای / خانم به شماره دانشجویی:

تحت عنوان:

در تاریخ توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد
مورد پذیرش قرار گرفت.
مورد ارزیابی و با درجه

امضاء	اساتید مشاور	امضاء	اساتید راهنما
	نام و نام خانوادگی :		نام و نام خانوادگی :
	نام و نام خانوادگی :		نام و نام خانوادگی :

امضاء	نماینده تحصیلات تکمیلی	امضاء	اساتید داور
	نام و نام خانوادگی :		نام و نام خانوادگی :
			نام و نام خانوادگی :
			نام و نام خانوادگی :
			نام و نام خانوادگی :

پاس کزاری



پدر عزیزم:

اعتیاجی به تسلیج نیست...

دستت را که به من بدهی...

با انگشتت ذکر دوست داشتن می گویم...

پدرم، دستانت، چشمت و صدایت را عاشقانه دوست دارم...

به پاس اولین بوسه ای که پدرم به رسم مهر در اولین ساعات تولد به گونه ام زد و من نفهمیدم و به یاد سوزنده ترین بوسه ی آخر که من به رسم وداع به صورتش

زدم و او نفهمید

یادش را کرامی میدارم

مادر عزیزم:

پاس یکران بر همدلی و بهمراهی و بهکامی مادر و سوز و مهربانم که سجده ی ایشان گل محبت را در وجودم پروراند و دلمان گهربارش بخت های مهربانی را به من

آموخت.

عاجزانه از خداوند منان خواستارم تا سایه ی مادرم را بر سرم حفظ نماید.

مشکرو قدردانی

به مصداق «من لم یسکر المخلوق لم یسکر الخالق» بسی شایسته است از استاد فرهیخته و فرزانه ام جناب آقای دکتر حمید حسن پور که با کرامتی چون خورشید، سرزمین دل را روشنی بخشید و گلشن سرای علم و دانش را با راهنمایی های کار ساز و سازنده بارور ساختند؛ تقدیر و تشکر نمایم.

و

از استاد فاضل و اندیشمند جناب آقای دکتر وحید ابوالقاسمی به عنوان استاد مشاور که همواره بنده را مورد لطف و

محبت خود قرار داده اند، کمال تشکر را دارم.

تعهد نامه

اینجانب سینا اخلاقی دانشجوی دوره کارشناسی ارشد رشته هوش مصنوعی دانشکده کامپیوتر و فناوری اطلاعات دانشگاه صنعتی شاهرود نویسنده پایان نامه مدل سازی چهره با استفاده از واژه نامه ی انسداد زاویه ای تحت راهنمایی آقای دکتر حمید حسن پور متعهد می شوم.

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود است و مقالات مستخرج با نام «دانشگاه صنعتی شاهرود» و یا «Shahrood University of Technology» به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه، در مواردی که از موجود زنده (یا بافتهای آن ها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است. تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه های رایانه ای، نرم افزارها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود است. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

چکیده

در نظارت ویدئویی، زاویه‌ی دید چهره نسبت به دوربین (انسداد زاویه‌ای) توانایی سیستم در شناسایی چهره را محدود می‌سازد. در این پایان‌نامه روشی برای رفع انسداد زاویه‌ای در تصاویر چهره پیشنهاد می‌شود که مبتنی بر دگرديسی تصویر بوده و با دریافت تعدادی تصویر زاویه‌دار از چهره‌ی یک شخص، تصویری از روبه‌رو مدل می‌سازد. روش پیشنهادی با تکنیک‌هایی ساده از قبیل درونیابی مکانی پیکسل‌های تصاویر ورودی با توابع انتقال، درونیابی خطی و میانگین‌گیری از شدت روشنایی آنها، مدل چهره از روبه‌رو را بدست می‌آورد.

برای بهبود تصویر مدل، روش پیشنهادی برای مدل‌سازی تصاویر چهره‌ی روبه‌رو، بر روی تصاویر ورودی واگرا (تصاویر چهره‌ی دارای زاویه‌ی قرینه نسبت به هم)، اعمال می‌شود. سپس تصاویر مدل شده، با تجزیه‌ی مرتبه پایین، نسبت به هم تراز می‌شوند. شبکه‌ی عصبی تابع پایه شعاعی به عنوان تابع انتقال برای این روش در نظر گرفته شده است. از مزایای این روش عدم نیاز به اطلاعات عمق در تصویر، کالیبره بودن تصاویر و اطلاعات زاویه‌ی چهره است که معمولاً در مدل‌سازی سه بعدی استفاده می‌شود. عملکرد الگوریتم روی تصاویر پایگاه داده PRIMA مورد بررسی قرار گرفته است. نتایج آزمایشات نشان می‌دهد که روش پیشنهادی، با دقت مناسبی تصویر نمای روبه‌روی چهره را از تصاویر زاویه‌دار مدل‌سازی می‌نماید.

واژه‌های کلیدی: انسداد زاویه‌ای، دگرديسی، تابع پیچ و تاب، ترازسازی، شبکه‌ی عصبی تابع پایه شعاعی.

لیست مقالات مستخرج از پایان نامه

- اخلاقی سینا، حسن پور حمید، ابوالقاسمی وحید، "ارائه‌ی روشی برای رفع انسداد زاویه‌ای در تصاویر چهره"، کنفرانس بین المللی پردازش سیگنال و سیستم‌های هوشمند، دانشگاه صنعتی امیر کبیر، تهران، ص ۸۱، (آذر ۱۳۹۵).
- اخلاقی سینا، حسن پور حمید، ابوالقاسمی وحید ۱۳۹۵، "مدل‌سازی چهره با استفاده از میانگین‌گیری بر پایه‌ی دگرذیسی تصویر و تجزیه‌ی مرتبه پایین"، نشریه مهندسی برق و الکترونیک ایران، (ارسال شده).

فهرست مطالب

فصل اول: مقدمات

۲	۱-۱- سیستم‌های شناسایی چهره بر اساس تصویر
۳	۲-۱- اهمیت وجود سیستم استخراج چهره
۴	۳-۱- کاربردهای سیستم‌های استخراج چهره
۵	۴-۱- چالش‌های سیستم‌های شناسایی چهره
۶	۵-۱- مدل‌سازی چهره
۷	۶-۱- بازنمایی تنک
۷	۷-۱- یادگیری واژه‌نامه
۸	۶-۱- تعریف مساله
۹	۷-۱- اهداف پایان‌نامه
۱۰	۸-۱- پایگاه داده‌ی مورد استفاده
۱۱	۹-۱- ساختار پایان‌نامه

فصل دوم: مروری بر روش‌های استخراج چهره و نرمال‌سازی

۱۴	مقدمه
۱۴	۱-۲- مروری بر روش‌های استخراج چهره
۱۶	۲-۱-۱- روش‌های بر پایه‌ی ویژگی
۱۷	۲-۱-۱-۲- اطلاعات سطح خاکستری و رنگ
۱۸	۲-۱-۱-۳- معیارهای عام
۲۰	۲-۱-۲- تجزیه و تحلیل ویژگی
۲۰	۲-۱-۲-۱- تجزیه و تحلیل بر اساس مدل‌های سلسله‌مراتبی چهره
۲۱	۲-۱-۲-۲- مدل‌های شکلی-فعال
۲۱	۲-۱-۳-۱- اسنیک‌ها
۲۳	۲-۱-۴- مشکلاتی که روش‌های بر پایه‌ی ویژگی با آن مواجه‌اند
۲۳	۲-۱-۶- مدل‌های بر پایه‌ی تصویر
۲۴	۲-۱-۶-۱- روش‌های بر پایه‌ی تبدیل فضای تصویر به فضای محدودتر
۲۴	۲-۱-۶-۲- تحلیل مؤلفه‌های اصلی

۲۶	۲-۱-۳- شبکه‌های عصبی
۲۸	۲-۱-۷- مشکلاتی که روش‌های مبتنی بر تصویر با آن مواجه‌اند
۲۸	۲-۲- روش‌های نرمال‌سازی
۲۸	۲-۲-۱- رفع انسداد از تصاویر چهره
۳۰	۲-۲-۲- تصحیح شدت روشنایی
۳۱	۲-۳- جمع‌بندی

فصل سوم: روش پیشنهادی

۳۴	مقدمه
۳۴	۳-۱- انتقال نسبی
۳۵	۳-۲- ترازسازی مقاوم با تجزیه‌ی تنک و مرتبه پایین
۴۲	۳-۳- پیچ و تاب تصویر
۴۳	۳-۴- دگردیسی تصویر
۴۴	۳-۵- انواع زوایای تصویر
۴۶	۳-۶- روش پیشنهادی
۵۵	۳-۸- جمع‌بندی
۵۸	مقدمه

فصل چهارم: جزئیات پیکربندی و ارزیابی روش پیشنهادی

۵۸	۴-۱- تحلیل اثر روش پیشنهادی
۶۲	۴-۲- تنظیم پارامترها برای روش پیشنهادی
۶۲	۴-۳- محدوده‌ی زاویه‌ی چهره‌ی قابل تصحیح توسط روش پیشنهادی
۶۴	۴-۴- چارچوب آزمایش و نتایج کمی
۶۸	۴-۵- جمع‌بندی

فصل پنجم: بحث و نتیجه‌گیری

۷۰	۵-۱- بحث
۷۱	۵-۲- نتیجه
۷۲	مراجع

فهرست اشکال

- شکل ۱-۱ تصاویر آموزشی سیستم‌های شناسایی چهره ۳
- شکل ۲-۱ تصاویر چهره در دنیای واقعی ۴
- شکل ۳-۱ فضای مساله ۹
- شکل ۴-۱ نمونه‌هایی از تصاویر پایگاه داده‌ی PRIMA ۱۰
- شکل ۱-۲ نمودار دسته‌بندی ساختاری روش‌های استخراج چهره [۱] ۱۶
- شکل ۲-۲ نتیجه‌ی استخراج چهره با روش [۱۲] ۱۹
- شکل ۳-۲ نتیجه‌ی استخراج چهره با استفاده از روش [۱۳] ۲۱
- شکل ۴-۲ نمونه‌ای از اسنیک قفل شده در تصویر چهره با روش [۱۴] ۲۳
- شکل ۵-۲ چهره‌های ویژه‌ی استخراج شده با روش [۱۵] ۲۵
- شکل ۶-۲ معماری شبکه‌ی عصبی [۱۶] برای استخراج چهره ۲۷
- شکل ۷-۲ داده‌های آموزشی به همراه خلاصه‌ی روش [۱۷] ۳۰
- شکل ۸-۲ شمای کلی روش [۱۸] ۳۱
- الگوریتم ۱-۳ حلقه‌ی خارجی RASL [۲۴] ۳۹
- الگوریتم ۲-۳ حلقه‌ی داخلی RASL [۲۴] ۴۰
- شکل ۳-۳ نمونه‌ای از ترازسازی تصاویر چهره با RASL ۴۱
- شکل ۴-۳ ترازسازی تصاویر زاویه‌دار چهره با RASL ۴۱
- شکل ۵-۳ نمونه‌ای از فرایند دگردیسی برای تصاویر چهره ۴۳
- شکل ۶-۳ انواع زاویه‌های موجود در تصویر ۴۴
- شکل ۷-۳ خلاصه‌ی روش پیشنهادی در سه مرحله ۴۵
- شکل ۸-۳ تصاویر ورودی روش پیشنهادی به همراه تصویر روبه‌روی پایگاه داده ۴۷

- ۴۸ شکل ۹-۳ نمودار گردشی روش پیشنهادی
- ۵۲ شکل ۱۰-۳ مدل سازی تصویر روبه رو با روش پیشنهادی
- ۵۳ شکل ۱۳-۳ نتیجه ی چهره با روش پیشنهادی و تراز سازی نتایج آن با RASL
- ۵۴ شکل ۱۴-۳ مقایسه ای میان روش های مطرح شده
- ۵۴ شکل ۱-۴ تحلیل اثر الگوریتم
- ۵۹ شکل ۲-۴ تحلیل اثر انتقال با توابع در نظر گرفته شده در روش پیشنهادی
- ۶۱ شکل ۳-۴ تصاویر روبه روی مدل شده از پایگاه داده ی PRIMA
- ۶۷ شکل ۱-۵. نمونه ای از خروجی روش های مکان یابی بخش های چهره [۳۵-۴۰]

فهرست جداول

- ۶۱ جدول ۱-۴ اندازه ی حرکت نقاط در تصاویر خروجی دگردیسی با تابع انتقال روش پیشنهادی
- ۶۳ جدول ۲-۴ مشخص سازی محدوده ی مناسب زاویه ی چهره برای مدل سازی در روش پیشنهادی
- ۶۴ جدول ۳-۴ زوایای تخمین زده شده برای چهره های مدل شده با استفاده از روش [۳۵]
- ۶۶ جدول ۴-۴ مقایسات شباهت چهره های مدل شده با چهره های روبه روی پایگاه داده
- ۶۷ جدول ۵-۴ جدول درصد نتایج مطلوب

فصل اول

مقدمات

۱-۱- سیستم‌های شناسایی چهره بر اساس تصویر

امروزه به دلیل کاربردهای گسترده از جمله تامین امنیت و ثبت ورود و خروج کارکنان در سازمان‌ها، نیاز به شناسایی هویت بیش از پیش احساس می‌شود. روش‌های متعددی برای تحقق این امر وجود دارند. روش‌های شناسایی هویت را می‌توان در سه دسته به صورت زیر قرار داد.

❖ شناسایی هویت بر پایه‌ی نشانه: این دسته از سیستم‌ها بر مبنای شی‌ئی هستند که کاربر به همراه خود دارد. مانند شناسنامه و کارت ملی.

❖ شناسایی هویت بر پایه‌ی دانش: سیستم‌های بر پایه‌ی رمز عبور (که عموماً کاربر آن‌ها را به خاطر دارد)، جزو این دسته از سیستم‌ها به شمار می‌روند.

❖ روش‌های بیومتریک شناسایی هویت: این سیستم‌ها با استفاده از ویژگی‌های فیزیولوژیکی و رفتاری انسان، هویت را تشخیص می‌دهند. از جمله‌ی این روش‌ها می‌توان به سیستم تشخیص اثر انگشت و تشخیص چهره اشاره نمود.

مفقودی نشانه یا همراه نداشتن آن توسط کاربر، فرسوده شدن نشانه و جعل یا فاش شدن اطلاعات از جمله نقاط ضعف شناسایی هویت بر پایه‌ی نشانه و دانش است که روش‌های بیومتریک با آن در چالش نیستند. به عنوان تنها مشکل روش‌های بیومتریک می‌توان به دقت این روش‌ها اشاره نمود که زمینه‌ی وسیعی از تحقیقات را به خود اختصاص داده است. در میان روش‌های بیومتریک شناسایی هویت، شناسایی چهره به دلیل دخالت کمتر انسان، نسبت به سایر روش‌ها بیشتر مورد توجه می‌باشد. این ویژگی شناسایی چهره باعث شده تا بتوان در موارد امنیتی و تعقیب افراد از آن استفاده نمود. هر سیستم شناسایی چهره را می‌توان متشکل از دو بخش دانست.

❖ سیستم استخراج چهره^۱ از تصویر: این بخش از سیستم، چهره(ها) را در تصویر مکان‌یابی و از پس‌زمینه جدا می‌کند.

❖ سیستم شناسایی چهره^۲: این بخش از سیستم، چهره(ها)ی بدست آمده از مرحله‌ی قبل را با پایگاه داده‌ی خود مقایسه نموده و در صورت تشخیص چهره نتیجه را اعلام می‌کند.

۱-۲- اهمیت وجود سیستم استخراج چهره

اکثر روش‌های شناسایی چهره، فرض چهره‌ی پرسنلی^۳ را دارا می‌باشند [۱]. اما در محیط‌های واقعی به جهت پویایی چهره، هنگام پدیدار شدن در تصویر و همچنین شرایط محیطی مختلف این فرض برقرار نخواهد ماند. شکل (۱-۱) را در نظر بگیرید.



شکل ۱-۱. تصاویر معمول در آموزش و آزمایش روش‌های شناسایی چهره. همان‌طور که مشخص است این تصاویر شبه پرسنلی با شدت روشنایی یکنواخت هستند. تغییرات در حالت چهره نیز در این تصاویر کم می‌باشد.

شکل (۱-۱) دارای تصاویری است که به صورت عمومی در تحقیقات مربوط به شناسایی چهره استفاده می‌شوند. اما در دنیای واقعی، معمولاً تصویر چهره در یک پس‌زمینه‌ی پیچیده با شرایط محیطی مختلف ظاهر می‌شود. شکل (۲-۱) نمونه‌ای از تصاویر در دنیای واقعی را نشان می‌دهد.

^۱ Face Detection

^۲ Face Recognition

^۳ در این پایان‌نامه، تصاویر پرسنلی به تصاویری گفته می‌شود که در آن چهره، به صورت روبه‌رو نسبت به دوربین فیلم‌بردای قرار گرفته است، طوری که تمام بخش‌های آن در تصویر دیده شود.



شکل ۱-۲. تعدادی تصویر چهره در دنیای واقعی. این چهره‌ها در پس زمینه‌ای پیچیده و با حالت‌های مختلف چهره به همراه انسدادهایی همچون عینک پدیدار شده‌اند.

شکل (۱-۲) اهمیت وجود سیستم استخراج چهره را نمایان می‌سازد. ممکن است بخش شناسایی چهره در مواجهه با تصاویر شکل (۱-۲) دچار خطا شود. بنابراین نیاز به روشی است که به عنوان پیش‌پردازش، قادر باشد تصویر چهره را از تصویر پس زمینه برای بخش شناسایی جدا کند [۲].

۱-۳ – کاربردهای سیستم‌های استخراج چهره

به غیر از شناسایی چهره، استخراج چهره در زمینه‌های دیگری نیز کاربرد دارد که از جمله‌ی آن‌ها می‌توان به موارد زیر اشاره نمود.

۱. بازیابی تصویر بر اساس محتوی^۱

۲. به عنوان بخشی از سیستم‌های تعقیب ویدئویی^۲

^۱ Content-based image retrieval

^۲ Video tracking systems

۱-۴- چالش‌های سیستم‌های شناسایی چهره

از جمله مواردی که سیستم‌های شناسایی چهره با آن در چالش می‌باشند، می‌توان به موارد زیر اشاره نمود.

- ❖ زاویه‌ی چهره^۱: زاویه‌ی چهره نسبت به دوربین باعث مخفی ماندن بخشی از چهره در تصویر می‌شود که بر نرخ شناسایی تاثیر بسزایی خواهد داشت.
- ❖ شدت روشنایی^۲: شدت روشنایی زیاد یا کم ممکن است بر روی چهره سایه افکنده و ویژگی‌هایی تصویر چهره را پنهان کند.
- ❖ حالت‌های مختلف چهره^۳: روش‌های شناسایی، از تصاویر چهره افراد با حالت عادی استفاده می‌نمایند. حالاتی مانند خنده و اخم ممکن است این سیستم‌ها را دچار اشتباه نماید.
- ❖ تغییر مقیاس^۴: این مشکل در زمان تصویربرداری از فاصله‌ی دوربین و فرد ناشی می‌شود. فاصله‌ی زیاد از دوربین باعث وضوح پایین چهره‌ی استخراج شده می‌شود و نرخ شناسایی را کاهش خواهد داد.
- ❖ تاری تصویر^۵: در تصاویر ویدئویی، هنگامی که فرد حرکت ناگهانی انجام دهد ممکن است تصویر تار شود.

^۱ Pose variation

^۲ Illumination variation

^۳ Expression variation

^۴ Scale variation

^۵ Motion blur

❖ انسداد^۱: هر نوع پوششی که بر روی چهره قرار بگیرد (مانند ماسک، ریش) به علت پوشاندن

ویژگی‌های چهره، روند شناسایی چهره را با مشکل مواجه می‌کند.

❖ وضوح پایین تصویر^۲: در تصاویر ویدئویی به جهت محدودیت در تکنولوژی دوربین‌های فیلم-

برداری، هزینه و فضای ذخیره‌سازی، تصاویر ویدئویی با کیفیت پایین اخذ می‌شوند.

❖ بلادرنگ بودن: بلادرنگ بودن سیستم یکی از پارامترهایی است که به دلیل حساسیت موارد

امنیتی، اخیراً بسیار مورد توجه قرار گرفته است [۲].

۱-۵- مدل‌سازی چهره

قبلاً ذکر شد که بعضی از شرایط محیطی را حین استخراج چهره نمی‌توان کنترل نمود به خصوص زمانی که این پردازش بر روی تصاویر ویدئوی مد نظر باشد. شدت روشنایی‌های غیریک‌نواخت و ایجاد زاویه بین تصویر چهره و دوربین از این موارد می‌باشند. در این بخش تمرکز بر روی زاویه‌ی چهره خواهیم داشت. این مشکل در اکثر سیستم‌هایی که در درگاه‌های ورودی به کار بسته شده‌اند وجود دارد و باعث پایین آمدن کارایی این سیستم‌ها می‌شود. اخیراً روش‌هایی برای مدل‌سازی چهره، تصاویری ویدئویی کالیبره شده و دارای اطلاعات عمق (که به تصاویر استریو^۳ معروف‌اند) را به کار گرفته‌اند. این روش‌ها با استفاده از مدل‌سازی سه‌بعدی، سعی بر ساخت مدل روبه‌روی چهره نموده‌اند. برخلاف عملکرد مناسبی که این سیستم‌ها در مدل‌سازی چهره دارند، دارای پیچیدگی بالایی از محاسبات بوده و معمولاً قابلیت بلادرنگ بودن را ندارند. همچنین هزینه‌ی بالای تهیه‌ی دوربین‌های کینکت^۴ نیز از عوامل بازدارنده برای استفاده از

^۱ Occlusion

^۲ Low resolution

^۳ Stereo

^۴ دوربین‌های کینکت قادر به تولید تصاویر دارای اطلاعات عمق بوده و قابلیت کالیبره شدن را دارند.

این روش‌ها در دنیای واقعی است. در این پایان‌نامه قصد بر آن داریم تا روشی را ارائه دهیم که قابلیت مدل‌سازی تصویر چهره از تصاویر زاویه‌دار دوبعدی را دارا می‌باشد. برای نیل به این هدف الگوریتمی پیشنهاد داده و پس از پردازش تصاویر زاویه‌دار چهره با آن، نتایج بدست آمده را با استفاده از بهینه‌سازی-های برپایه‌ی بازنمایی تنک و تجزیه‌ی مرتبه پایین تراز می‌نماییم. به همین دلیل ابتدا تعریفی از بازنمایی تنک^۱ و یادگیری واژه‌نامه^۲ خواهیم داشت.

۱-۶- بازنمایی تنک

در بازنمایی تنک یک سیگنال، با نمونه‌های سیگنال داده شده Y و ماتریس پایه D (واژه‌نامه) به دنبال یافتن فضای داده‌ی x هستیم که اکثر مقادیر آن صفر باشد. با فرض اینکه تبدیل فضا برای داده، خطی باشد در این صورت رابطه‌ی (۱-۱) را خواهیم داشت:

$$Y = Dx \quad \|y - Dx\| \leq \varepsilon \quad \forall i, \|x_i\|_0 \leq T \quad (1-1)$$

در رابطه‌ی (۱-۱)، x ماتریس ضرایب بازسازی x_i ، ستون i ام ماتریس ضرایب $\| \cdot \|_0$ ، تابع نرم صفر، T مقدار ثابت از پیش معین شده برای درجه‌ی تنک بودن، Y نمونه‌های سیگنال و D واژه‌نامه است [۳]. در بازنمایی تنک فرض بر این است که مجموعه‌ی داده (Y) و همچنین ماتریس پایه (D) داده شده‌اند و تنها مجهول رابطه‌ی (۱-۱) ماتریس ضرایب تنک (x) هستند.

۱-۷- یادگیری واژه‌نامه

اگر نمونه‌های سیگنال به صورت $Y = [y_1, \dots, y_N]$ باشند، پیدا کردن یک ماتریس واژه‌نامه D با تعداد ستون K که $K \ll N$ و هر y_i از ترکیب خطی ستون‌های D بدست آید، مساله‌ی یادگیری واژه‌نامه نام دارد

^۱ Sparse representation

^۲ Dictionary

[۳]. در زمینه‌های تحقیقاتی پردازش تصویر، روش‌های بازنمایی تنک و یادگیری واژه‌نامه، در شناسایی چهره [۴،۵] و بازسازی تصاویر با وضوح بالا^۱ [۶،۷] مورد استفاده قرار می‌گیرند.

اکثر الگوریتم‌های یادگیری واژه‌نامه، از بهینه‌سازی بر پایه‌ی بازنمایی تنک برای، بدست آوردن ماتریس واژه‌نامه استفاده می‌کنند. بنابراین مساله‌ی یادگیری واژه‌نامه، تفاوت چندانی با بازنمایی تنک نداشته و غالباً، روش‌های یادگیری واژه‌نامه به معادلات بهینه‌سازی توامان بر روی ضرایب تنک و ماتریس پایه تعریف می‌شوند. از جمله الگوریتم‌های بازنمایی تنک می‌توان به، تبدیل فوریه، تبدیل موجک، پیگیری انطباق^۲ و پیگیری انطباق متعامد^۳ اشاره نمود [۳]. از جمله روش‌های یادگیری واژه‌نامه نیز می‌توان به KSVD^۴ و روش‌های مبتنی بر تجزیه مرتبه پایین اشاره نمود^۵ [۴].

۱-۶- تعریف مساله

فرض کنیم که تعدادی تصویر متعلق به چهره یک شخص با زاویه در راستاهای مختلف نسبت به دوربین داریم. این تصاویر از فریم‌های یک فایل ویدئویی که توسط دوربین نصب شده در درگاه ورودی، ضبط شده است، بدست آمده‌اند. در این تصاویر کنترلی بر روی زاویه دید شخص وجود ندارد. همچنین فرض کنیم پس از تهیه‌ی این تصاویر، چهره‌ی شخص از آن‌ها استخراج شده است. شکل (۱-۳) فضای مساله و شیوه-ی اخذ تصاویر، به همراه تعدادی از تصاویر بدست آمده از فایل ویدئو را نشان می‌دهد. زاویه چهره را با جفت $(\pm y^\circ \pm x^\circ)$ نشان می‌دهیم که در آن x و y به ترتیب میزان چرخش چهره در راستای افقی و عمودی می‌باشند. چرخش به راست و پایین چهره در راستای محور مختصات، منفی در نظر گرفته می-
شود.

^۱ Image super resolution

^۲ Matching Pursuit

^۳ Orthogonal Matching Pursuit

^۴ K-means singular value decomposition

^۵ Low rank decomposition

شود. زاویه تصویر چهره را با $I^{\pm y^{\circ} \pm x^{\circ}}$ نشان می‌دهیم. مجموعه‌ای از تصاویر با وضوح یکسان به صورت $C = \{i \in I^{\pm y^{\circ} \pm x^{\circ}}, x \text{ or } y \neq 0\}$ به عنوان ورودی داریم. می‌خواهیم با تصاویر مجموعه‌ی C ، تصویری از روبه‌روی چهره‌ی شخص را مدل کنیم. به عبارت دیگر، می‌خواهیم با استفاده از مجموعه‌ای از تصاویر (۱-۳-ب)، تصویر چهره‌ای را مدل سازیم تا بیشترین شباهت به تصویر شکل (۱-۳-پ) را داشته باشد.



شکل ۱-۳. فضای مساله و تصاویر زاویه‌دار. الف: فضای مساله. ب: تصاویر بدست آمده از دوربین ویدئویی. پ: تصویر روبه‌رو با زاویه‌ی (+++) که در تصاویر ب جای آن خالی گذاشته شده است.

۱-۷- اهداف پایان‌نامه

در این پایان‌نامه، هدف از ارائه‌ی روش پیشنهادی، ساخت مدلی از تصویر روبه‌روی چهره با استفاده از تصاویر ویدئویی است که از دوربین تصویربرداری نصب شده در درگاه ورودی استخراج می‌شوند. در روش پیشنهادی فرض بر این است که در ابتدا با یک روش استخراج چهره، تصاویر چهره از پس‌زمینه جدا می‌شوند. از جمله نیازمندی‌های روش پیشنهادی، دقت بالا در مدل‌سازی و همچنین بلادرنگ بودن است. هدف این است که با ارائه‌ی روش پیشنهادی، بر دقت سیستم شناسایی چهره افزوده شود.

۸-۱- پایگاه داده‌ی مورد استفاده

در این پایان‌نامه از پایگاه داده‌ی PRIMA استفاده شده است [۸]. این مجموعه‌ی داده شامل ۲۷۹۰ تصویر چهره‌ی یک‌چشمه^۱ از پانزده نفر است. از هر شخص دو مجموعه‌ی ۹۳ تایی تصویر با تغییرات زاویه‌ای در دو راستای X و Y در محدوده‌ی $(-90, +90)$ موجود می‌باشد. در هر دو راستا گام تغییر زاویه ۱۵ درجه، است. از دو مجموعه‌ی مذکور، یک مجموعه حاوی تصاویر ساده‌ای است که فقط دارای زاویه نسبت به دوربین بوده و مجموعه‌ی دیگر شامل مواردی همچون حالات مختلف، انسداد یا شدت روشنایی-های متفاوت است. در این پایگاه داده هر تصویر دارای برچسب زاویه است که البته کاملاً دقیق نمی‌باشد و دارای مقداری خطاست [۸]. شکل (۱-۴) نمونه‌ای از تصاویر پایگاه داده‌ی PRIMA را نمایش می‌دهد.



شکل ۱-۴. نمونه‌ای از تصاویر پایگاه داده‌ی PRIMA.

^۱ Monocular

۱-۹- ساختار پایان نامه

این پایان نامه در پنج فصل تدوین شده که ساختار آن به شرح زیر می باشد. در فصل دوم، به روش های استخراج چهره و نقاط ضعف و قوت آن ها در برابر شرایط محیطی، پرداخته می شود. با توجه به اینکه در سیستم های جدید شناسایی چهره، بخشی دیگر بین تشخیص چهره و شناسایی آن برای بهبود تصویر چهره افزوده شده است، به صورت خلاصه، به چند روش از روش های موجود در این زمینه نیز اشاره خواهد شد. در فصل سوم روش پیشنهادی و الگوریتم های پیش نیاز آن مطرح خواهند شد. فصل چهارم به آزمایشات کمی و کیفی روش پیشنهادی پرداخته و در آخر، در فصل پنجم، نتیجه گیری و کارهای آتی نیز مطرح می شود.

فصل دوم

مروری بر روش های استخراج چهره و نمال سازی

مقدمه

در این فصل، ابتدا به صورت مختصر به روش‌های استخراج چهره پرداخته می‌شود. برای اخذ نتایج بهتر در بخش شناسایی، معمولاً پیش‌پردازش و پس‌پردازش‌هایی به ترتیب بر روی تصویر ورودی و خروجی سیستم استخراج چهره صورت می‌گیرد. این روش‌ها نرمال‌سازی تصاویر چهره نامیده می‌شوند و می‌توان آن‌ها را به دو گروه تفکیک نمود [۹]:

❖ روش‌های بهبود کیفیت تصویر: این روش‌ها جهت مقابله با مشکلاتی همچون تاری یا وضوح پایین

تصویر چهره ارائه شده و با تکنیک‌های به‌سازی تصویر، سعی بر بهبود آن دارند.

❖ روش‌های نرمال‌سازی: این روش‌ها جهت ارائه تصویری واحد از مجموعه تصاویر چهره شخص، به

منظور رفع محدودیت‌هایی از جمله شدت روشنایی‌های غیریک‌نواخت، وجود انسداد و حالات

مختلف چهره ارائه می‌شوند. فرایندهایی از این قبیل موجب ارائه مدلی واحد می‌شوند و فرضیات

متعددی را که باعث پیچیدگی روش‌های شناسایی می‌شوند از میان می‌برند.

در ادامه فصل، پس از معرفی روش‌های استخراج چهره، کلیات تعدادی از روش‌های نرمال‌سازی مطرح خواهد شد.

۲-۱- مروری بر روش‌های استخراج چهره

پویایی چهره‌ی انسان و عوامل محیطی باعث می‌شوند چهره به شیوه‌های گوناگونی در تصویر ظاهر شود و

به همین دلیل روش‌های مختلفی برای استخراج چهره از تصویر ارائه شده‌اند. این روش‌ها به دو دسته‌ی ۱.

روش‌های مبتنی بر ویژگی^۱ و ۲. روش‌های مبتنی بر تصویر^۱ بخش‌بندی می‌شوند [۱]. روش‌های دسته‌ی

^۱ Feature base approaches

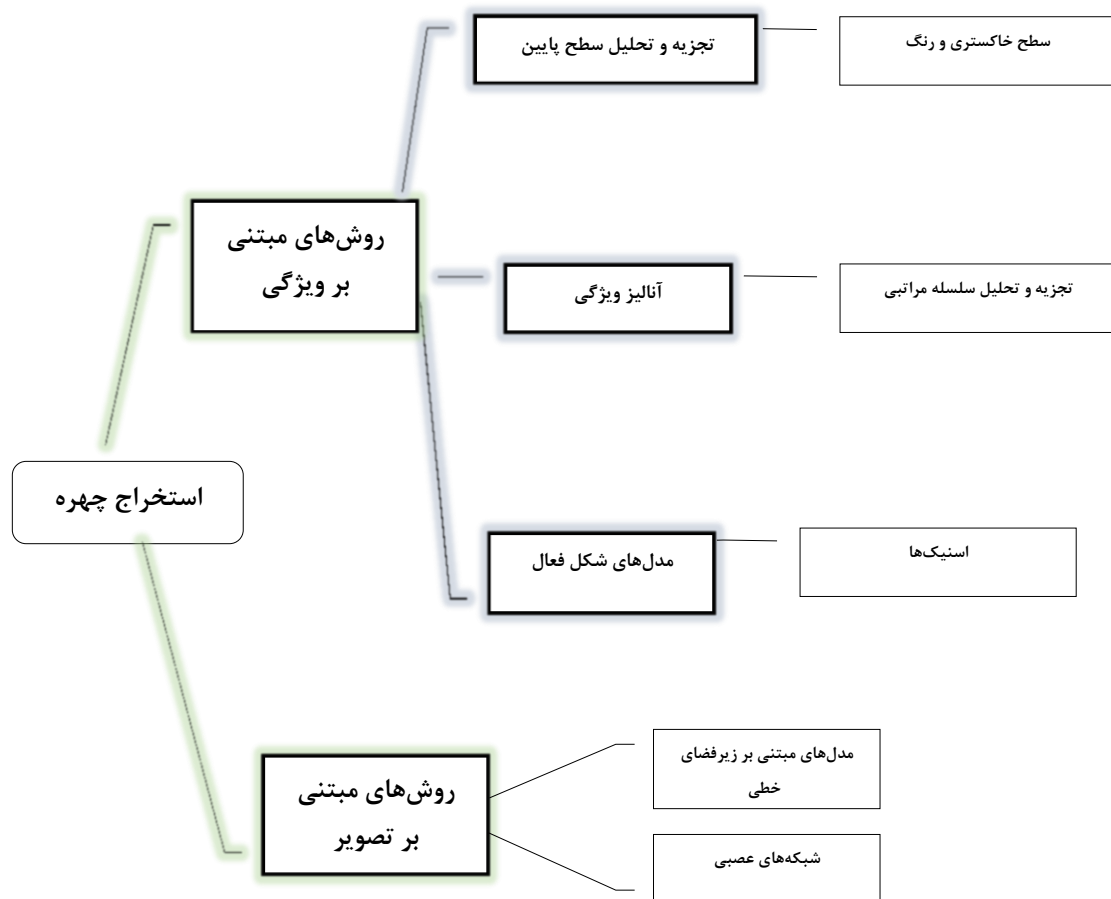
اول از دانش چهره به صورت صریح استفاده می‌کنند. از جمله‌ی این دانش‌ها می‌توان به ویژگی‌های سطح پایین^۲ همچون اطلاعات سطح خاکستری تصویر، رنگ و لبه‌ها اشاره نمود. روش‌های متعلق به دسته‌ی دوم، به استخراج چهره به عنوان یک مسأله‌ی شناسایی الگو می‌نگرند. این روش‌ها با استفاده از یادگیری، الگوی چهره را آموخته و با آن به استخراج چهره می‌پردازند. در دو دهه‌ی اخیر تحقیقات زیادی در زمینه-ی استخراج چهره صورت گرفته که هر یک در صدد پاسخ به یک یا تعدادی از چالش‌های مطرح شده در این حوزه (چالش‌های مطرح شده در بخش ۱-۴) هستند. دیاگرام شکل (۲-۱) شمایی کلی از روش‌های مبتنی بر ویژگی، روش‌های مبتنی بر تصویر و انواع آن‌ها را نشان می‌دهد [۱].

در این بخش سعی نموده‌ایم تا این دو دسته از روش‌ها را به صورت مختصر شرح دهیم و مثالی معروف برای هر یک از آن‌ها بیاوریم. نقاط ضعف هر روش در بخش مربوط به آن مورد بحث قرار می‌گیرد. هر کدام از روش‌های پیشنهادی جدید را می‌توان در یکی از دسته‌های مطرح شده قرار داد. چون روش‌های مطرح شده در زمینه استخراج بر پایه‌ی مکاشفه‌هایی^۳ بنا شده‌اند، طراحی سیستمی یک‌پارچه که مقاوم در برابر تمامی مشکلات استخراج چهره باشد نیز خود چالشی بزرگ محسوب می‌شود.

^۱ Image base approaches

^۲ چون بار معنایی کمی به لحاظ بصری برای انسان دارند، ویژگی سطح پایین نامیده می‌شوند.

^۳ Heuristic



شکل ۱-۲. دسته‌بندی روش‌های استخراج چهره بر اساس نوع عملکرد [۱].

۲-۱-۱- روش‌های بر پایه‌ی ویژگی

این روش‌ها را می‌توان در سه زیرمجموعه طبقه‌بندی نمود.

۱. آنالیز سطح پایین^۱: روش‌های آنالیز سطح پایین بر پایه‌ی قطعه‌بندی چهره یا ویژگی‌های آن (مانند بینی، چشم‌ها، دهان) با استفاده از اطلاعات پیکسل هستند.

۲. آنالیز ویژگی^۱: این روش‌ها از ویژگی‌های پیکسل و دانشی پیشین از هندسه‌ی چهره‌ی انسان^۲ برای استخراج چهره استفاده می‌نمایند.

^۱ Low-level analysis

رنگی‌ای استفاده می‌کنند که بتوانند به صورت مستقل، شدت روشنایی را دستکاری کنند. یکی از فضا‌های مناسب برای این منظور، فضای رنگی HSV است [۱۱].

۲-۱-۱-۳- معیارهای عام

روش‌های بر پایه‌ی معیارهای عام^۱، از سیستم بینایی انسان الگوبرداری می‌کنند. در روش ارائه شده [۱۲]، Resifeld و Yeshuruu یک عملگر متقارن را معرفی نمودند که بر مبنای پیکسل‌های لبه عمل می‌نماید. عملگری که آن‌ها معرفی نمودند از تقارن چشم‌ها و دهان استفاده نموده و برای پیکسل‌های این نواحی در چهره، مقدار بزرگتری نسبت پیکسل‌های دیگر بخش‌های چهره تولید می‌نماید. این معیار تقارن با رابطه‌ی (۱-۲) برای هر پیکسل تعریف می‌شود که به مکان هر پیکسل بر اساس پیکسل‌های محاط بر آن مقداری نسبت می‌دهد.

$$M_{\sigma}(p) = \sum_{(i,j) \in \Gamma(p)} C(i,j), \quad (1-2)$$

در رابطه‌ی (۱-۲)، $C(i,j)$ سهم پیکسل‌های محاط کننده‌ی پیکسل p در مجموعه‌ای از پیکسل‌های $\Gamma(p)$ است. C و Γ به صورت روابط (۲-۲) و (۳-۲) تعریف می‌شوند.

$$\Gamma(p) = \left[(i,j) \mid \frac{p_i + p_j}{2} = p \right] \quad (2-2)$$

$$C(i,j) = D_{\sigma}(i,j)P(i,j)r_i r_j, \quad (3-2)$$

که $D(i,j)$ تابع وزن فاصله و $P(i,j)$ تابع وزن فاز و r_i و r_j با روابط (۴-۲) تعریف می‌شوند.

$$D_{\sigma}(i,j) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{\|p_i - p_j\|}{2\sigma}}$$

$$P(i,j) = (1 - \cos(\theta_i + \theta_j - 2\alpha_{ij}))(1 - \cos(\theta_i - \theta_j))$$

^۱ Generalized Measures

$$r_k = \log(1 + \|\nabla_{p_k}\|) \quad (۴-۲)$$

$$\theta_k = \arctan \left[\frac{\frac{\partial}{\partial_y} p_k}{\frac{\partial}{\partial_x} p_k} \right]$$

$P_k = (x_k, y_k), k = 1, \dots, K$ نشان‌دهنده‌ی پیکسل، ∇_{P_k} مقدار گرادیان شدت روشنایی در نقطه‌ی P_k و $\alpha_{i,j}$ زاویه‌ی خط عبور کننده از P_i و P_j و محور افقی در خلاف جهت عقربه‌های ساعت است. شکل (۲-۲) مثالی از این روش را پس از محاسبه‌ی $M_\sigma(i, j)$ از تصویر گرادیان چهره نشان می‌دهد.



شکل ۲-۲. تصویر اصلی و M_σ بدست آمده از تصویر اصلی بر اساس روش **Resifeld** و همکارانش [۱۲].
از تصویر حاصل مشخص است که این معیار به خوبی مکان چشم‌ها و دهان را نشان می‌دهد. با استفاده از این معیار **Resifeld** و همکارانش به موفقیت استخراج ۹۵٪ ویژگی‌های چهره (چشم‌ها و دهان) در یک پایگاه داده چهره به همراه پس‌زمینه‌ی پیچیده دست یافتند.

۲-۱-۲- تجزیه و تحلیل ویژگی

در تکنیک‌های بر پایه‌ی تجزیه و تحلیل ویژگی^۱، دانشی در رابطه با هندسه‌ی چهره، جهت قطعیت بخشیدن به نتیجه‌ی استخراج چهره و بخش‌های آن به کار گرفته می‌شود. ذیلاً یک مثال از این روش توضیح داده می‌شود.

۲-۱-۲-۱- تجزیه و تحلیل بر اساس مدل‌های سلسله‌مراتبی چهره

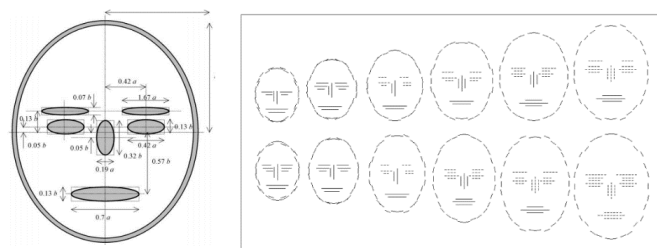
به عنوان مثالی برای روش استخراج چهره مبتنی بر مدل سلسله‌مراتبی^۲ می‌توان به روش ارائه شده در [۱۳] اشاره نمود. در این روش، تصویر ورودی، به یک تصویر جهت‌دار با استفاده از نوعی عملگر بر پایه‌ی گرادیان با اندازه‌ی پنجره‌ی کانولوشن 7×7 تبدیل می‌شود. سپس تبدیل هاف^۳ و یک مجموعه الگوی 12×12 تایی دودویی که طرح سلسله‌مراتبی چهره را می‌سازد، روی تصویر جهت‌دار اعمال می‌شود. از تبدیل هاف برای پیدا کردن اشکال بیضوی در تصویر به عنوان کاندیداهای چهره استفاده می‌شود. در نهایت این کاندیداها به مرحله‌ی تطبیق الگو^۴ جهت رد یا قبول بخش چهره داده می‌شوند. شکل (۲-۳) الگوی مورد استفاده شده در این روش، مجموعه‌ی 12×12 تایی الگو و همچنین مراحل این روش را برای استخراج چهره در یک تصویر نشان می‌دهد.

^۱ Feature Analysis

^۲ Constellation analysis

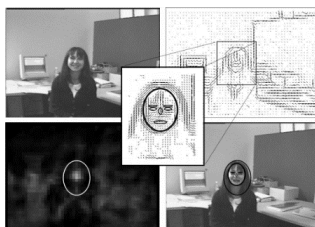
^۳ Hough transform

^۴ Template matching



الف

ب



پ

شکل ۲-۳. الف: الگوی استفاده شده برای استخراج چهره در [۱۳] به همراه نسبت‌های آن. ب: یک مجموعه دوازده تایی از این الگوها برای پیدا نمودن چهره با اندازه‌های مختلف در تصویر. پ: نحوه‌ی استخراج چهره با روش [۱۳].

۲-۱-۳- مدل‌های شکلی - فعال

مدل‌های شکلی فعال^۱ دارای کانتورهایی هستند که با استفاده از ویژگی‌های محلی چهره مانند لبه‌ها به صورت تدریجی تغییر شکل می‌دهند تا شکل چهره یا ویژگی‌های آنرا به خود بگیرند. اسنیک‌ها^۲ و الگوهای دگردیس‌پذیر^۳ از جمله روش‌های مبتنی بر مدل‌های شکلی - فعال می‌باشند [۱۴]. در این بخش به صورت خلاصه به یک مورد روش دارای اسنیک پرداخته می‌شود.

۲-۱-۳-۱- اسنیک‌ها

اسنیک به کانتورهایی گفته می‌شود که برای مشخص نمودن مرزهای چهره از آن‌ها استفاده می‌شود [۱۴]. یک اسنیک تقریباً حوالی مرز چهره قرار گرفته، سپس نزدیک لبه‌های چهره، حالت چهره را به خود

^۱ Active Shape Models

^۲ Snake

^۳ Deformable template

گرفته و قفل می‌شود. این عمل توسط کمینه‌سازی تابع انرژی E_{snake} که با رابطه‌ی (۵-۲) مشخص شده، انجام می‌شود.

$$E_{snake} = E_{internal} + E_{external} \quad (۵-۲)$$

در رابطه‌ی (۵-۲) $E_{internal}$ انرژی داخلی و $E_{external}$ انرژی خارجی هستند. انرژی داخلی نحوه‌ی تغییر شکل اسنیک را به صورت انبساطی و انقباضی مشخص می‌نماید. انرژی خارجی، اثر انرژی داخلی را خنثی نموده و از کمینه‌سازی بیش از حد اسنیک جلوگیری می‌نماید. این دو تابع انرژی بر یکدیگر تاثیر می‌گذارند تا به نقطه‌ی تعادل برسند. روش ارائه شده در [۱۴]، ابتدا با استفاده از لبه‌یابی بر اساس رنگ و شدت روشنایی و استفاده از اطلاعات رنگ پوست چهره، مکان تقریبی چهره را مشخص می‌نماید. سپس برای چهره و هر یک از بخش‌های آن (مانند چشم‌ها، ابروها، دهان و بینی) یک کانتور در نظر می‌گیرد. با استفاده از اطلاعاتی که در بخش قبل بدست آمده بود، این کانتورها در حوالی بخش‌های مذکور قرار گرفته و با کمینه‌سازی رابطه‌ی (۵-۲) کانتورها، شکل بخش‌های چهره را به خود می‌گیرند. در روش ارائه شده‌ی [۱۴] تابع انرژی داخلی، میزان هموار بودن کانتور را بر اساس رنگ پوست کنترل نموده و تابع انرژی خارجی، برای خنثی نمودن اثر آن به صورت جداگانه برای کانتورهای مختلف تعریف می‌شود. با وجود توانایی این کانتورها در استخراج چهره و بخش‌های آن، امکان قرار گرفتن آن‌ها در دیگر بخش‌های غیر چهره‌ای تصویر وجود دارد. به علاوه اسنیک‌ها قادر به استخراج ویژگی‌های غیرمحدب^۱ (با توجه به محدودیت‌های تکنیک‌های کمینه‌سازی) نیستند. شکل (۴-۲) نمونه‌ای از یک اسنیک بدست آمده در چهره‌ی شخص را با روش [۱۴] نشان می‌دهد.

^۱ ناحیه‌هایی که بتوان در داخل ناحیه‌ی آن‌ها دو نقطه گذاشت و با خطی این دو نقطه را به هم متصل نمود، به صورتی که بخشی از این خط در داخل ناحیه نباشد نواحی غیرمحدب به شمار می‌روند.



شکل ۲-۴. نمونه‌ای از یک اسنیک قفل شده بر روی چهره‌ی یک شخص با استفاده از روش [۱۴].

۲-۱-۴- مشکلاتی که روش‌های بر پایه‌ی ویژگی با آن مواجه‌اند

روش‌های مطرح شده در بخش قبل، عموماً در تصاویر دنیای واقعی کارایی محدودی دارند. همچنین بعضی از آن‌ها دارای مدل‌های پیچیده و پارامترهای زیاد هستند که باعث می‌شود نتوان در کاربردهای بلادرنگ از آن‌ها استفاده نمود و اینکه در مواجهه با تصاویر دارای تعدادی چهره، کارایی خود را از دست می‌دهند.

۲-۱-۶- مدل‌های بر پایه‌ی تصویر

برای رفع مشکلات مطرح در روش‌های قبلی، در تحقیقات اخیر مساله‌ی استخراج چهره، با استفاده از روش‌های یادگیری ماشین مدل می‌شود. با این روش حالات بیشتری از تصاویر چهره، در قالب نمونه‌های آموزشی مورد پوشش قرار می‌گیرند و بدین صورت از اطلاعات چهره به صورت ضمنی در استخراج چهره استفاده می‌شود. روش‌های یادگیری، تصاویر ورودی را به تصاویر چهره و غیر چهره دسته‌بندی می‌نمایند. برای استخراج چهره، این روش‌ها مقایسه‌ای بین تصویر ورودی و الگوی یادگرفته شده خود (الگوی چهره-ای که از نمونه‌های آموزشی یاد گرفته‌اند) انجام داده و اقدام به تصمیم‌گیری می‌نمایند. معمولاً این روش-ها یک الگوریتم پوشش تصویر توسط پنجره را به کار می‌گیرند که نوعی جست‌وجوی سراسری برای

مشخص نمودن مکان‌های چهره است. ذیلاً به روش تحلیل مؤلفه‌های^۱ اصلی (PCA) و شبکه‌های عصبی^۲ که برای استخراج چهره به کار گرفته شده‌اند پرداخته می‌شود.

۲-۱-۶-۱- روش‌های بر پایه‌ی تبدیل فضای تصویر به فضای محدودتر

روش‌های مبتنی بر تبدیل خطی فضا^۳، به تصویر به عنوان یک فضای دارای بعد می‌نگرند (مثلاً یک تصویر با ابعاد 200×200 دارای 40000 بعد است) و فرض می‌کنند چهره‌ها قابلیت نمایش در یک زیر فضای خطی با ابعاد کمتر را دارند. از جمله‌ی این روش‌ها می‌توان به تحلیل مؤلفه‌های اصلی، اشاره نمود.

۲-۱-۶-۲- تحلیل مؤلفه‌های اصلی

روش ارائه شده در [۱۵] از یک مجموعه تصاویر چهره که متعلق به اشخاص متفاوت است، با استفاده از PCA، مؤلفه‌های اصلی چهره‌ها را به دست می‌آورد. به این مؤلفه‌ها، بردارهای ویژه^۴ گفته می‌شود. بردارهای ویژه از ماتریس کوواریانس تصاویر مجموعه نسبت به هم بدست می‌آیند. تصویر هر چهره در مجموعه با ترکیب خطی وزن‌دار این بردارها ساخته می‌شود. بردارهای ویژه معمولاً چهره‌های ویژه^۵ نیز نامیده می‌شوند. بازیابی چهره اصلی با استفاده از PCA تقریبی از آن بوده و مقدار خطای باقی‌مانده در الگوریتم (اختلاف میان تصویر اصلی و تصویر ساخته شده توسط PCA) به عنوان معیار "عدم چهره بودن" در نظر گرفته می‌شود. طریقه محاسبه‌ی این معیار که "فاصله از فضای چهره"^۶ (DFFS) نامیده شد به صورت زیر می‌باشد.

^۱ Principle Component Analysis

^۲ Neural Networks

^۳ Linear Subspace Methods

^۴ Eigenvectors

^۵ Eigenfaces

^۶ Distance From Face Space

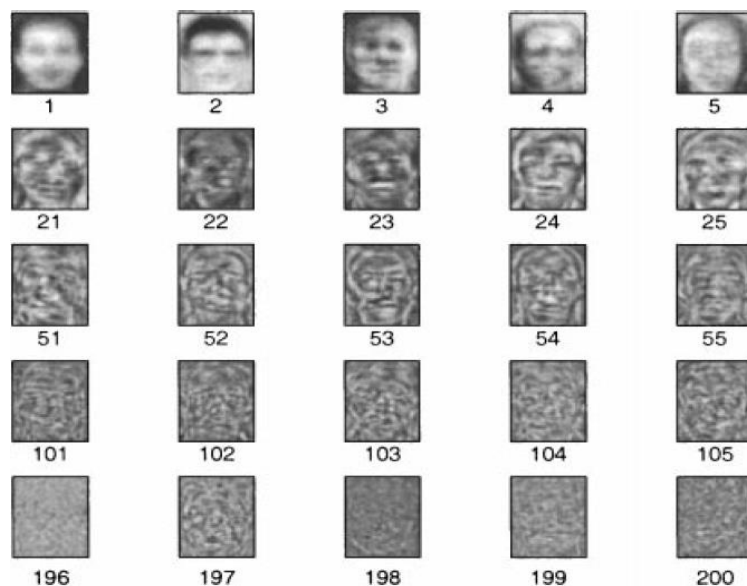
۱. با فرض در دست داشتن n تصویر چهره $\Gamma_1, \Gamma_2, \dots, \Gamma_n$ ، میانگین تصاویر با رابطه‌ی (۶-۲) محاسبه می‌شود.

$$\Psi = \frac{1}{n} \sum_{i=1}^n \Gamma_i \quad (6-2)$$

۲. سپس مطابق رابطه‌ی (۷-۲) تصویر میانگین از هر تصویر کم شده و تصویر تفریق حاصل، به بردار تبدیل می‌شود.

$$\Phi_i = \text{vec}(\Gamma_i - \Psi) \quad (7-2)$$

در رابطه‌ی (۷-۲) Φ_i تصویر تفریق است و vec عملگر تبدیل تصویر تفریق به بردار می‌باشد. با فرض $D = [\varphi_1 \varphi_2 \dots \varphi_n]$ و $C = DD^t$ ، بردارهای ویژه‌ی C ، با نشان داده شده و مؤلفه‌های اصلی D نامیده می‌شوند. اگر آن‌ها را تبدیل به تصویر نماییم، چهره‌های ویژه‌ی پایگاه داده بدست می‌آیند. به شکل (۵-۲) توجه شود.



شکل ۵-۲. نمونه‌هایی از تصاویر چهره‌های ویژه که از روی پایگاه داده‌ی ORL توسط روش [۱۵] بدست آمده است. این تصاویر ویژه بر حسب مقدار ویژه‌ی خود به صورت نزولی مرتب شده‌اند که اعداد زیر آن‌ها بیانگر جایگاه آن‌ها در ترتیب است.

۳. چهره‌های ویژه، فضایی با ابعاد کمتر از فضای تصویر را تشکیل می‌دهند که به آن فضای چهره گفته می‌شود. یک تصویر چهره Φ با رابطه‌ی (۸-۲) به فضای چهره نگاشت می‌شود.

$$\omega_k = u_k^T \Phi, \quad k = 1, \dots, m, \quad (8-2)$$

$m < n$ تعداد مؤلفه‌های اصلی است و T به معنی ترانهاده بردار می‌باشد. همان‌طور که از تصاویر چهره‌ی ویژه‌ی شکل (۵-۲) مشخص است، تصاویر ویژه با مقدار ویژه‌ی بزرگتر حاوی اطلاعات ارزشمند برای بازسازی چهره می‌باشند.

۴. تصویر چهره، Φ ، با رابطه‌ی زیر از تصاویر ویژه قابل بازیابی است.

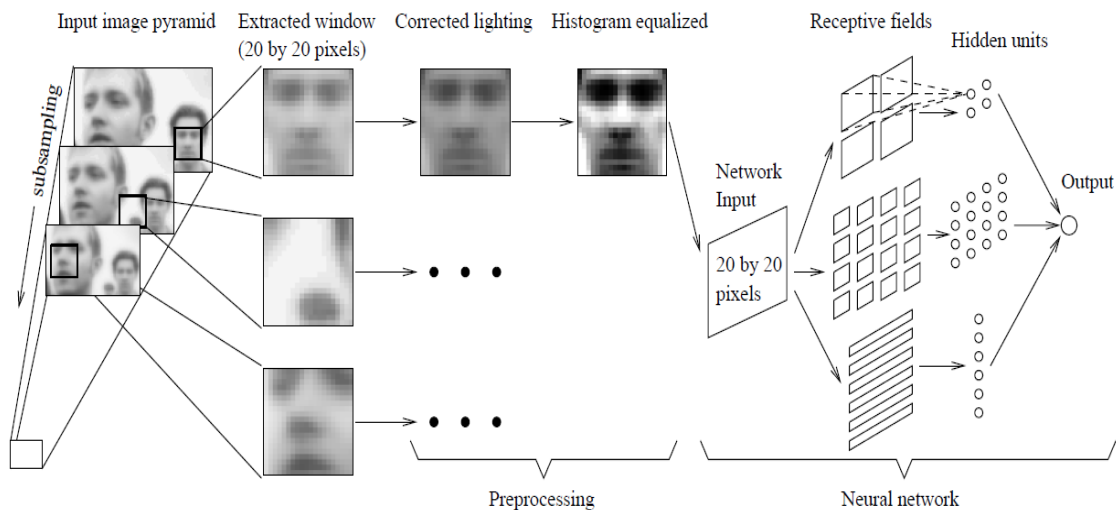
$$\Phi_r = \sum_{k=1}^m \omega_k u_k \quad (9-2)$$

۵. خطای بازسازی، (DFFS) با رابطه‌ی $\|\Phi - \Phi_r\|^2 = \varepsilon$ محاسبه می‌شود.

۲-۱-۶-۳- شبکه‌های عصبی

از جمله تکنیک‌های شناسایی الگو که در زمینه‌ی استخراج چهره مورد استفاده قرار می‌گیرد شبکه‌های عصبی می‌باشد. Rowley و همکارانش در [۱۶] روش استخراج چهره‌ی مبتنی بر شبکه‌ی عصبی ارائه دادند که بر روی پایگاه داده‌ی بزرگ آموزش دیده بود. در روش آن‌ها برای استخراج چهره، یک شبکه‌ی عصبی پرسپترون چند لایه^۱ به کار گرفته شده است که معماری آن در شکل (۶-۲) نمایش داده شده است.

^۱ Multi Layer Perceptron



شکل ۲-۶. شبکه عصبی پایه جهت استخراج چهره، ارائه شده توسط Rowley و همکارانش [۱۶].

این شبکه‌ی عصبی برای پوشش نمودن یک تصویر 20×20 (مجموعاً ۴۰۰ ورودی) طراحی شده است. شبکه یک لایه‌ی مخفی دارای ۲۶ نرون دارد. چهار نرون ناحیه‌های 10×10 ، ۱۶ نرون ناحیه‌های 5×5 و شش نرون ناحیه‌های 20×5 را که با هم هم‌پوشانی دارند پوشش می‌نماید. تحت پیش‌پردازشی، شدت روشنایی تصویر تصحیح شده و پس از آن با برابرسازی هیستوگرام، تصویر بهبود کیفیت داده می‌شود. از جمله مشکلات این روش، تشخیص چهره در بخش‌هایی است که با هم هم‌پوشانی دارند. Rowley و همکارانش به این مشکل توسط مکاشفه‌های زیر پاسخ دادند.

۱. آستانه‌گذاری: تعداد شناسایی‌های چهره در یک همسایگی کوچک شمرده شده و اگر بیشتر از

یک مقدار خاص (مثلاً پنج تا) بود، یک چهره در آن مکان وجود دارد.

۲. حذف هم‌پوشانی‌ها: وقتی که یک ناحیه به عنوان چهره تشخیص داده شد، نواحی که چهره

تشخیص داده شده‌اند و با این ناحیه هم‌پوشانی دارند تشخیص‌های "مثبت کاذب" بوده و حذف

می‌شوند.

۲-۱-۷- مشکلاتی که روش‌های مبتنی بر تصویر با آن مواجه‌اند

بزرگترین چالشی که روش‌های مبتنی بر تصویر با آن مواجه‌اند این است که ساخت پایگاه داده‌ی مناسب برای آن‌ها مشکل می‌باشد. چون نحوه‌ی عملکرد این روش‌ها بر پایه‌ی دسته‌بندی تصاویر به کلاس چهره و غیر چهره است، ساخت کلاس چهره برای نمونه‌های آموزشی ساده است اما ساخت کلاس غیر چهره دشوار می‌باشد. همچنین ظرفیت یادگیری این روش‌ها و اینکه چه تعداد الگو را می‌توانند یاد بگیرند نیز به سادگی قابل اندازه‌گیری نیست. در آخر برای کاربردهای بلادرنگ نیاز به توسعه دارند.

۲-۲- روش‌های نرمال‌سازی

همان‌طور که در بخش مقدمه‌ی این فصل اشاره شد، روش‌های نرمال‌سازی برای ارائه‌ی تصویری واحد از چهره‌ی شخص، به جهت ساده‌سازی فرضیات شناسایی مطرح می‌شوند. در این بخش تعدادی از این روش‌ها به صورت مختصر معرفی می‌شوند و در فصل بعد، دو روش نرمال‌سازی که مورد استفاده‌ی روش پیشنهادی این پایان‌نامه می‌باشد، با جزئیات توضیح داده خواهد شد.

۲-۲-۱- رفع انسداد از تصاویر چهره

روش ارائه شده در [۱۷] بعد از استخراج چهره، با استفاده از باقیمانده بازسازی^۱ در کلاس‌بندی بر پایه‌ی بازنمایی تنک^۲، ماسک دودویی برای از بین بردن تاثیر انسداد در تصویر چهره می‌سازد. بعد از رفع انسداد، شناسایی چهره صورت می‌پذیرد. خلاصه‌ی مراحل این روش به صورت زیر است.

^۱ Reconstruction Residual

^۲ Sparse Representation Based Classification

گام اول: دریافت ورودی، ماتریس نمونه‌های آموزشی $D \in \mathbb{R}^{m \times n}$ و یک نمونه چهره‌ی دارای انسداد که قرار است انسداد آن رفع شود، $y \in \mathbb{R}^m$. در این مدل ماتریس آموزشی به صورت $D = [D_1, D_2, \dots, D_C]$ بوده که دارای C کلاس آموزشی است و هر یک از D_i ها، نمونه‌های آموزشی آن کلاس را در بر دارد. ماتریس D همان ماتریس واژه‌نامه است.

گام دوم: نرمال‌سازی ستون‌های ماتریس D و y نسبت به نرم اقلیدسی.

گام سوم: استفاده از معادلات (۱۰-۲) و (۱۱-۲) برای برآورد پیکسل‌های دارای انسداد در تصویر آزمایشی و ساخت ماسک برای آن.

$$\hat{X} = \arg \min_x \{\|y - DX\|_2^2 + \lambda \|X\|_2^2\} \quad (10-2)$$

$$m = |y - D\hat{X}| < \sigma \quad (11-2)$$

در این روابط λ پارامتر تعمیم بازسازی بوده و مقداردهی آن با مقادیر بزرگ، انسدادها را در تصویر ماسک نمایان می‌سازد. X ماتریس ضرایب بازسازی کلاس‌های مختلف و $\hat{X}$ ماتریس ضرایب بازسازی چهره‌ی فرد مورد نظر با استفاده از واژه‌نامه‌ی D است. $y - D\hat{X}$ باقی‌مانده‌ی بازسازی و σ مقدار آستانه است که با آزمایش بدست می‌آید. پیکسل‌هایی که در رابطه‌ی (۱۱-۲) صدق کنند، به عنوان پیکسل‌های بدون انسداد شناخته می‌شوند.

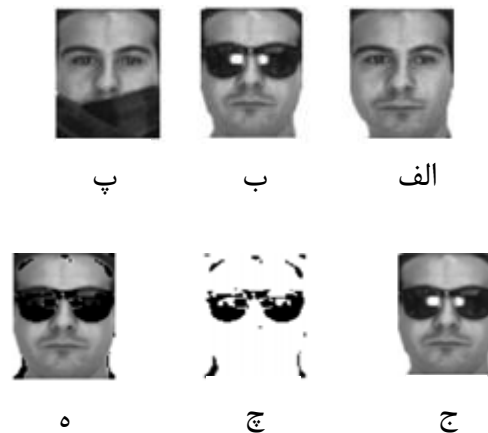
گام چهارم: پاسخ به مساله حداقل مربعات رابطه‌ی (۱۲-۲) با ماسک باینری بدست آمده از رابطه‌ی (۱۱-۲).

$$\hat{X} = \arg \min_x \{\frac{1}{2} \|\text{dig}(m)(y - DX)\|_2^2 + \lambda \|X\|_2^2\} \quad (12-2)$$

گام پنجم: رفع انسداد از تصویر آزمایشی با پاسخ به معادله‌ی (۱۳-۲).

$$\text{identity}(y) = \arg \min_i \{\frac{1}{2} \|\text{dig}(m)(y - D_i \hat{X}_i)\|_2 / \|\hat{X}_i\|_2\} \quad (13-2)$$

شکل (۷-۲) نمونه‌های آموزشی و آزمایشی را به همراه ساخت یک ماسک و قرار دادن آن بر روی چهره‌ی شخص توسط این روش نمایش می‌دهد.



شکل ۷-۲. داده‌های آموزشی، آزمایشی و شیوه‌ی عملکرد روش [۱۷]. الف: یک نمونه‌ی آزمایشی از پایگاه داده. ب و پ: دو نمونه‌ی دارای انسداد. ج: یک نمونه از تصویر چهره‌ی دارای انسداد که برای آن ماسک دودویی ساخته شده است. چ: ماسک دودویی تصویر ج. ه: ماسک دودویی که بر روی تصویر قرار گرفته است.

۲-۲-۲- تصحیح شدت روشنایی

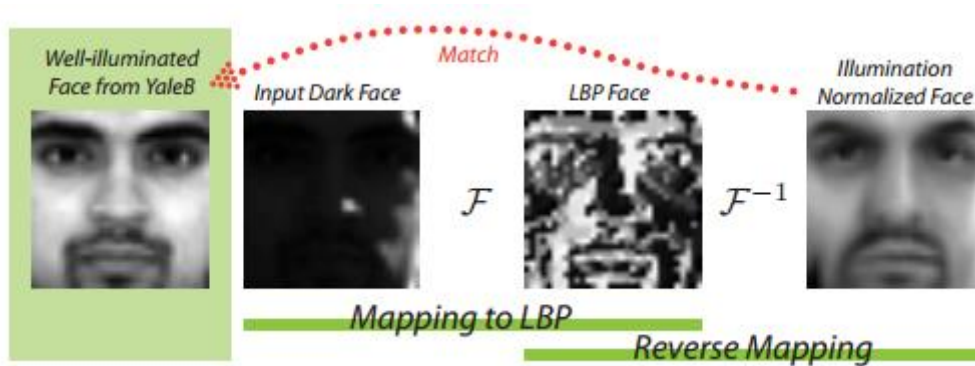
در روش پیشنهادی ارائه شده در [۱۸]، برای تصحیح شدت روشنایی، ابتدا تصاویر چهره به حوزه‌ی الگوی دودویی محلی^۱ (LBP) نگاشت می‌شوند که باعث از بین بردن مقدار زیادی از تغییرات شدت روشنایی در تصویر چهره‌ی تیره خواهد شد و بدین صورت تصویر چهره بهبود داده می‌شود. در مرحله‌ی بعد، با استفاده از تصویر LBP میانی، نرمال‌سازی نهایی صورت می‌پذیرد. بدین صورت که در طی یک نگاشت معکوس از فضای LBP به فضای پیکسل، که از طریق یادگیری واژه‌نامه توامان^۲ میان دو حوزه‌ی مذکور بدست می‌آید، نرمال‌سازی انجام می‌شود. تابع هدف برای این پژوهش با رابطه‌ی (۲-۱۴) تعریف می‌شود.

^۱ Local Binary Pattern

^۲ Joint Dictionary Learning

$$\underset{D, X, D_{LBP}}{\operatorname{argmin}} \|Y - DX\|_F^2 + \|Y_{LBP} - D_{LBP}X\|_F^2 \quad (14-2)$$

در رابطه‌ی (۱۷-۲) D و D_{LBP} به ترتیب واژه‌نامه‌های حوزه‌ی پیکسل و حوزه‌ی LBP تصویر چهره هستند. همین‌طور Y و Y_{LBP} نمونه‌های آموزشی دو حوزه می‌باشند و X ضرایب تنک مشترک میان دو حوزه است. برای این بهینه‌سازی، از روش یادگیری واژه‌نامه KSVD [۱۹, ۲۰] استفاده می‌شود. شکل (۸-۲) شمای کلی روش ارائه شده در [۱۸]، را نمایش می‌دهد.



شکل ۸-۲. شمای کلی روش ارائه شده در [۱۸]. ابتدا تصاویر تیره به حوزه‌ی LBP نگاشت می‌شوند و سپس از طریق نگاشت معکوس به فضای پیکسل انتقال داده می‌شوند.

۲-۳- جمع‌بندی

در این فصل روش‌های استخراج چهره در قالبی دسته‌بندی شده و مورد بررسی قرار گرفتند. این روش‌ها به دو دسته‌ی مبتنی بر ویژگی و مبتنی بر تصویر دسته‌بندی شدند. روش‌های پایه برای هر دسته به صورتی مختصر شرح داده شد. در هر بخش سعی شد تا نقاط ضعف و قدرت روش‌های مطرح شده بیان شود. روش‌های استخراج چهره بر اساس ترکیب این روش‌ها و با استفاده از روش‌های یادگیری که اخیراً مطرح شده نیز قابل افزودن بر دیاگرام شکل (۱-۲) می‌باشند.

فصل سوم

روش پیمایشی

مقدمه

در این فصل به ارائه‌ی روش پیشنهادی خواهیم پرداخت. در ابتدا به شرح ملزومات و تکنیک‌های مورد استفاده در روش پیشنهادی این پایان‌نامه می‌پردازیم. روش پیشنهادی از دو تکنیک استفاده می‌کند. تکنیک اول با انتقال مکان پیکسل‌های تصویر، بدون تغییر در شدت روشنایی آن‌ها، تصویر را دست‌خوش تغییر قرار می‌دهد. این تکنیک پیچ و تاب دادن تصویر نام دارد. تکنیک دوم، با دریافت دو تصویر ورودی به نام‌های تصویر مبدا و مقصد، تصویر مبدا را به صورت تدریجی به سمت تصویر مقصد تغییر می‌دهد که حاصل آن، دنباله‌ای از تصاویر میانی است و به دگرذیسی معروف است. تکنیک پیچ و تاب، بخشی از فرایند دگرذیسی است. با بیان تمامی تکنیک‌ها و ملزومات، به شرح روش پیشنهادی، تحلیل اثر آن و نتایج بصری تصاویر تولید شده خواهیم پرداخت. در ادامه، روش پیشنهادی با هدف مدل‌سازی تصاویر روبه‌رو به کار گرفته می‌شود و سپس تصاویر حاصل با یک روش ترازسازی تصاویر توسط تجزیه مرتبه پایین، تراز می‌شوند. به همین خاطر در بخش ملزومات، به تفصیل به آن الگوریتم خواهیم پرداخت.

۳-۱- انتقال نسبی

انتقال نسبی^۱ نوعی انتقال در صفحه‌ی تصویر است که با جابه‌جایی پیکسل‌های تصویر، شی را به گونه‌ای منتقل می‌نماید که ساختار آن حفظ می‌شود. انتقال نسبی، خط‌ها و نسبت فواصل را ثابت نگه می‌دارد ولی جهت، اندازه و مکان شی را تغییر می‌دهد [۲۱]. در فضای اقلیدسی معادله‌ی انتقال نسبی تصویر به صورت رابطه‌ی (۳-۱) است.

$$\hat{p}^T = p^T A + c \quad (3-1)$$

^۱ Affine

که p و $\hat{p}$ نقاط هندسی به ترتیب قبل و بعد انتقال می‌باشند و $A \in \mathbb{R}^{2 \times 2}$ ماتریس انتقال، مقیاس و چرخش بوده و c مقدار جابه‌جایی است. با استفاده از مختصات همگن^۱ [۲۲]، ماتریس انتقال نسبی به صورت $A = \begin{pmatrix} a_{11} & a_{12} & c_x \\ a_{21} & a_{22} & c_y \\ p_x & p_y & 1 \end{pmatrix}$ نمایش داده می‌شود که c_x و c_y بیانگر مقدار جابه‌جایی و a_{ij} عناصر انتقال نسبی هستند. بردار تصویرسازی است و در انتقال $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ (انتقال از فضای دوبعدی به دوبعدی) آنرا صفر در نظر می‌گیریم. انتقال یک نقطه در این مختصات با رابطه‌ی (۲-۳) صورت می‌پذیرد.

$$\hat{p}_h^T = p_h^T A \quad (2-3)$$

همانند قبل p_h و $\hat{p}_h$ به ترتیب نقاط هندسی قبل و بعد انتقال در دستگاه مختصات همگن می‌باشند.

۲-۲- ترازسازی مقاوم با تجزیه‌ی تنک و مرتبه پایین

ترازسازی مقاوم با تجزیه‌ی تنک و مرتبه پایین^۲ [۲۳] (RASL) روشی برای ترازسازی دسته‌ای از تصاویر یک شی نسبت به هم است که با هم دارای همبستگی خطی^۳ می‌باشند. این روش در حضور انسداد، شدت روشنایی غیر یکنواخت، ناترازی تصاویر و همچنین نسبت به حالات مختلف چهره مقاوم است. اساس کار RASL این است که به دنبال تبدیلاتی بهینه در حوزه‌ی مکانی تصویر است، به‌طوری که بعد از تبدیل، ماتریس تصویر قابلیت تجزیه به حاصل جمع ماتریس مرتبه‌ی پایین و ماتریس خطا را داشته باشد. این مساله نوعی بهینه‌سازی غیرمحدب^۴ محسوب می‌شود که می‌توان تحت شرایطی آنرا با مجموعه‌ای از

^۱ سیستم مختصاتی می‌باشد که در کاربردهایی همچون هندسه‌ی نگاشتی و انتقالات استفاده می‌شود. هندسه‌ی نگاشتی چهارچوبی برای یکپارچه‌سازی انتقالات هندسی در ماتریس‌هایی به نام ماتریس‌های انتقال و استفاده از جبر ماتریس‌ها ارائه می‌دهد.

^۲ Robust Alignment by Sparse and Low Rank decomposition

^۳ Linearly Correlated

^۴ Non Convex Optimization

بهینه‌سازی‌های محدب تقریب زد. روش RASL، تعدادی از تصاویر یک شی را نسبت به یک قاب و دو نقطه‌ی کانونی تراز می‌کند. به صورت دقیق، مساله‌ی روش RASL این‌گونه بیان می‌شود: فرض کنیم که n تصویر از یک شی که به خوبی تراز شده‌اند و با هم دارای همبستگی خطی هستند، در یک مجموعه مانند، $\{I_i^0\}_{i=1}^n$ داریم. $\text{vec}: \mathbb{R}^{w \times h} \rightarrow \mathbb{R}^m$ عملگری است که ناحیه‌ی m پیکسلی از شی مورد نظر را استخراج می‌نماید و در یک ماتریس به صورت $A \doteq [\text{vec}(I_1^0) | \dots | \text{vec}(I_n^0)] \in \mathbb{R}^{m \times n}$ ذخیره می‌کند (صفر در I_i^0 نشان‌دهنده‌ی مرتبه‌ی پایین^۱ ماتریس تصویر است). وجود همبستگی خطی در تصاویر به معنی داشتن مرتبه‌ی پایین A می‌باشد. خرابی‌های تصویر مانند انسداد، شدت روشنایی غیر یک‌نواخت و ناترازی‌های ناشی از حرکت شی یا دوربین، می‌توانند باعث نقض ساختار مرتبه‌ی پایین A شوند. به همین جهت Peng و همکارانش [۲۳] برای مدل‌سازی تصاویر دنیای واقعی از رابطه‌ی $I_i = (I_i^0 + e_i) \circ \tau_i$ استفاده نمودند. سپس با تشکیل ماتریس‌های تنک خرابی‌ها و ماتریس تصاویر تراز شده به صورت $D = [\text{vec}(I_1) | \dots | \text{vec}(I_m)] \in \mathbb{R}^{m \times n} = A + E$ و $E = [\text{vec}(e_1) | \dots | \text{vec}(e_n)] \in \mathbb{R}^{m \times n}$ (تصاویری که فرض می‌شود نسبت به هم تراز بوده و قابلیت تجزیه به ماتریس خرابی و ماتریس مرتبه پایین را دارند) تابع هدف را تشکیل دادند. τ_i ها تبدیلات پارامتری همچون تشابه^۲، تبدیل نسبی و هموگرافی صفحه‌ای^۳ هستند. Peng و همکارانش [۲۳] جهت بدست آوردن پاسخ مساله یا $\{I_i^0\}_{i=1}^n$ ، از این ایده استفاده نمودند که اگر تصاویر نسبت به هم تراز شوند، باید دارای ساختار مرتبه‌ی پایین نسبت به مقداری خطای تنک باشند. بنابراین روش آن‌ها به دنبال گروه تبدیلات $T = \{\tau_i\}_{i=1}^n$ می‌گردد که اگر خطای تصاویر انتقال یافته تحت انتقال معکوس جبران شود، رتبه‌ی آن‌ها تا حد ممکن کاهش یابد. تابع هدف روش RASL به صورت رابطه‌ی (۳-۳) تعریف می‌شود.

^۱ Low Rank

^۲ Similarity

^۳ Homography

$$\min_{A,E,T} \text{rank}(A) \quad \text{s.t.} \quad D \circ T = A + E, \|E\|_0 < k \quad (3-3)$$

یا به صورت فرم لاگرانژین

$$\min_{A,E,T} \text{rank}(A) + \lambda \|E\|_0 \quad \text{s.t.} \quad D \circ T = A + E \quad (4-3)$$

که $\| \cdot \|_0$ نرم صفر بوده و با ι^0 نمایش داده می‌شود. $\lambda > 0$ در این رابطه، مصالحه‌ای میان مرتبه‌ی ماتریس پاسخ و میزان تنگی خطا برقرار می‌سازد. این مساله یک بهینه‌سازی غیرمحدب و غیرخطی ناپیوسته محسوب می‌شود که با استفاده از الگوریتم‌های تقریبی [۲۳]، تابع غیرمحدب هدف با تابع محدب جایگزین شده و همین‌طور محدودیت غیرخطی به محدودیت خطی تبدیل می‌شود. با این روش تابع هدف به مجموعه‌ای از بهینه‌سازی‌های محدب تبدیل می‌شود. مراحل ساده‌سازی تابع هدف^۱ این روش به صورت خلاصه ذیلا آورده شده است. لازم به ذکر است که این ساده‌سازی وقتی مقدار ناترازی بسیار زیاد نمی‌باشد، تقریبی قابل قبول از تابع هدف است.

مراحل ساده‌سازی تابع هدف به دو بخش تقسیم می‌شود.

۱. در ابتدا تابع هدف نامحدب (رابطه‌ی (۴-۳)) با یک تابع محدب تقریب زده می‌شود.

❖ ثابت شده که اگر مرتبه‌ی ماتریس A بزرگ نباشد و تعداد خطاهای تنگ نیز کم باشد، تابع

نامحدب $\text{rank}(A) + \lambda \|E\|_0$ ، قابلیت تخمین با تابع محدب $\min \|A\|_* + \lambda \|E\|_1$ را دارد

[۲۳]. در این رابطه، $\| \cdot \|_*$ نرم هسته‌ای یا مجموع مقادیر تکیه یک ماتریس است و برای

ماتریس A ، شکل آن به صورت $\sum_{ij} \sigma_i(A)$ می‌شود. همچنین، $\| \cdot \|_1$ تابع نرم یک (ι^1) است

که برای ماتریس E به صورت $\sum_{ij} |E_{ij}|$ می‌شود.

^۱ Target function relaxation

۲. سپس قید غیرخطی رابطه‌ی (۴-۳)، $(D \circ T = A + E)$ تبدیل به قید خطی خواهد شد که

ماحصل آن، تبدیل تابع هدف نامحدب به مجموعه‌ای از بهینه‌سازی‌های محدب می‌باشد.

❖ برای خطی‌سازی قید فرض می‌شود که تبدیلات دارای p پارامتر بوده و مجموعه‌ی تبدیلات

به صورت $\tau \in \mathbb{G} = [\tau_1 | \dots | \tau_1] \in \mathbb{R}^{p \times n}$ در نظر گرفته می‌شود. اگر تغییر در مقدار

پارامترهای تبدیلات کم باشد، می‌توان قید غیرخطی رابطه‌ی (۴-۳) را با مجموع مشتقات

جزئی مرتبه اول توابع انتقال، به صورت $D \circ (\tau + \Delta\tau) \approx D \circ \tau + \sum_{i=1}^n J_i \Delta\tau_i \varepsilon_i^T$ تخمین

زد. در این رابطه، $J_i \triangleq \frac{\partial}{\partial \zeta} \text{vec}(I_i \circ \zeta)|_{\zeta=\tau_i} \in \mathbb{R}^{m \times p}$ ماتریس جاکوبین i امین تصویر نسبت

به پارامترهای τ_i بوده و $\{\varepsilon_i\}$ پایه^۱ برای فضای $\mathbb{R}^n$ را مشخص می‌نماید. بنابراین، با جایگزینی

معادله‌ی محدب و قید خطی مذکور، تابع هدف RASL، تبدیل به رابطه‌ی (۵-۳) می‌شود.

$$\min_{A, E, \tau} \|A\|_* + \lambda \|E\|_1 \quad s. t. \quad D \circ \tau + \sum_{i=1}^n J_i \Delta\tau_i \varepsilon_i^T = A + E \quad (5-3)$$

با استفاده از تکنیک‌های بهینه‌سازی مدرن، تابع هدف (۵-۳) (که اکنون محدب بوده و دارای قید خطی

می‌باشد) قابل حل است [۲۳]. همگرایی روش بهینه‌سازی که Peng و همکارانش برای حل تابع هدف

RASL ارائه دادند، تضمین شده است [۲۳].

الگوریتم (۱-۳) و (۲-۳) به ترتیب حلقه‌ی خارجی و داخلی الگوریتم RASL را نمایش می‌دهند. شکل

(۳-۳) نیز نمونه‌ای از خروجی روش RASL را برای مجموعه‌ای از تصاویر ورودی که نسبت به هم تراز

نبوده و دارای حالات مختلف چهره، انسداد و شدت روشنایی‌های متفاوت می‌باشند نشان می‌دهد. اما این

روش مقاومتی در مقابل تصاویر چهره‌ی دارای زاویه نسبت به دوربین ندارد. شکل (۴-۳) نتیجه‌ی

ترازسازی مجموعه‌ای از تصاویر چهره دارای زاویه را با RASL نسبت به هم را نشان می‌دهد.

^۱ Basis

Algorithm 1(Outer loop of RASL)

INPUT: Images $I_1, \dots, I_n \in \mathbb{R}^{w \times h}$, initial transformations $\tau_1, \dots, \tau_n$ in certain parametric group G , weight $\lambda > 0$.

WHILE not converged DO

Step 1 : compute Jacobian matrices w.r.t transformation:

$$J_i \leftarrow \frac{\partial}{\partial \zeta} \left(\frac{\text{vec}(I_i \circ \zeta)}{\|\text{vec}(I_i \circ \zeta)\|_2} \right) \Big|_{\zeta=\tau_i}, i = 1, \dots, n;$$

Step 2 : warp and normalize the images:

$$D \circ \tau \leftarrow \left[\frac{\text{vec}(I_1 \circ \zeta)}{\|\text{vec}(I_1 \circ \zeta)\|_2} \mid \dots \mid \frac{\text{vec}(I_n \circ \zeta)}{\|\text{vec}(I_n \circ \zeta)\|_2} \right];$$

Step 3 (inner loop): solve the linearized convex optimization:

$$\begin{aligned} (A^*, E^*, \Delta \tau^*) &\leftarrow \arg \min_{A, E, \Delta \tau} \|A\|_* + \lambda \|E\|_1 \\ \text{s.t. } & D \circ \tau + \sum_{i=1}^n J_i \Delta \tau \mathcal{E}_i^T = A + E ; \end{aligned}$$

Step 4: update transformations: $\tau \leftarrow \tau + \Delta \tau$;

END WHILE

OUTPUT: solution $A^*, E^*, \Delta \tau^*$ problem (5-3)

الگوریتم ۳-۱ حلقه‌ی خارجی روش RASL [۲۳].

Algorithm 2 (Inner loop of RASL)

INPUT: $(A^0, E^0, \Delta\tau^0) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}$, $\lambda > 0$.

WHILE not converged DO

$$(U, \Sigma, V) = \text{svd}(D \circ \tau + \sum_{i=1}^n J_i \Delta\tau \varepsilon_i \varepsilon_i^T + \frac{1}{\mu_k} Y_k - E_k);$$

$$A_{k+1} = U \Sigma \frac{1}{\mu_k} [V]^T;$$

$$E_{k+1} = S \frac{\lambda}{\mu_k} \left[D \circ \tau + \sum_{i=1}^n J_i \Delta\tau \varepsilon_i \varepsilon_i^T + \frac{1}{\mu_k} Y_k - A_{k+1} \right];$$

$$\Delta\tau_{k+1} = \sum_{i=1}^n J_i^\dagger (A_{k+1} + E_{k+1} - D \circ \tau - \frac{1}{\mu_k} Y_k) \varepsilon_i \varepsilon_i^T;$$

$$Y_{k+1} = Y_k + \mu_k h(A_{k+1}, E_{k+1}, \Delta\tau_{k+1});$$

END WHILE

OUTPUT: solution $(A^*, E^*, \Delta\tau^*)$.

الگوریتم ۳-۲ حلقه‌ی داخلی روش RASL [۲۳].

در حلقه‌ی داخلی روش RASL، Y ماتریس ضرایب لاگرانژ و μ مقدار مثبت عددی است. در رابطه‌ی

(۵-۳)، تابع $h(A, E, \Delta\tau) = D \circ \tau + \sum_{i=1}^n J_i \Delta\tau \varepsilon_i \varepsilon_i^T - A - E$ به جای قید خطی استفاده شده است.

این امر به جهت جست‌وجو برای نقطه‌ی زینی در عوض یافتن پاسخ مستقیم معادله، در فرم ضرایب

افزوده‌ی لاگرانژی^۱ رابطه‌ی (۵-۳) است. $\dagger$ عملگر موور-پنروز^۲ می‌باشد.

^۱ Augmented Lagrange Multiplier

^۲ Moore – Penrose Pseudoinverse (تعمیمی برای عمل معکوس ماتریس با هر ابعادی است و در سیستم‌های خطی کاربرد دارد)

در این روش ماتریس A ، واژه‌نامه متشکل از تصاویر مرتبه پایین است که پس از بهینه‌سازی تابع هدف (۵-۳) بدست آمده‌اند.



شکل ۳-۳. نمونه‌ای از تصاویر ورودی روش RASL و تصاویر خروجی آن. تصویر سمت راست، تصاویر ورودی با حالات، انسداد و شدت روشنایی‌های متفاوت از یک شخص هستند. تصویر پایین آن نیز میانگین تصاویر ورودی است. شکل سمت چپ خروجی روش RASL (I_1^0) و تصویر پایین آن میانگین تصاویر تراز شده است.



شکل ۳-۴. نتیجه‌ی ترازسازی تصاویر زاویه‌دار چهره با RASL

۳-۳- پیچ و تاب تصویر

فرض کنید تصویری مانند $I \in \mathbb{R}^{w \times h}$ به عنوان تصویر ورودی داریم و مجموعه‌ای از توابع انتقال بر روی پیکسل‌های تصویر به صورت $w \subseteq W = \{T_i(x, y): \mathbb{R}^2 \rightarrow \mathbb{R}, i \in \{u, v\}\}$ موجود باشد. فرایند پیچ و تاب دادن تصویر با رابطه‌ی (۶-۳) تعریف می‌شود.

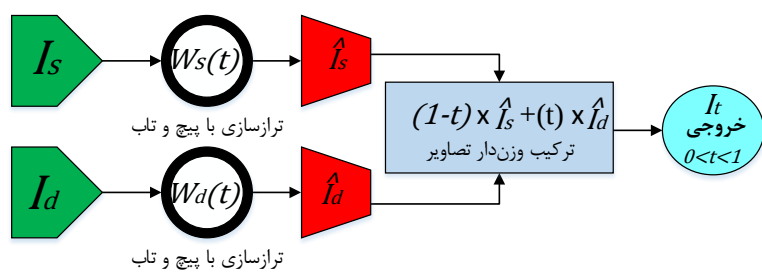
$$\hat{I}(u, v) = w \circ I(x, y) = I(T_u(x, y), T_v(x, y)) \quad (6-3)$$

که $\hat{I}$ تصویر انتقال یافته است. در پیچ و تاب دادن تصویر ممکن است که مکان چند پیکسل به یک نقطه نگاشت شود، یا یک پیکسل به خارج از محدوده تصویر انتقال یابد. با توجه به محدود بودن تعداد پیکسل‌های تصویر، ممکن است که به تعدادی از پیکسل‌ها مقداری به عنوان شدت روشنایی تعلق نگیرد. این پیکسل‌ها باعث به وجود آمدن مصنوعات در تصویر انتقال یافته می‌شوند.^۱ برای حذف مصنوعات به وجود آمده، با معکوس نمودن توابع پیچ و تاب (با فرض یک‌به‌یک بودن آن‌ها) و مکان‌یابی پیکسل‌های تصویر خروجی در تصویر ورودی، انتقال به صورت معکوس انجام می‌شود. در انتقال معکوس، برای پیکسل‌هایی که به بیرون از صفحه‌ی تصویر نگاشت می‌شوند، می‌توان از مکاشفه‌ی ساده‌ی نگاشت به نزدیکترین پیکسل استفاده نمود. همچنین می‌توان با درونیابی شدت روشنایی پیکسل‌های همسایه، به پیکسل‌هایی که شدت روشنایی به آن‌ها تعلق نگرفته، عددی نسبت داد و با این روش اثر مصنوعات را از بین برد [۲۱]. از دیگر کاربردهای پیچ و تاب تصویر می‌توان به نرمال‌سازی، و تطبیق تصاویر اشاره نمود [۲۱]. در اکثر انتقال‌ها، نقاطی بر روی تصاویر انتخاب می‌شوند و تابع انتقال پیچ و تاب از این نقاط استخراج می‌شود. یکی از روش‌های استخراج توابع انتقال از نقاط، درونیابی می‌باشد [۲۵].

^۱ Aliasing

۳-۴- دگردیسی تصویر

فرایند دگردیسی متشکل از سه مرحله می‌باشد: (۱) استخراج توابع انتقال از نقاط ایجاد شده. (۲) اعمال توابع انتقال بر روی تصاویر مبدا و مقصد برای ترازسازی مکان نقاط (۳) ترکیب وزن‌دار تصاویر تراز شده‌ی مبدا و مقصد نسبت به پارامتر زمان [۲۶-۲۸]. بخش مهم دگردیسی، انتخاب مناسب تابع انتقال برای پیچ و تاب است. اگر w مجموعه‌ی توابع پیچ و تاب، t پارامتر زمان و I_s و I_d به ترتیب تصاویر مبدا و مقصد باشند، دگردیسی این تصاویر با I_t نشان داده شده و به صورت شکل (۷-۳) می‌باشد.



(۷-۳)

وقتی دو تصویر I_s و I_d داده می‌شوند، مساله‌ی دگردیسی تصویر به صورت تولید دنباله‌ای از تصاویر میانی $I(t)$ تعریف می‌شود که $I(0) = I_s$ و $I(1) = I_d$. با تغییر t از صفر به یک، $0 < t < 1$ تصویر مبدا I_s به صورت پیوسته تغییر می‌کند تا به I_d تبدیل شود. فرض کنیم که w_s تابع پیچ و تاب‌ی است که مکان و شکل ویژگی‌های مطابق I_s بر روی I_d را تراز می‌کند. همچنین w_d تابع پیچ و تاب برای ترازسازی مکان و شکل ویژگی‌های مطابق I_d بر روی I_s باشد. برای ساخت یک تصویر میانی $I(t)$ در ابتدا باید دو تابع $w_s(t)$ و $w_d(t)$ را از w_s و w_d با وابسته‌سازی آنها به پارامتر زمان^۱، بدست آوریم. سپس I_s و I_d به وسیله‌ی $w_s(t)$ و $w_d(t)$ نسبت به هم تراز می‌شوند که نتیجه‌ی آن تصاویر $I_s(t)$ و $I_d(t)$ است. در نهایت با جمع وزن‌دار تصاویر مبدا و مقصد تراز شده که به ترتیب با $I_s(t)$ و $I_d(t)$ نشان داده می‌شوند، نسبت به

^۱ روش‌های مختلفی برای این کار وجود دارد. مثلاً اگر از درونیایی برای استخراج توابع انتقال استفاده شود، نقاط آموزشی توابع، توسط جمع وزن‌دار نقاط تصاویر مبدا و مقصد تعریف می‌شوند و بدین صورت توابع انتقال به زمان وابسته می‌شوند.

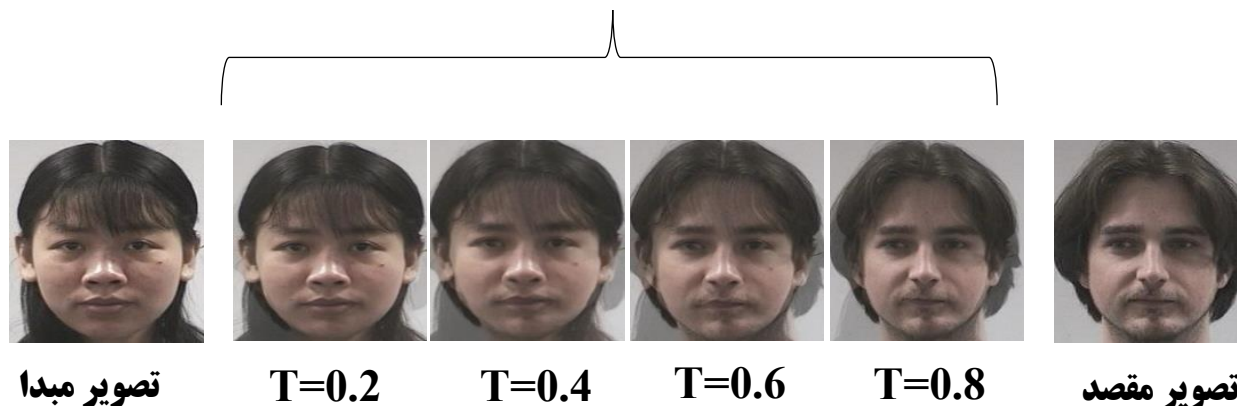
پارامتر زمان، تصویر جدید حاصل می‌شود. روابط (۸-۳) ترازسازی و ترکیب وزن دار تصاویر ورودی و خروجی را نشان می‌دهند. شکل (۵-۳) نیز نمونه‌ای از فرایند دگردیسی را برای دو تصویر چهره نشان می‌دهد.

$$I_0(t) = W_0(t) \circ (I_0),$$

$$I_1(t) = W_1(t) \circ (I_1), \quad (۸-۳)$$

$$I(t) = (1 - t) \cdot I_0(t) + t \cdot (I_1(t))$$

تصاویر میانی



شکل ۵-۳. نمونه‌ای از فرایند دگردیسی برای دو تصویر چهره

۵-۳- انواع زوایای تصویر

در تهیه‌ی تصاویر مختلف از یک صحنه، میان اشیای درون صحنه و همچنین صفحه‌ی تصویر ممکن است سه نوع زاویه ایجاد شود. در نوع اول، شیء موجود در تصویر ساختار خود را حفظ نموده و نسبت بین اجزای آن ثابت باقی می‌ماند. تصحیح این نوع زاویه تحت تبدیل نسبی امکان‌پذیر است [۲۹،۳۰]. نوع دوم هنگامی رخ می‌دهد که صحنه‌ی تصویر هم‌راستا با صفحه‌ی دوربین نباشد. در این نوع از زاویه، هنگام

نگاشت تصویر از فضای سه‌بعدی به دو بعدی، ساختار اشیا حفظ می‌شود، اما ممکن است نسبت اجزای شیء به هم و اندازه‌ی آن‌ها ثابت باقی نماند. این نوع زاویه تحت تبدیل معکوس پرسپکتیو^۱ [۲۱] قابل تصحیح است. روش ارائه شده در [۲۳] (RASL) قادر به تصحیح تبدیلات نسبی و پرسپکتیو است. نوع سوم با چرخش شیء در محیط به وجود می‌آید. چون بخشی از شیء در تصویر پنهان می‌شود، تصحیح این نوع زاویه دشوار است. شکل (۳-۶) انواع زاویه‌های مطرح شده در فوق را نشان می‌دهد.



پ

ب

الف

شکل ۳-۶. انواع زاویه‌های موجود در تصویر. الف: زاویه نسبی. ب: زاویه پرسپکتیو. ج: زاویه حاصل از چرخش سر

در تصاویر ویدئویی، امکان ایجاد زاویه از نوع اول و سوم در چهره، به دلیل چرخش سر بر محور گردن وجود دارد. اگر دوربین تصویربرداری قادر به کالیبره‌سازی^۲ تصاویر باشد یا تصاویر استریو^۳ موجود باشد، روش‌هایی وجود دارند که با استفاده از مدل‌های سه‌بعدی قادر به ساخت تصویر روبه‌روی چهره می‌باشند [۳۱، ۳۲]. اما به دلیل هزینه‌بر بودن تامین دوربین‌هایی که قادر به تولید این نوع از تصاویر هستند، کاربرد روش‌های مبتنی بر مدل‌سازی سه‌بعدی در نظارت ویدئویی محدود می‌شود. لازم به ذکر است که روش-

^۱ Perspective

^۲ در تصاویر کالیبره، امکان تخمین فاصله‌ی اشیا تا دوربین تحت پارامتری به نام فاکتور کالیبره که دوربین تصویربرداری در اختیار می‌گذارد وجود دارد.

^۳ تصاویر استریو جفت تصاویری هستند که توسط دوربین‌های ثابت شده بر اساس طول یک خط پایه اخذ می‌شود. این تصاویر حاوی اطلاعاتی مانند عمق و زاویه صفحه‌ی تصویر نسبت به دوربین می‌باشند و با این اطلاعات امکان مدل‌سازی سه‌بعدی را فراهم می‌آورند.

های مبتنی بر مدل سه بعدی معمولاً بلادرنگ نبوده و در مدل سازی آنها، امکان بازسازی بخش هایی از چهره که از تصاویر ورودی قابل استخراج نیستند وجود دارد. بنابراین ممکن است که مدل چهره ی ارائه شده توسط این روش ها با چهره ی شخص تفاوت هایی داشته باشد.

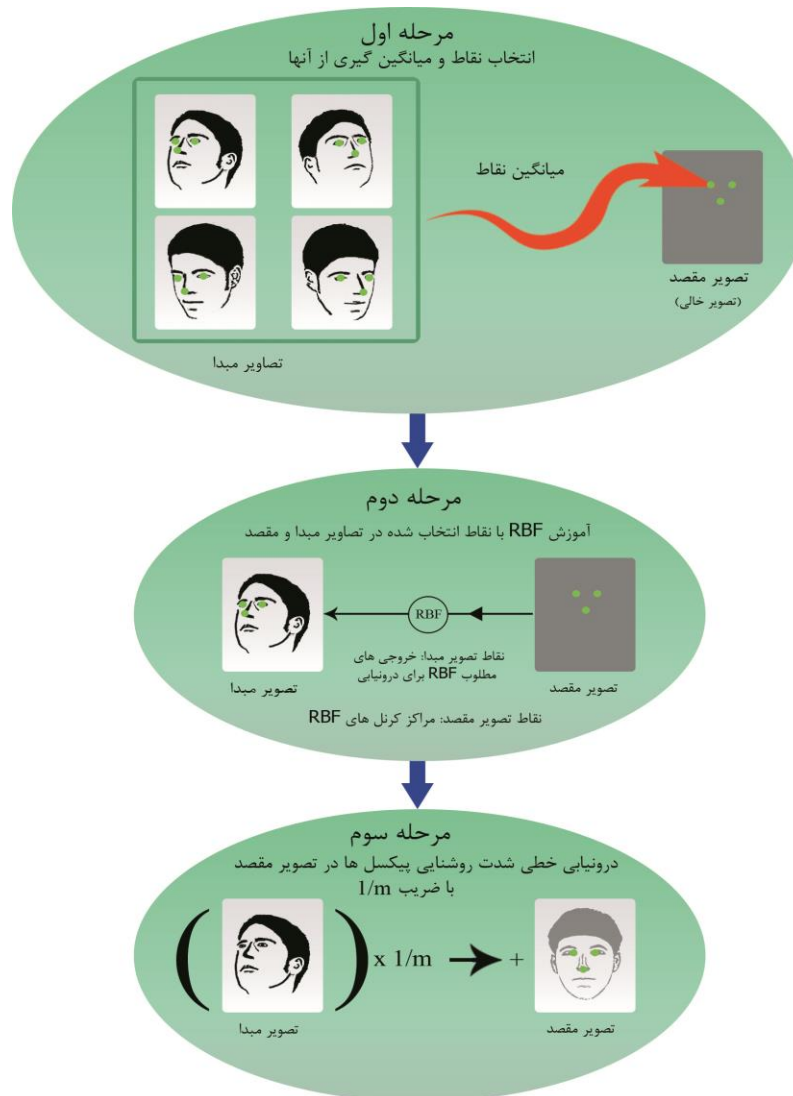
روش پیشنهادی در این پایان نامه، با مدل سازی چهره از تصاویر زاویه دار، به تصحیح زاویه های نوع اول و سوم چهره در تصاویر ویدئویی خواهد پرداخت. روش پیشنهادی، بدون در نظر گرفتن فرض های سه بعدی و تنها با استفاده از تصاویر ساده ی دوبعدی، مدل سازی چهره ی روبه رو را بهبود می بخشد.

۳-۶- روش پیشنهادی

روش پیشنهادی مطرح شده در این پایان نامه مبتنی بر دگردیسی بوده و از چند تصویر ورودی به عنوان تصاویر مبدا استفاده می نماید. در ابتدا تعدادی نقاط از بخش های چهره در تصاویر مختلف ورودی انتخاب می شوند. فرض کنید تعداد این نقاط n عدد در هر تصویر باشد و m تصویر ورودی نیز در اختیار باشد. برای تراز سازی تصاویر مبدا نسبت به هم، نیاز است تا نقاط انتخاب شده در تصاویر مبدا با هم دارای تطابق باشد. به این معنی که مجموعه نقاط نشان دهنده ی یک بخش خاص از تصویر چهره ی فرد (مثلاً چشم راست) که از تمامی تصاویر مبدا انتخاب شده اند، در یک بردار سازمان دهی شوند. طول این بردارها برابر با m (تعداد تصاویر ورودی) می باشد و تعداد این بردارها بستگی به انتخاب نقاط در بخش های مختلف چهره دارد. به نقاط این مجموعه، نقاط مطابق^۱ می گوئیم. نقاط مطابق در تصویر مقصد، از میانگین بردارهای نقاط مطابق تصاویر مبدا بدست می آیند. سپس توابع انتقال میان تصویر مقصد و هر تصویر مبدا محاسبه می شوند (به ازای هر تصویر مبدا یک مجموعه تابع انتقال). در نهایت با استفاده از انتقال معکوس، مکان تمامی پیکسل های تصویر مقصد، در تصویر مبدا محاسبه شده و مقدار شدت روشنایی آنها با درونیایی

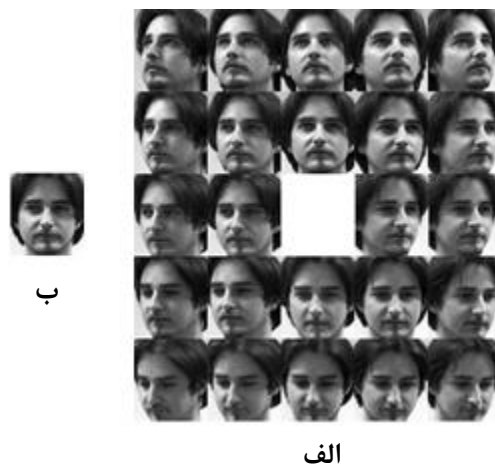
^۱ Correspond points

خطی بدست می‌آید. این روند برای هر تصویر ورودی انجام می‌شود و هر تصویر به اندازه‌ی $1/m$ به شدت روشنایی پیکسل‌های تصویر مقصد می‌افزاید. شکل (۷-۳) شمای کلی روش پیشنهادی را برای چهار تصویر ورودی و سه نقطه‌ی مطابق، در سه گام خلاصه می‌نماید.



شکل ۷-۳. خلاصه‌ی روش پیشنهادی در سه گام. گام اول: ایجاد مجموعه نقاط مطابق و میانگین‌گیری از آن‌ها. گام دوم: آموزش تابع انتقال و استفاده از انتقال معکوس برای مکان‌یابی پیکسل‌های تصویر مدل در تصاویر مبدا. گام سوم: بدست آوردن مقدار شدت روشنایی پیکسل‌های تصویر مدل.

در فصل اول صورت مساله بدین صورت بیان شد که مجموعه‌ای از تصاویر با اندازه‌ی یکسان به صورت $C = \{i \in I^{\pm y^{\circ} \pm x^{\circ}}, x \text{ or } y \neq 0\}$ به عنوان ورودی داریم. می‌خواهیم با تصاویر مجموعه‌ی C ، تصویری از روبه‌روی چهره‌ی شخص را مدل کنیم. برای یادآوری در شکل (۳-۸) یک نمونه از این مجموعه تصاویر ورودی نشان داده شده است. فرض می‌کنیم که تصاویر مجموعه‌ی C فقط دارای زاویه‌ی چهره باشند. به علاوه مجموعه‌ی C ، خاصیت $\{ \forall i: i^{-y^{\circ} - x^{\circ}} \in C \Rightarrow i^{+y^{\circ} + x^{\circ}} \in C \}$ (قرینگی نسبت به زاویه) را دارد.



شکل ۳-۸. تصاویر مبدا و تصویر مقصد. الف: تصاویر مبدا، تصویر گوشه‌ی بالا سمت چپ دارای زاویه‌ی $(+30+30)$ است. میزان چرخش چهره در راستای افقی و عمودی تصاویر متوالی به اندازه‌ی ۵- است. ب: تصویر مقصد که در شکل (الف) جای آن خالی گذاشته شده و دارای زاویه $(++0)$ است.

همان‌طور که پیش‌تر بیان شد یکی از روش‌های استخراج توابع انتقال از مجموعه نقاط مطابق، درونیابی می‌باشد. شبکه‌های عصبی تابع پایه شعاعی (RBF)^۱ ابزاری قدرتمند جهت درونیابی داده‌های چند متغیره‌ی پراکنده^۲ به شمار می‌روند. برای تابع انتقال، از معماری شبکه‌ی عصبی تابع پایه شعاعی معرفی شده در [۲۵] استفاده می‌کنیم. این تابع با رابطه‌ی (۳-۹) نشان داده شده است.

$$RBF(x, y): R^2 \rightarrow R = \sum_{i=1}^n \omega_i k \|c_i - \rho\|_2 + P_m(x, y),$$

^۱ Radial Basis Function

^۲ Scattered data interpolation

$$P_m(x, y) \in \sum_{i=0}^m \sum_{j=0}^{m-i} a_{ij} x^i y^j \quad (9-3)$$

با شرط ارضای قید زیر

$$\sum_{i=1}^n \omega_i = \sum_{i=1}^n \omega_i x_i = \dots = \sum_{i=1}^n \omega_i y_i = \dots = 0$$

که در آن RBF شبکه عصبی ω_i و a_{ij} وزن های شبکه عصبی، P_m تابع تعمیم چندجمله ای، c_i مراکز کرنل ها، $\rho = (\rho^x, \rho^y)$ مختصات پیکسل مورد نظر برای درونیایی و k تابع کرنل می باشند. $\| \cdot \|_2$ فاصله ی اقلیدسی است.

آموزش شبکه ی عصبی رابطه ی (9-3)، با عبارت تعمیم آن، باعث می شود که اگر فاصله نقاطی که با این شبکه درونیایی می شوند تا مراکز شبکه زیاد باشد، مقداری که شبکه به عنوان خروجی برای آن ها تولید می نماید از یک چندجمله ای پیروی نمایند [25]. در روش پیشنهادی مطلوب است که تابع انتقال تاثیرات سراسری بر حرکت پیکسل ها داشته باشد و حداقل خمیدگی را به صفحه ی تصویر وارد کند. تحقیقات انجام شده توسط [25]، انتخاب تابع پایه شعاعی $k(x) = x^2 \log x$ (TPS) با $k(0)=0$ به عنوان هسته شبکه ی عصبی و تنظیم درجه ی چندجمله ای تابع تعمیم آن (P_m) به 1 را برای رسیدن به این هدف ملزم می داند. پس مجموعه توابع انتقال با رابطه ی (10-3) مشخص می شوند.

$$w = \{(T_u, T_v) \in R(x, y)\}: R^2 \rightarrow R^2 = \begin{pmatrix} T_u(\rho) = \sum_{i=1}^n \omega_i k(\|c_i - \rho\|_2) + a_{00} + a_{01}\rho^y + a_{10}\rho^x \\ T_v(\rho) = \sum_{i=1}^n \alpha_i k(\|c_i - \rho\|_2) + b_{00} + b_{01}\rho^y + b_{10}\rho^x \end{pmatrix} \quad (10-3)$$

¹ Thin Plate Spline

برای بدست آوردن توابع انتقال، مجموعه بردارهای نقاط مطابق تصاویر مبدا را به صورت مجموعه نقاط $S = \{s_i^d \mid d \in \{x, y\}, i = 1, 2, \dots, m\}$ در نظر می‌گیریم. نقاط مقصد، D ، نیز همین‌طور نمایش داده می‌شوند.

با تشکیل ماتریس S از بردارهای نقاط مطابق، با رابطه‌ی (۱۰-۳) به صورت زیر

$$S^x = \begin{bmatrix} s_{11}^x & \cdots & s_{1m}^x \\ \vdots & \ddots & \vdots \\ s_{n1}^x & \cdots & s_{nm}^x \end{bmatrix}, S^y = \begin{bmatrix} s_{11}^y & \cdots & s_{1m}^y \\ \vdots & \ddots & \vdots \\ s_{n1}^y & \cdots & s_{nm}^y \end{bmatrix} \quad (۱۱-۳)$$

نقاط متناظر تصویر مقصد را با بردارهای (۱۲-۳) تعریف می‌کنیم.

$$D^x = \begin{bmatrix} \frac{\sum_{j=1}^m s_{1,j}^x}{m} \\ \vdots \\ \frac{\sum_{j=1}^m s_{n,j}^x}{m} \end{bmatrix}, D^y = \begin{bmatrix} \frac{\sum_{j=1}^m s_{1,j}^y}{m} \\ \vdots \\ \frac{\sum_{j=1}^m s_{n,j}^y}{m} \end{bmatrix} \quad (۱۲-۳)$$

فرض کنید $s_1 = \{(s_{i1}^x, s_{i1}^y)\}_{i=1}^n$ نقاط انتخاب شده یکی از تصاویر مبدا باشند. تفاوتی در ترتیب ورود تصاویر مبدا به الگوریتم وجود ندارد. علت آن در بخش تحلیل اثر الگوریتم (بخش ۳-۷) ذکر خواهد شد. با در نظر گرفتن مجموعه D به عنوان مراکز کرنل‌های توابع انتقال و با درونیایی دقیق نقاط s_1 ، توابع انتقال w تشکیل می‌شوند. اگر از انتقال معکوس استفاده کنیم رابطه‌ی (۱۳-۳) را خواهیم داشت:

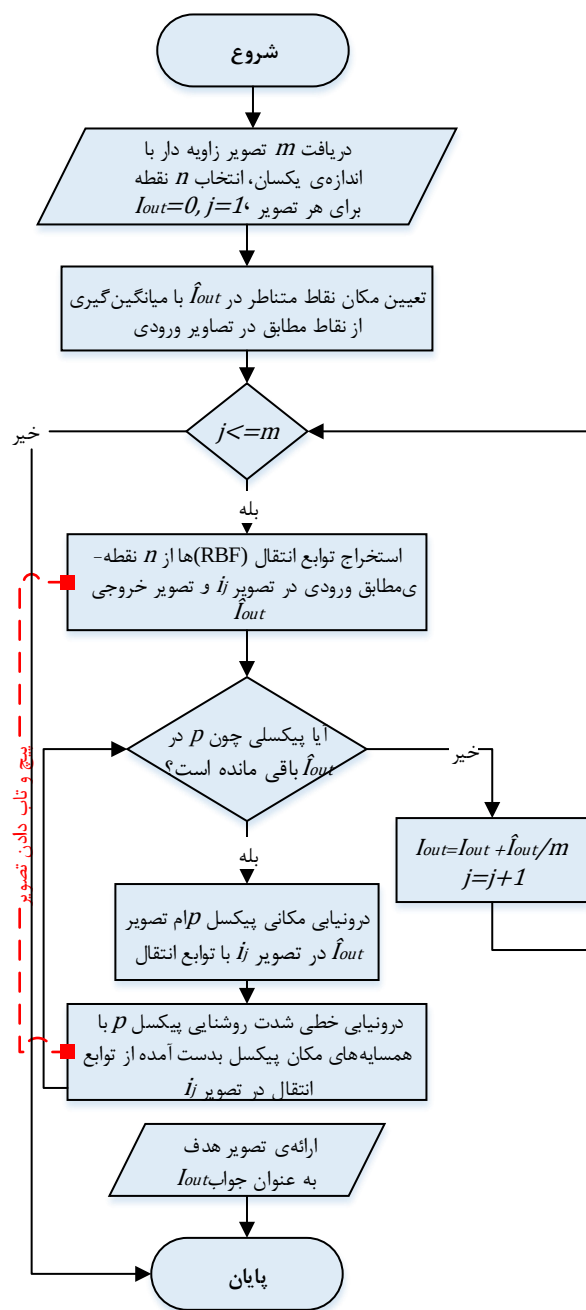
$$w = (T(D)) = \{(T_u(d_i), T_v(d_i))\}_{i=1}^n = \{(s_i^x, s_i^y)\}_{i=1}^n \quad (۱۳-۳)$$

مطابق رابطه‌ی (۱۳-۳) برای بعد X پیکسل‌های تصویر مدل، سیستم خطی (۱۴-۳) را خواهیم داشت:

$$\{s_1^x, d_1^x\}_1^n = \begin{bmatrix} T_u(d_1) = \sum_{i=1}^n \omega_i k(\|d_i - d_1\|_2) + \\ a_{00} + a_{01}d_1^y + a_{10}d_1^x = s_{1,1}^x \\ \vdots \\ T_u(d_n) = \sum_{i=1}^n \omega_i k(\|d_i - d_n\|_2) + \\ a_{00} + a_{01}d_n^y + a_{10}d_n^x = s_{n,1}^x \\ d_1^x \omega_1 + \cdots + d_n^x \omega_n = 0 \\ d_1^y \omega_1 + \cdots + d_n^y \omega_n = 0 \\ \omega_1 + \cdots + \omega_n = 0 \end{bmatrix} \quad (۱۴-۳)$$

به صورتی مشابه می‌توان برای مجموعه‌ی $\{s_1^y, d_1^n\}$ رابطه‌ی (۱۴-۳) را نوشت. پس با داشتن n نقطه‌ی متناظر، برای ساخت توابع انتقال، نیاز به حل سیستم خطی با $n + 3$ مجهول برای هر بعد مختصات داریم. با در نظر گرفتن قید تابع رابطه‌ی (۹-۳) سیستم خطی، رابطه‌ی (۱۴-۳) قابل حل می‌باشد. برای اینکه سیستم فوق دارای پاسخ باشد، تعداد نقاط انتخابی حداقل باید سه عدد باشند.

بعد از حل سیستم‌های خطی رابطه‌ی (۱۴-۳) توابع انتقال بدست خواهند آمد. این الگوریتم میانگین-گیری بر پایه‌ی دگرדיسی نامیده می‌شود. شکل (۹-۳) نمودار گردشی میانگین‌گیری دگرדיسی را نمایش می‌دهد. برای $m = 2$ ، الگوریتم برابر است با دگرדיسی تصویر با همان توابع انتقال در $t = 0.50$. شکل (۱۰-۳) نمونه‌ای از خروجی الگوریتم را برای تصاویر شکل (۸-۳) نشان می‌دهد.



شکل (۳-۹) نمودار گردش الگوریتم پیشنهادی، میانگین گیری بر پایه ی دگر دیسی.



شکل ۳-۱۰. نتایج حاصل از اعمال الگوریتم پیشنهادی. میانگین‌گیری بر روی تمامی تصاویر شکل (۳-۸-الف). در این آزمایش در هر تصویر چهره شخص با ایجاد ۲۰ نقطه متناظر (مجموعاً ۴۸۰ نقطه) مجموعه D تشکیل شد.

برای بهبود تصویر مدل، اگر تصویر مدل را از تصاویر واگرا بدست آوریم، مجموعه‌ای از تصاویر روبه‌رو خواهیم داشت. در این حالت نیز تاثیر درونیابی میانگین‌گیری دگردیسی، مانند حالات مختلف چهره بوده و دارای اثرات روح نیز هست. در ابتدای این فصل، در بخش (۳-۲) راجع به روش ترازسازی RASL [۲۳] بحث شد و با آزمایشی نشان داده شد که این روش مقاومتی در مقابل زاویه‌ی چهره ندارد.

به دلیل استفاده از مکاشفه‌ی ساده برای پیکسل‌هایی که به بیرون از صفحه‌ی تصویر نگاشت می‌شوند، عدم اطلاع از تعداد نقاط مکفی و اینکه نمی‌توان در بخش‌های پوشیده شده، نقاط مطابق ایجاد نمود، به وجود آمدن اثرات روح [۲۶] در تصویر خروجی اجتناب ناپذیر است. همچنین ایجاد نقاط مطابق بر روی تعداد زیادی از تصاویر ورودی (به علت مشکل بودن انتخاب دقیق مکان نقاط) دارای خطای مکانی زیادی برای نقاط مطابق می‌باشد.

حال اگر تصاویر بدست آمده به صورت مطرح شده در بالا را با روش RASL نسبت به هم تراز کنیم، اثرات به وجود آمده در خروجی میانگین دگردیسی، خطا محسوب شده و تصاویر مرتبه‌ی پایینی از روبه‌رو بدون اثرات هاله‌ی و حالات مختلف بدست می‌آید. شکل (۳-۱۱) نشان‌دهنده‌ی نتایج میانگین دگردیسی بر روی تصاویر واگرا و سپس ترازسازی آن‌ها با روش RASL نسبت به هم است. پارامترهای روش RASL برای تولید تصاویر شکل (۳-۴) و (۳-۱۱) یکسان تنظیم شده است. با این روش، انسداد زاویه‌ای چهره به حالت‌های چهره و خطاهای با انرژی پایین (هاله) تبدیل شده و قابلیت تصحیح با روش RASL را دارد.

شکل (۳-۱۲) نتیجه‌ی ترازسازی با روش RASL، میانگین دگردیسی و ترکیب میانگین دگردیسی و ترازسازی با روش RASL را به همراه تصویر روبه‌روی پایگاه داده‌ی شخص برای مقایسه‌ی کیفی مورد نمایش قرار می‌دهد.



الف ب

شکل ۳-۱۱. تصاویر خروجی میانگین‌گیری دگردیسی و تصاویر مرتبه‌ی پایین الگوریتم RASL. الف: تصاویر مبدا شکل (۳-۸-الف)، دوبه‌دو میانگین دگردیسی آن‌ها تولید شده و حاصل آن ۱۲ تصویر است. به عنوان مثال تصویر بالا سمت چپ حاصل میانگین دگردیسی تصاویر واگرای $(+30+30)$ و $(-30-30)$ است. ب: تصاویر مرتبه‌ی پایین بعد از ترازسازی با الگوریتم RASL.



شکل ۳-۱۲. مقایسه‌ای میان خروجی روش‌های مطرح شده. تصویر بالا سمت چپ میانگین تصاویر تراز شده با الگوریتم RASL با تصاویر ورودی شکل (۳-۸-الف) است. تصویر بالا سمت راست خروجی میانگین‌گیری دگردیسی بر روی همان تصاویر ورودی است. تصویر پایین سمت چپ: میانگین تصاویر شکل (۳-۱۳-ب) است. تصویر پایین سمت راست تصویر روبه‌روی پایگاه داده است.

۳-۸- جمع‌بندی

در این فصل روش پیشنهادی مطرح شده و به جزئیات آن پرداخته شد. در ابتدا ملزومات و تکنیک‌هایی که در روش پیشنهادی مورد استفاده قرار گرفته بود شرح داده شد. از جمله‌ی این ملزومات، انتقال نسبی و تکنیک‌های دگردیسی و پیچ و تاب تصویر است. همچنین روش RASL [۲۳] که برای ترازسازی نتایج روش پیشنهادی به کار گرفته شد، مفصل شرح داده شد. پس از آن، تحلیلی راجع به اثر انتقال پیکسل-های تصاویر با روش پیشنهادی انجام داده شد و یک مورد مقایسه‌ی کیفی از نتایج روش‌های مختلف نیز به تصویر در آمد. در فصل آینده، مقایسات و ارزیابی روش پیشنهادی به صورت کمی مطرح می‌شود.

فصل چهارم

جرتیات پیکربندی و ارزیابی روش پیشنهادی

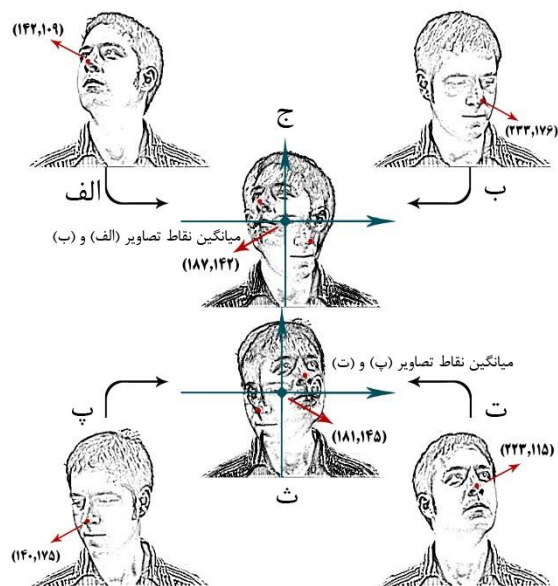
مقدمه

در این فصل، در ابتدا به تحلیل اثر روش پیشنهادی پرداخته می‌شود. سپس برای مدل‌سازی چهره با استفاده از روش پیشنهادی، نحوه‌ی پیکربندی روش RASL گزارش می‌شود. پس از مدل روبه‌روی چهره، برای تمامی افراد پایگاه داده ساخته می‌شود. سپس زاویه‌ی تمامی تصاویر مدل شده، با استفاده از روشی که قادر به سنجش زاویه چهره در تصاویر دوبعدی است، اندازه‌گیری خواهد شد. در ادامه میزان شباهت تصاویر مدل شده با تصویر روبه‌روی پایگاه داده، تحت معیار شباهتی سنجیده خواهد شد تا کارایی روش پیشنهادی را در سیستم‌های شناسایی نشان دهد. در انتها نیز به بحث و نتیجه‌گیری خواهیم پرداخت.

۴-۱- تحلیل اثر روش پیشنهادی

برای تحلیل اثر این الگوریتم، $i_1^{+y_1+x_1}$ و $i_2^{-y_1-x_1}$ را از مجموعه‌ی تصاویر C (مجموعه‌ی ورودی که در ابتدای بخش معرفی شد) در نظر بگیرید. این تصاویر را واگرا می‌نامیم. بردارهای D در رابطه‌ی (۳-۱۲) مطابق با مکان نقاط در تصویر روبه‌روی چهره می‌باشند. شکل (۴-۱) این مطلب را نشان می‌دهد. در شکل (۴-۱)، بر روی چهار تصویر واگرا، با زوایای $(+30^\circ+30^\circ)$ ، $(+30^\circ-30^\circ)$ ، $(-30^\circ+30^\circ)$ و $(-30^\circ-30^\circ)$ ، یک نقطه بر روی بینی ایجاد شده است. پس مجموعه‌ی D دارای یک نقطه است. اگر محور مختصات را بر روی این نقطه قرار دهیم، نقاط مطابق در مبدا محور مختصات تصاویر واگرا، نسبت به هم قرینه خواهند بود. اگر دو تصویر واگرا را به عنوان ورودی الگوریتم در نظر بگیریم، در گذر از یک تصویر واگرا به تصویر دیگر، یک تصویر چهره از روبه‌رو که حداکثر دارای مقداری زاویه از نوع اول مانند شکل (۳-۶-الف) است، ایجاد می‌شود [۳۳]. این تصویر روبه‌رو در میانگین مکان نقاط مطابق دو تصویر واقع می‌شود. چون جابه‌جایی در میانگین‌گیری تاثیری ندارد، می‌توانیم به ورودی‌های الگوریتم به صورت دوبه‌دو واگرا نگاه کنیم. در این

صورت اثر انتقال با الگوریتم میانگین‌گیری دگردیسی، به میانگین تصاویر روبه‌روی بدست آمده از تصاویر مبدا و مقصد واگرا، توجیه می‌شود.



شکل ۴-۱. چهار تصویر (الف)، (ب)، (پ) و (ت) به ترتیب با زوایای $(+30+30)$ ، $(-30-30)$ ، $(-30+30)$ و $(+30-30)$. تصاویر (الف) و (ب) جفت واگرای هم هستند. تصاویر (پ) و (ت) نیز همین‌طور. از روی مختصات بینی در تصاویر (ث) و (ج) که به ترتیب مکان نگاشت نقطه در تصویر مدل برای جفت‌های (الف و ب) و (پ و ت) را نشان می‌دهد، مشخص است که بینی در خروجی به صورت تقریبی، به یک نقطه نگاشت می‌شود و حذف دو تصویر واگرا از مجموعه باعث جابه‌جایی نقاط مقصد، D، نخواهد شد.

چون زاویه نوع سوم مانند شکل (۳-۶-الف) از چرخش سر حول محور گردن به وجود می‌آید، نحوه‌ی جابه‌جایی نقاط مختلف چهره با توجه به فاصله‌ی آن‌ها تا محور گردن مشخص می‌شود. در میان مجموعه نقاط مطابق، آنهایی که دارای فاصله‌ی بیشتری از هم هستند، بر روی مداری دورتر از محور گردن خواهند بود (برای مثال در چرخش سر بر محور گردن در راستای افقی و عمودی، میزان جابه‌جایی نوک بینی بیشتر از چشم یا گوش می‌باشد). پس نسبتی مستقیم میان فواصل نقاط مطابق و فاصله‌ی آن‌ها با محور گردن وجود دارد. در انتقال با استفاده از شبکه‌ی عصبی، فاصله‌ی بین نقاط مطابق رابطه مستقیم با اندازه‌ی حرکت نقطه در تصویر مدل دارد. به عبارت دیگر، در تصویر مدل، نقاط دور از هم، بیشتر از نقاط

نزدیک به هم تغییر مکان می‌دهند. اگر تابع انتقال به خوبی عمل درونیایی را تحت تابعی هموار یاد بگیرد به منزله‌ی این است که الگوی اندازه‌ی جابه‌جایی نقاط مطابق نسبت به هم را آموزش دیده است. همچنین چون این شبکه‌ی عصبی تابعی پیوسته می‌باشد، تصاویر میانی که تولید می‌کند، دارای زاویه‌های مختلف نسبت به دوربین می‌باشند. برای نشان دادن نحوه حرکت نقاط، یک فرایند دگردیسی با رابطه‌ی (۳-۷) و استفاده از توابع انتقال رابطه‌ی (۳-۹) تشکیل دادیم و دو تصویر چهره با زوایای $(-30^\circ, -0^\circ)$ و $(+30^\circ, +0^\circ)$ را به ترتیب به عنوان تصاویر مبدا و مقصد دگردیسی در نظر گرفتیم. شکل (۴-۲) نحوه‌ی حرکت نقاط در تصاویر میانی تولید شده توسط فرایند دگردیسی را نشان می‌دهد. در شکل (۴-۲) تصاویر (پ ۱)، (پ ۲) و (پ ۳) بدست آمده از دگردیسی به ترتیب در $t = 1/4$ ، $t = 1/2$ و $t = 3/4$ نشان داده شده است. با توجه به نحوه‌ی انتقال توابع پیچ و تاب رابطه‌ی (۳-۹) انتظار می‌رود تصاویر تولید شده به ترتیب دارای زوایای $(-15^\circ, +0^\circ)$ ، $(+0^\circ, +0^\circ)$ و $(+15^\circ, +0^\circ)$ باشند. به همین خاطر تصاویر چهره‌ی شخص از پایگاه داده با همان زاویه‌ها، برای مقایسه در شکل (۴-۲) آورده شده‌اند. جدول (۴-۱) مختصات دو نقطه‌ی مطابق قرار داده شده بر روی بینی و چشم راست شخص در تصاویر شکل (۴-۲) را نشان می‌دهد. همچنین در جدول (۴-۱)، اختلاف نقاط مطابق در تصاویر متناظر که در بالا مطرح شد، در دو راستای محور مختصات و فاصله‌ی اقلیدسی آن‌ها نیز درج شده است.



شکل ۴-۲. نمایش نحوه حرکت نقاط در تصاویر میانی دگردیسی با تابع انتقال (۳-۹). (الف) و (ت) به ترتیب تصاویر مبدا و مقصد هستند. در این شکل (ب ۱)، (ب ۲) و (ب ۳) به ترتیب تصاویر با زوایای (+۰+۱۵)، (+۰+۳۰) و (+۰-۱۵) از پایگاه داده هستند. انتظار داریم در تصاویر (ب ۱)، (ب ۲) و (ب ۳) و تصاویر متناظر آن‌ها یعنی (پ ۱)، (پ ۲) و (پ ۳) که به ترتیب حاصل دگردیسی در $t = 1/4$, $t = 1/2$, $t = 3/4$ هستند، نقاط اختلاف مکانی ناچیزی داشته باشند.

جدول ۱-۴. مقایسه مکان نقاط در تصاویر متناظر. در این جدول، مختصات نقاط تصاویر متناظر، در شکل (۳-۱۲) برای مقایسه در یک ستون درج شده است. در این جدول اختلاف نقاط دو راستای محور مختصات و همچنین فاصله اقلیدسی آن‌ها درج شده است.

تصاویر	ب ۱	فاصله میان	ب ۲	فاصله میان	ب ۳	فاصله میان
	پ ۱	نقاط	پ ۲	نقاط	پ ۳	نقاط
مختصات چشم	(۲۲۶،۱۲۳)	(۲،۱)	(۲۱۴،۱۲۲)	(۲،۳)	(۱۹۹،۱۲۴)	(۱،۲)
	(۲۲۴،۱۲۴)		(۲۱۲،۱۲۵)		(۱۹۸،۱۲۶)	
مختصات بینی	(۲۰۲،۱۵۳)	(۳،۰)	(۱۸۵،۱۵۶)	(۳،۳)	(۱۶۸،۱۵۶)	(۲،۳)
	(۲۰۵،۱۵۳)		(۱۸۸،۱۵۳)		(۱۷۰،۱۵۳)	
فاصله اقلیدسی میان نقاط بینی	۲.۲۳۶		۳.۶۰۶		۲.۲۳۶	
فاصله اقلیدسی میان نقاط چشم	۳		۴.۲۴۲		۳.۶۰۶	

۴-۲- تنظیم پارامترها برای روش پیشنهادی

در ابتدا به پیکربندی الگوریتم RASL [۲۳] می‌پردازیم. در روش RASL پارامتر λ مصالحه‌ای میان مرتبه‌ی ماتریس و میزان تنگی خطا برقرار می‌کند. در آزمایشات این پژوهش مقدار λ به ۰.۰۱ تنظیم شد. همچنین روش RASL ترازسازی تصاویر را نسبت به دو نقطه‌ی کانونی و فاصله‌ی میان آن‌ها، در یک قاب کانونی انجام می‌دهد. برای انجام آزمایش، ابعاد قاب کانونی را 100×100 قرار دادیم. نقاط کانونی را به گوشه‌های داخلی چشم تنظیم نموده و فاصله‌ی کانونی این نقاط را ۳۲ پیکسل قرار دادیم. از دیگر پارامترهای روش RASL تعیین نوع تابع انتقال است. تابع انتقال روش RASL، به انتقال نسبی برای ترازسازی تنظیم شد. از تصاویر مرتبه‌ی پایین، (I_i^0) ، تصویری که دارای خطای تنک کمتری (e_i) در نرم صفر (l^0) باشد به عنوان بهترین تصویر روبه‌رو انتخاب می‌شود. تعداد تصاویر ورودی الگوریتم برای تمامی اشخاص، ۲۴ تصویر با تغییر زاویه از $(-30^\circ - 30^\circ)$ تا $(+30^\circ + 30^\circ)$ مانند شکل (۳-۸-الف) در نظر گرفته می‌شود. برای استخراج چهره، از مرکز چهره به همراه اندازه‌ی ابعادی برای قاب دور چهره که پایگاه داده در برچسب تصاویر در اختیار گذاشته بود استفاده شد [۸]. برای انتخاب نقاط مطابق در تصاویر مبدا، از نواحی گوشه‌های چشم، ابروها، گوشه‌های دهان، بینی و حاشیه‌ی صورت استفاده شد. این عمل به صورت دستی انجام گرفت.

۴-۳- محدوده‌ی زاویه‌ی چهره‌ی قابل تصحیح توسط روش پیشنهادی

بحث راجع به اینکه روش پیشنهادی قابلیت تحمل زاویه چهره در چه محدوده‌ای را دارد، وابسته به پارامترهای زیادی است. از جمله‌ی این پارامترها می‌توان به میزان بستگی خطی تصاویر، پس از مدل-سازی تصاویر و اگر اشاره نمود. از دیگر پارامترها گام تغییر زاویه است. مسلم است که هرچه تعداد تصاویر بیشتر باشد، تصویر مدل شده به تصویر ایده‌آل شبیه‌تر خواهد بود. در یک سناریو بر روی تصاویر پایگاه

داده‌ی PRIMA، از مجموعه تصاویر در بازه‌های $(-15, 15)$ ، $(-30, 30)$ ، $(-60, 60)$ و $(-90, 90)$ به ترتیب با تعداد تصاویر ورودی ۸، ۲۴، ۸۶ و ۹۲ تصویر روبه‌رو مدل سازی شد. سپس شباهت تصویر مدل شده با تصویر روبه‌روی پایگاه داده تحت رابطه‌ی (۴-۱) (تفاضل نرمال شده) سنجیده شد. این آزمایشات بر روی ماشین با پردازنده‌ی core i5 با فرکانس ۲.۸GHz و مقدار حافظه‌ی اصلی ۶GB صورت گرفت و مقدار زمان ترازسازی ثبت شد. همانطور که قبلاً ذکر شد، چون ایجاد نقاط مطابق بر روی تصاویر ورودی به صورت دستی انجام شد، از سنجش زمان این مرحله اجتناب شد. خوش‌بختانه روش‌هایی وجود دارند که مکان‌یابی بخش‌های چهره را به صورت بلادرنگ انجام می‌دهند. راجع به این روش‌ها در فصل پنجم بحث شده است. جدول (۴-۲) شباهت تصویر مدل با تصویر روبه‌ی پایگاه داده برای ۵ شخص که به صورت تصادفی از پایگاه داده انتخاب شده‌اند به همراه زمان ترازسازی را نشان می‌دهد.

$$E(I_1^{m \times n}, I_2^{m \times n}) = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n |I_{1,i,j} - I_{2,i,j}| \quad (۴-۱)$$

جدول ۴-۲ مشخص‌سازی محدوده‌ی مناسب زاویه‌ی چهره برای مدل‌سازی در روش پیشنهادی

شخص	محدوده‌ی زاویه $(-60, 60)$		محدوده‌ی زاویه $(-60, 60)$		محدوده‌ی زاویه $(-30, 30)$		محدوده‌ی زاویه $(-15, 15)$	
	تعداد تصاویر ورودی ۹۸ عدد	زمان ترازسازی	میزان شباهت	زمان ترازسازی	تعداد تصاویر ورودی ۲۴ عدد	زمان ترازسازی	میزان شباهت	تعداد تصاویر ورودی ۸ عدد
دوم	۸۶,۲۴ ثانیه	۰,۱۰۶۴	۶۹ ثانیه	۰,۱۰۶۹	۱۴,۳ ثانیه	۰,۱۱۳۷	۶,۲۱ ثانیه	۰,۱۴۵۶
چهارم	۸۶,۲۴ ثانیه	۰,۱۱۵۰	۶۹,۱ ثانیه	۰,۱۱۵۹	۱۴,۲۹ ثانیه	۰,۱۲۰۶	۶,۲۳ ثانیه	۰,۱۵۳۲
هفتم	۸۶,۲۵ ثانیه	۰,۰۸۱۲	۶۹,۱ ثانیه	۰,۰۸۲۳	۱۴,۳۱ ثانیه	۰,۰۸۶۹	۶,۲۲ ثانیه	۰,۱۱۲۹
یازدهم	۸۶,۲۵ ثانیه	۰,۱۰۳۴	۶۹,۱ ثانیه	۰,۱۰۳۹	۱۴,۳۱ ثانیه	۰,۱۱۰۳	۶,۲۱ ثانیه	۰,۱۴۸۹
دوازدهم	۸۶,۲۵ ثانیه	۰,۱۱۷۰	۶۹ ثانیه	۰,۱۲۰۱	۱۴,۳۱ ثانیه	۰,۱۲۶۰	۶,۲۱ ثانیه	۰,۱۶۰۱

با توجه به جدول (۴-۲) تعداد تصاویر ورودی برای مدل‌سازی در این پایان‌نامه ۲۴ عدد در محدوده‌ی $(-30, 30)$ در نظر گرفته شد.

۴-۴- چارچوب آزمایش و نتایج کمی

در این بخش ابتدا زاویه‌ی چهره سنجیده شده و سپس میزان شباهت چهره‌ی مدل شده با چهره‌ی روبه-روی پایگاه داده گزارش می‌شود. روش‌های زیادی برای تخمین زاویه‌ی چهره وجود دارند [۳۴,۳۵]. برای تخمین زاویه‌ی چهره از روش مبتنی بر رگرسیون احتمالی با ابعاد بالا [۳۴] استفاده نمودیم. پس از آموزش روش [۳۴] بر روی پایگاه داده‌ی PRIMA با تعداد تبدیلات نسبی، $k = 50$ ، زاویه تصاویر مدل شده سنجیده شد. جدول (۳-۴) نشان‌دهنده‌ی زاویه‌ی چهره برای تصاویر مدل شده است که با استفاده از روش [۳۴] تخمین زده شده است.

جدول ۳-۴. جدول زاویه‌های سنجیده شده در تصاویر مدل شده پایگاه داده PRIMA با استفاده از روش [۳۴]

زاویه	شخص	زاویه	شخص	زاویه	شخص
(+++)	یازدهم	(+++)	ششم	(+۰-۵)	اول
(+۰+۵)	دوازدهم	(+++)	هفتم	(+++)	دوم
(+++)	سیزدهم	(+++)	هشتم	(+++)	سوم
(+++)	چهاردهم	(+++)	نهم	(+++)	چهارم
(+۰-۵)	پانزدهم	(+++)	دهم	(+++)	پنجم

همان‌طور که مشاهده می‌شود به غیر از اشخاص اول، دوازدهم و پانزدهم، که تنها ۵ درجه زاویه در راستای محور افقی برای چهره‌ی آن‌ها تخمین زده شده، برای باقی اشخاص، زاویه‌ی چهره مقدار (+++۰) بدست آمده است. در آزمایشی دیگر خروجی الگوریتم را با چهره‌ی روبه‌رو مقایسه نمودیم. برای سنجش میزان شباهت دو تصویر، از رابطه‌ی (۴-۱) استفاده شد. جهت سنجش مقاومت الگوریتم، از ۲۴ تصویر ورودی، ۶ شبیه‌ترین تصویر به تصویر وسط را برای هر شخص، با زوایای $(+۱۵+۰)$ ، $(-۱۵+۰)$ ، $(+۰+۱۵)$ ،

(+۰-۱۵)، (+۰+۳۰) و (+۰+۳۰) از ورودی الگوریتم حذف نمودیم. از ۱۸ تصویر ورودی باقی مانده، ۴ شبیه‌ترین تصویر به تصویر چهره‌ی روبه‌رو با زوایای (+۱۵+۱۵)، (+۱۵-۱۵)، (-۱۵+۱۵) و (+۱۵+۱۵) شناسایی شدند. سپس چهره‌ی چهار تصویر مذکور به همراه تصویر روبه‌رو با استفاده از نقاط و فاصله‌ی کانونی ذکر شده نرمال‌سازی شدند. در جدول (۴-۴) میزان شباهت هر یک از این چهار تصویر و تصویر مدل شده، با تصویر روبه‌روی پایگاه داده گزارش شده است.

در جدول (۴-۴) رنگ زرد نشان‌دهنده‌ی تفاوتی کمتر از ۱ درصد، رنگ خاکستری نشان‌دهنده‌ی تفاوت بین ۱ تا ۱۰ درصد و رنگ قرمز نشان‌دهنده‌ی تفاوت بیش از ۱۰ درصد است. جدول (۴-۵) نتایج مطلوب با توجه به میزان پذیرش خطا را نشان می‌دهد. ستون سمت راست این جدول نشان می‌دهد که اگر از خطای ۵ درصد صرف نظر کنیم، در حدود ۹۳ درصد مواقع، مدل بدست آمده، مدلی قابل اعتماد برای شناسایی است.

بیشترین اختلافی که تصویر مدل شده با شبیه‌ترین تصویر به تصویر روبه‌رو دارد، بعد از شخص سوم، مربوط به شخص پانزدهم است که در حدود ۳.۵۸ درصد است. دلیل ساخت مدلی ناکارآمد برای چهره‌ی شخص سوم، وجود مقدار زیادی انسداد (موی شخص که کل پیشانی آنرا پوشش داده) در تمامی تصاویر ورودی این شخص است. ستون سوم جدول نشان‌دهنده‌ی ۱۴ وضعیت مطلوب از ۱۵ مورد است. شکل (۳-۴) نیز تصاویر مدل شده‌ی پایگاه داده را به همراه تصویر روبه‌رو نمایش می‌دهد.

جدول ۴-۴. جدول مقایسات چهار شبیه‌ترین تصویر با زوایای مطرح شده به همراه تصویر مدل شده با تصویر روبه‌روی پایگاه داده. هر سطر نشان‌دهنده‌ی نتیجه‌ی مقایسات یک شخص است. ■ نشان‌دهنده‌ی شبیه‌ترین به تصویر روبه‌روی پایگاه داده است. ■ نشان‌دهنده‌ی اختلاف با دقت دو رقم اعشار در محدوده‌ی (0.001, 0.01) با شبیه‌ترین تصویر به تصویر روبه‌رو است. ■ نشان‌دهنده‌ی اختلاف با یک رقم اعشار در محدوده‌ی (0.01, 0.1) است. ■ نشان‌دهنده‌ی دقت بدون رقم اعشار در محدوده‌ی (0.1, 1) است. همچنین برای تصویر هر شخص، رتبه‌ی آن بر اساس شباهت با تصویر روبه‌رو مشخص شده است. A دارای بیشترین شباهت و E دارای کمترین شباهت است.

زاویه شخص	(۱۵+۱۵)		(۱۵-۱۵)		(۱۵+۱۵)		(۱۵-۱۵)		تصویر مدل شده	
اول	۰.۱۱۲۱	A	۰.۱۴۰۰	D	۰.۱۲۹۸	C	۰.۱۶۵۱	E	۰.۱۱۳۲	B
دوم	۰.۱۲۳۷	D	۰.۱۲۳۰	C	۰.۱۰۲۷	A	۰.۱۴۱۵	E	۰.۱۱۳۷	B
سوم	۰.۰۸۵۰	A	۰.۰۹۲۸	C	۰.۱۳۶۴	D	۰.۰۹۱۲	B	۰.۳۰۷۰	E
چهارم	۰.۱۲۲۳	C	۰.۱۵۲۵	E	۰.۱۴۷۵	D	۰.۱۰۰۸	A	۰.۱۲۰۶	B
پنجم	۰.۱۱۰۲	C	۰.۱۲۸۵	E	۰.۱۲۴۶	D	۰.۱۰۹۴	B	۰.۱۰۳۳	A
ششم	۰.۱۲۴۱	B	۰.۱۳۸۲	D	۰.۱۷۳۱	E	۰.۱۲۶۷	C	۰.۰۷۲۷	A
هفتم	۰.۱۳۱۹	C	۰.۱۳۴۶	D	۰.۱۴۷۵	E	۰.۰۸۶۹	A	۰.۰۸۹۹	B
هشتم	۰.۱۱۰۶	A	۰.۱۲۱۸	B	۰.۱۷۱۹	E	۰.۱۲۶۸	C	۰.۱۳۰۹	D
نهم	۰.۰۸۸۰	B	۰.۱۰۹۵	D	۰.۰۶۹۳	A	۰.۱۱۱۴	C	۰.۰۹۳۴	E
دهم	۰.۱۰۴۸	A	۰.۱۳۳۶	C	۰.۱۵۶۳	E	۰.۱۳۵۰	D	۰.۱۰۵۸	B
یازدهم	۰.۰۹۳۳	A	۰.۱۰۴۶	C	۰.۱۳۴۰	E	۰.۱۱۰۳	D	۰.۱۰۳۹	B
دوازدهم	۰.۰۸۱۵	A	۰.۱۱۹۶	C	۰.۱۳۰۱	E	۰.۱۲۶۰	D	۰.۱۰۱۶	B
سیزدهم	۰.۰۹۱۴	B	۰.۱۰۷۹	C	۰.۱۲۹۶	E	۰.۱۱۵۰	D	۰.۰۸۹۶	A
چهاردهم	۰.۱۲۲۳	D	۰.۱۲۹۴	E	۰.۱۱۰۱	C	۰.۱۰۰۵	B	۰.۰۹۸۳	A
پانزدهم	۰.۱۳۵۹	D	۰.۱۱۲۷	B	۰.۱۱۹۰	C	۰.۱۰۵۵	A	۰.۱۴۱۳	E

جدول ۴-۵. درصد نتایج مطلوب نسبت به پارامترهای مطرح شده در عنوان ستون‌ها

پذیرش خطای کمتر از ۵ درصد با شبیه‌ترین تصویر	داشتن رتبه‌ی A یا B	شبیه‌ترین تصویر به تصویر روبه‌رو
۹۳.۳۳٪	۷۳.۳۳٪	۲۶.۶۷٪



شکل ۴-۴. تصاویر مدل شده‌ی پایگاه داده. سطر اول تصاویر چهره با زاویه‌ی (+++) اشخاص پایگاه داده بوده و سطر زیرین آن، تصاویر مدل شده قرار گرفته است. این ترتیب در سطرهای بعدی تصویر نیز جاری است. اولین سطر این تصویر برای اشخاص اول تا پنجم پایگاه داده به ترتیب از چپ به راست است. سطرهای بعدی به ترتیب برای اشخاص ششم تا پانزدهم پایگاه داده هستند.

۴-۵- جمع‌بندی

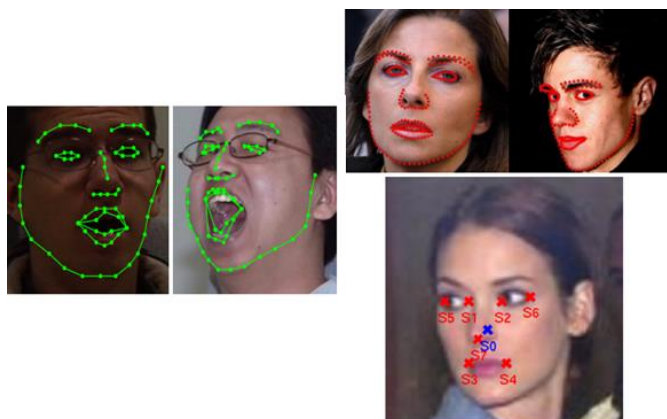
در این فصل به ارزیابی روش پیشنهادی پرداخته شد. در ابتدا تحلیلی راجع به نحوه‌ی انتقال روش پیشنهادی ارائه شد. سپس نحوه‌ی تنظیم پارامترهای روش RASL و پیکربندی روش پیشنهادی مطرح شد. در ادامه تمامی تصاویر چهره‌ی پایگاه داده‌ی PRIMA با روش پیشنهادی، از تصاویر ورودی مطرح شده مدل شدند. در ادامه زاویه چهره‌های مدل شده توسط روش [۳۴] سنجیده شد. پس از آن با استفاده از معیار سنجش شباهت، چهره‌های مدل شده با چارچوب مطرح شده مورد آزمایش قرار گرفتند.

فصل پنجم

بحث و نتیجه گیری

۵-۱- بحث

پایگاه داده‌های زیادی همچون FERET وجود دارند که برای آزمایشات روش پیشنهادی این مقاله مناسب می‌باشند. از میان این پایگاه داده‌ها، PRIMA دارای دقت کمتری در اندازه‌گیری زاویه‌ی چهره نسبت به سایر پایگاه‌های داده بود و تنها حاوی تصاویر دو بعدی ساده است [۳۴]. با توجه به عملکرد مناسب روش پیشنهادی بر روی پایگاه داده‌ی PRIMA، می‌توان به نتایج این روش در تصاویر ویدئویی نصب شده در درگاه‌های ورودی اتکا نموده و همچنین از تصاویر تقریباً واگرا (تصاویری که در گذر از یکی به دیگری یک تصویر چهره‌ی روبه‌رو وجود دارد) نیز استفاده نمود. با استفاده از چهارچوب ساخت تابع (۳-۹) قابل اثبات است که از یک تابع چندجمله‌ای بدست آمده با حداقل مربعات نیز می‌توان به عنوان تابع انتقال استفاده نمود. همچنین با استفاده از روش‌هایی همچون [۴۰-۳۵]، بخش ایجاد نقاط مطابق، قابلیت خودکار سازی دارد. می‌توان با استفاده از روش‌هایی مانند [۳۵] که مرکز صورت را مشخص می‌کنند، در طی فرایند برقراری تناظر بین نقاط، با میانگین‌گیری مکانی، جفت‌های واگرا را پیدا نمود و بخش ایجاد نقاط مطابق را خودکار نمود. شکل (۵-۱) نمونه‌ای از خروجی این روش‌ها برای تصاویر چهره را نشان می‌دهد.



شکل ۵-۱. نمونه‌ای از خروجی روش‌های مکان‌یابی بخش‌های چهره [۴۰-۳۵]

۵-۲- نتیجه

در این پایان‌نامه روشی برای مدل‌سازی تصویر چهره‌ی شخص از روبه‌رو، با استفاده از مجموعه تصاویر زاویه‌دار ارائه شد. این روش میانگین‌گیری دگردیسی نامیده شد و نشان داد که قادر به نرمال‌سازی تصاویر چهره نسبت به زاویه می‌باشد. همچنین نشان داده شد که با استفاده از تصاویر دوبه‌دوی واگرا و با میانگین‌گیری دگردیسی، قادریم که انسداد زاویه‌ای را به حالات چهره تبدیل کنیم. با تبدیل خطای تصویر، امکان تصحیح آن با استفاده از تجزیه‌ی مرتبه‌ی پایین RASL فراهم شد. نتایج کمی و کیفی نشان می‌دهند که روش پیشنهادی به صورتی مناسب تصویر روبه‌رو را از تصاویر زاویه‌دار می‌سازد. این روش از اطلاعات زاویه‌ی چهره استفاده نمی‌کند و قادر به مدل‌سازی تصویر چهره از تصاویر ساده‌ی دوبعدی که ۳۰ درجه زاویه در هر دو جهت دارند، می‌باشد. سعی بر آن داریم تا در آینده این مساله‌ی حداقل مربعات را با تابع هدف الگوریتم RASL در قالب یک معادله‌ی بهینه‌سازی توامان، یکپارچه‌سازی نماییم و با خودکارسازی ایجاد نقاط مطابق، روش پیشنهادی را کاملاً خودکار سازیم.

- [1] Hjeltnæs, E. and Low, B.K., 2001. Face detection: A survey. *Computer vision and image understanding*, 83(3), (pp.236-274).
- [2] Viola, P. and Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-I). IEEE.
- [3] Mairal, J., Bach, F. and Ponce, J., 2014. Sparse modeling for image and vision processing. *Foundations and Trends® in Computer Graphics and Vision*, 8(2-3), (pp.85-283).
- [4] Zhang, Q. and Li, B., 2010, June. Discriminative K-SVD for dictionary learning in face recognition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (pp. 2691-2698). IEEE.
- [5] Xu, Y., Zhang, Z., Lu, G. and Yang, J., 2016. Approximately symmetrical face images for image preprocessing in face recognition and sparse representation based classification. *Pattern Recognition*, 54, (pp.68-82).
- [6] Yang, J., Wright, J., Huang, T. and Ma, Y., 2008, June. Image super-resolution as sparse representation of raw image patches. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1-8). IEEE.
- [7] Tang, Y., Yuan, Y., Yan, P. and Li, X., 2013. Greedy regression in sparse coding space for single-image super-resolution. *Journal of visual communication and image representation*, 24(2), (pp.148-159).
- [8] "PRIMA," [Online]. Available: <http://www-prima.inrialpes.fr/perso/Gourier/Faces/HPDatabase.html>. [Accessed on Tuesday, June 14, 2016]
- [9] Haghighat, M., Abdel-Mottaleb, M. and Alhalabi, W., 2016. Fully automatic face normalization and single sample face recognition in unconstrained environments. *Expert Systems with Applications*, 47, (pp.23-34).
- [10] van Beek, P.J., Reinders, M.J., Sankur, B. and van der Lubbe, J.C., 1992, November. Semantic segmentation of videophone image sequences. In *Applications in Optical Science and Engineering. International Society for Optics and Photonics*, (pp. 1182-1193).
- [11] Wong, C., Kortenkamp, D. and Speich, M., 1995, November. A mobile robot that recognizes people. In *Tools with Artificial Intelligence, 1995. Proceedings, Seventh International Conference on* (pp. 346-353). IEEE.

- [12] Reisfeld, D. and Yeshurun, Y., 1992, August. Robust detection of facial features by generalized symmetry. In *International Conference on Pattern Recognition* (pp. 117-117). *IEEE COMPUTER SOCIETY PRESS*.
- [13] Maio, D. and Maltoni, D., 2000. Real-time face location on gray-scale static images. *Pattern Recognition*, 33(9), (pp.1525-1539).
- [14] Wu, H., Yokoyama, T., Pramadiahanto, D. and Yachida, M., 1996, October. Face and facial feature extraction from color image. In *Automatic Face and Gesture Recognition, 1996. Proceedings of the Second International Conference on* (pp. 345-350). IEEE.
- [15] Sirovich, L. and Kirby, M., 1987. Low-dimensional procedure for the characterization of human faces. *Josa a*, 4(3), (pp.519-524).
- [16] Rowley, H.A., Baluja, S. and Kanade, T., 1998. Neural network-based face detection. *IEEE Transactions on pattern analysis and machine intelligence*, 20(1), (pp.23-38).
- [17] Guo, T., Tan, X.H. and Chao, C.X., 2016, January. Robust Face Recognition via Occlusion Detection and Masking. In *MATEC Web of Conferences* (Vol. 44). EDP Sciences.
- [18] Juefei-Xu, F. and Savvides, M., 2015, September. Encoding and decoding local binary patterns for harsh face illumination normalization. In *Image Processing (ICIP), 2015 IEEE International Conference on* (pp. 3220-3224). IEEE.
- [19] Elad, M. and Aharon, M., 2006. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12), (pp.3736-3745).
- [20] Rubinstein, R., Zibulevsky, M. and Elad, M., 2008. Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit. *Cs Technion*, 40(8), (pp.1-15).
- [21] Wolberg, G., 1988. *Geometric transformation techniques for digital images: a survey*. Department of Computer Science, Columbia University.
- [22] Bloomenthal, J. and Rokne, J., 1994. Homogeneous coordinates. *The Visual Computer*, 11(1), (pp.15-26).
- [23] Peng, Y., Ganesh, A., Wright, J., Xu, W. and Ma, Y., 2012. RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11), (pp.2233-2246).
- [24] Zhang, Z., Xu, Y., Yang, J., Li, X. and Zhang, D., 2015. A survey of sparse representation: algorithms and applications. *IEEE Access*, 3, (pp.490-530).

- [25] Arad, N., Dyn, N., Reisfeld, D. and Yeshurun, Y., 1994. Image warping by radial basis functions: Application to facial expressions. *CVGIP: Graphical models and image processing*, 56(2), (pp.161-172).
- [26] Beier, T. and Neely, S., 1992, July. Feature-based image metamorphosis. In *ACM SIGGRAPH Computer Graphics* (Vol. 26, No. 2, pp. 35-42). ACM.
- [27] Lee, S.Y., Chwa, K.Y., Hahn, J.K. and Shin, S.Y., 1996. Image morphing using deformation techniques. *Journal of Visualization and Computer Animation*, 7(1), (pp.3-23).
- [28] Prince, T., Fragoso, Y. and Gaye, O., 2016. Applications of Image Morphing Techniques to Analyze Changes in Our Environment. *Open Access Library Journal*, 3(09), p.1.
- [29] Baker, S. and Matthews, I., 2004. Lucas-kanade 20 years on: A unifying framework. *International journal of computer vision*, 56(3), (pp.221-255).
- [30] Brown, L.G., 1992. A survey of image registration techniques. *ACM computing surveys (CSUR)*, 24(4), (pp.325-376).
- [31] Fanelli, G., Dantone, M., Gall, J., Fossati, A. and Van Gool, L., 2013. Random forests for real time 3D face analysis. *International Journal of Computer Vision*, 101(3), (pp.437-458).
- [32] Jain, Hardik, Olaf Hellwich, and R. S. Anand. "Improving 3D face geometry by adapting reconstruction from stereo image pair to generic Morphable Model." In *Information Fusion (FUSION), 2016 19th International Conference on*, (pp. 1720-1727). IEEE, 2016.
- [33] Seitz, S.M. and Dyer, C.R., 1997. View morphing: Uniquely predicting scene appearance from basis images. In *Proc. Image Understanding Workshop* (pp. 881-887).
- [34] Drouard, V., Ba, S., Evangelidis, G., Deleforge, A. and Horaud, R., 2015, September. Head pose estimation via probabilistic high-dimensional regression. In *Image Processing (ICIP), 2015 IEEE International Conference on* (pp. 4624-4628). IEEE.
- [35] Zhu, X. and Ramanan, D., 2012, June. Face detection, pose estimation, and landmark localization in the wild. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 2879-2886). IEEE.
- [36] Asthana, A., Zafeiriou, S., Tzimiropoulos, G., Cheng, S. and Pantic, M., 2015. From pixels to response maps: Discriminative image filtering for face alignment in the wild. *IEEE transactions on pattern analysis and machine intelligence*, 37(6), (pp.1312-1320).
- [37] Lee, D., Park, H. and Yoo, C.D., 2015, June. Face alignment using cascade gaussian process regression trees. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4204-4212). IEEE.

- [38] Zhang, Z., Luo, P., Loy, C.C. and Tang, X., 2016. Learning deep representation for face alignment with auxiliary attributes. *IEEE transactions on pattern analysis and machine intelligence*, 38(5), (pp.918-930).
- [39] Uříčář, M., Franc, V. and Hlaváč, V., 2012. Detector of facial landmarks learned by the structured output SVM. *VIAPP*, 12, (pp.547-556).
- [40] Deng, Z., Li, K., Zhao, Q., Zhang, Y. and Chen, H., 2017. Effective face landmark localization via single deep network. *arXiv preprint arXiv:1702.02719*

Abstract:

In video surveillance, the vision angle of face with respect to camera, called angular occlusion (also referred as head pose) will limit system's ability in face recognition. In this thesis, a method for angular occlusion elimination in face images is proposed, which is based on image morphing. The proposed method models a head-on face from a batch of images with different head pose, which belongs to a person. By simple techniques such as spatial interpolation of input images pixels with translation functions, linear interpolation and averaging from pixels intensity, the proposed method be able to model frontal face. To improvement modeling result, the proposed method apply on divergent face images (face images with symmetry poses) and then by low rank decomposition, modeled faces are aligned with respect to each other. Radial basis function neural network is considered for translation function. No need for depth information, calibrated images and head pose data i.e. (yaw,pitch,roll) that usually used in 3D modelling are some advantages of proposed method. The algorithm's performance on PRIMA dataset is investigated. Experiment results show that the proposed method, as the face-modelling approach using face images with different head poses, will model frontal face image with proper accuracy.

Keywords: angular occlusion, images morphing, warp function, aligning, radial basis function neural network.



Faculty of computer Engineering

MSc Thesis in Artificial Intelligence Engineering

Face Modeling Via Angular Occlusion Dictionary

By:

Sina akhlaghi

Supervisor:

Prof. H. Hassnpour

Advisor:

Dr V. Abolghasemi

Feb 2017