بررسی روش‌های مختلف استخراج ویژگی به منظور طبقه‌بندی تصاویر پرطیفی

نویسندگان: مجید رحیم‌زادگان[۱]، محمدجواد ولدان‌زوج[۲]، مسعود ورشوساز[۳]، یاسر مقصودی مهرانی[۴]
[۱]دانشجوی کارشناسی ارشد دانشگاه صنعتی خواجه نصیرالدین طوسی maj_r2002@yahoo.com
[۲]دانشیار گروه سنجش از دور، دانشکده مهندسی نقشه برداری، دانشگاه صنعتی خواجه نصیر الدین طوسی
valadanzouj@yahoo.com
[۳]استادیار گروه فتوگرامتری، دانشکده مهندسی نقشه برداری، دانشگاه صنعتی خواجه نصیر الدین طوسی
varshosazm@yahoo.com
[۴]کارشناس ارشد سنجش از دور، دانشگاه صنعتی خواجه نصیر الدین طوسی ymaghsoudi@yahoo.com
آدرس: تهران، خیابان ولیعصر، تقاطع میرداماد، دانشکده مهندسی نقشه‌برداری، گروه سنجش‌از دور

چکیده

طبقه‌بندی تصاویر پرطیفی یکی از مهم ترین و پرکاربردترین روش های استخراج اطلاعات در تصاویر ماهواره ای است. استخراج ویژگی یکی از مهم‌ترین روش‌های کاهش تعداد باندها در تصاویر ماهواره ای پرطیفی به منظور انجام یک طبقه بندی مناسب است. در تصاویر ماهواره‌ای پرطیفی، به علت وجود تعداد زیاد باند و همچنین محدود بودن نمونه‌های آموزشی، طبقه‌بندی این تصاویر با مشکل مواجه خواهد شد. این مشکل به علت پدیده Hughes اتفاق می‌افتد. روش‌های مختلف استخراج ویژگی با در نظر گرفتن تعداد ابعاد بهینه برای طبقه‌بندی تصاویر پرطیفی، مشکل به‌وجود آمده به علت پدیدهٔ Hughes را حل می‌نمایند. در این مقاله به بررسی روش‌های استخراج ویژگی در تصاویر ماهواره‌ای پرطیفی پرداخته می شود. نتایج به دست آمده نشان می‌دهد که روش NWFE مناسب‌ترین روش برای استخراج ویژگی در تصاویر ماهواره‌ای پرطیفی می‌باشد.

۱- مقدمه

تصاویر پرطیفی[۱] به علت جمع آوری داده در تعداد ابعاد بسیار بالا نگاشت کاملی از فضای طیفی، در طول موج های مورد نظر، در اختیار ما قرار می دهند. در واقع به علت تراکم بسیار بالای داده در این تصاویر می توان به یک منحنی پیوسته در فضای طیفی، برای هر پیکسل تصویری، دست یافت [۱] و [۲] . در نتیجه در این تصاویر به

---

[1] Hyperspectral images

علت اختلافات کوچک باند های طیفی مجاور، می توان کلاس هایی را که شباهت طیفی بالایی به هم دارنـد، از هـم تفکیک نمود.

استخراج ویژگی یکی از مهمترین روش های کم کردن ابعاد در تصاویر ماهواره ای پرطیفی به منظور طبقـه بندی این تصاویر است [۱] و طبقه‌بندی یکی از مهمترین و در عین حال قوی ترین روش‌های استخراج اطلاعـات از تصاویر ماهواره‌ای است. در واقع طبقه بندی یک فرآیند تصمیم گیری است که مقادیر پیکسل ها در تصاویر اولیـه را برچسب‌گذاری نموده و با کلاس های نهایی جایگزین می کند.

در طبقه بندی تصاویر پرطیفی بایستی پارامترهای آماری مورد نیاز برای کلاس ها برآورد شود. برای برآورد این پارامترهای آماری یک سری نمونه های آموزشی نیاز است. در سنجش از دور این نمونه هـای آموزشـی معمـولاً اندازه محدودی دارند و چون در تصاویر پرطیفی با ابعاد بالا مواجه هستیم، بایـستی بـه طریقـی ابعـاد ایـن تـصاویر کاهش یابد.برای کاهش ویژگی در تصاویر پرطیفـی دو روش وجـود دارد [۱]: (۱) انتخـاب ویژگـی[۱] (۲) اسـتخراج ویژگی[۲]

در انتخاب ویژگی تعدادی از ویژگی ها که دارای بیشترین اطلاعات و کمترین نـویز هـستند، انتخـاب مـی شود، در حالی که در استخراج ویژگی، از تمام ویژگی هـای موجـود، اسـتفاده شـده و بـا اسـتفاده از مینـیمم کـردن ماتریس پراکندگی داخل کلاسی و ماکزیمم کردن ماتریس پراکندگی بین کلاسی نمونه های موجود، تعداد ویژگی ها کاهش می‌یابد. مسلماً روش های استخراج ویژگی به علـت اسـتفاده از اطلاعـات تمـام بانـدها، اطلاعـاتی در مـورد جدایی کلاس ها از دست نمی دهند، زیرا حتی باندهای نویزی هم، فاقد اطلاعات نمی باشند.

یکی از روش های کاهش ابعاد در تصاویر پرطیفی، استخراج ویژگی اسـت کـه در آن بایـستی پارامترهـای آماری به روشی تخمین زده شوند. در واقع در این تصاویر به علت کمبود نمونه‌های آموزشی، بایـستی از روش‌هـای برآورد کوواریانس با استفاده از نمونه‌های آموزشی کوچک پرداخته شود. برای انجام این برآورد هـا، از روش هـایی مانند RDA[۳] و LOOC[۴] استفاده می شود [۳] و [۱۳].

در بخش اول این مقاله در مقدمه توضیحاتی راجع به پدیدهٔ Hughes بیان شده، زیرا پدیـده Hughes علـت اصلی پرداختن به روش‌های استخراج ویژگی است. در بخش دوم به بیان توضـیحاتی راجـع بـه طبقـه‌بنـدی کننـده پرداخته شده و روش توابع تشخیص برای طبقه‌بندی توضیح داده خواهد شـد. بخـش سـوم بـه روش‌هـای بـرآورد کوواریانس با استفاده از نمونه آموزشی کوچک پرداخته است، چون همانطور کـه ذکـر شـد، در تـصاویر پرطیفـی بـا کمبود نمونه‌های آموزشی مواجه هستیم. بخش چهارم به بیان روش‌های مختلف استخراج ویژگـی پرداختـه و نهایتـاً بخش پنجم به بیان نتیجه‌گیری می‌پردازد.

Feature selection [1]
Feature extraction [2]
Regularized Discriminant Analysis [3]
Leave One Out Covarinace [4]

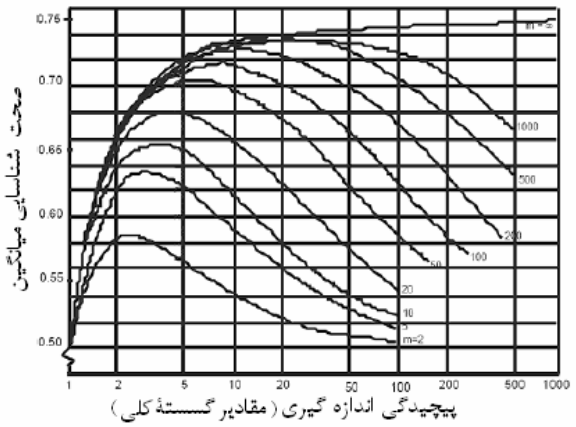
۲-۱- پدیده Hughes

پدیده Hughes بیان می کند که در تصاویر پرطیفی، با افزایش تعداد باندها، به علت افزایش جدایی بین باندها، تا حد مشخصی افزایش دقت طبقه‌بندی اتفاق می‌افتد، ولی بعد از این حد به دلیل کاهش دقت برآورد پارامترهای آماری، کاهش دقت خواهیم داشت. در واقع این پدیده بیان می‌کند که در تصاویر ماهواره‌ای پرطیفی، همیشه یک حد بهینه تعداد باندها برای انجام طبقه‌بندی وجود دارد و افزایش تعداد باندها لزوماً به دقت طبقه‌بندی بالاتر منجر نمی‌شود.

پیچیدگی اندازه گیری ها به این معنی است که یک اندازه گیری تا چه حد جزئی انجام شده است. اگر تعداد باندهای طیفی برابر n باشد و تعداد درجات روشنایی گسسته در هر باند طیفی k باشد، تعداد مکان های گسسته در فضای ویژگی n بعدی، k^n خواهد بود[۱].

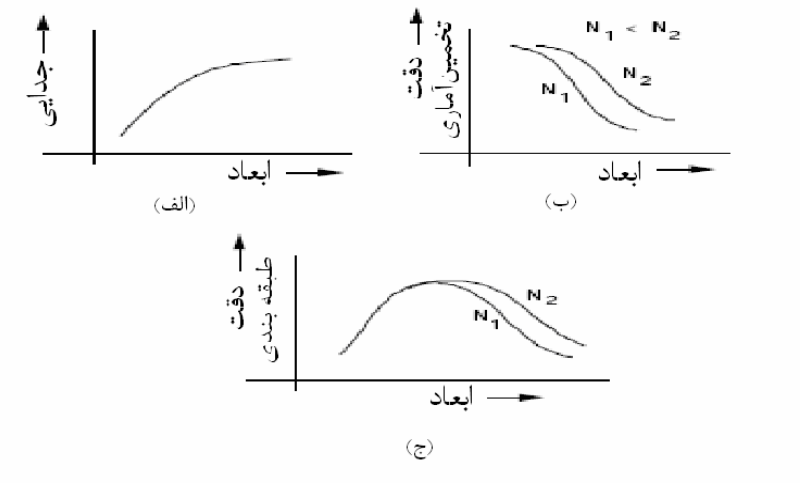
در تعداد نمونه های آموزشی محدود، هنگامی که پیچیدگی اندازه گیری ها افزایش یابد، می توان انتظار داشت دقت افزایش یابد. البته با افزایش پیوسته تعداد باندها، ممکن است یک اثر اشباع اتفاق بیفتد، یعنی با افزایش پیچیدگی اندازه گیری، دقت اندازه گیری به میزان بسیار کمی افزایش یابد.

موردی که ذکر شد، در حالتی است که تعداد نمونه های آموزشی نامحدود باشد. برای تعداد نمونه های آموزشی محدود یک حد پیچیدگی اندازه گیری بهینه وجود دارد. در واقع تعداد بسیار زیاد باندهای طیفی یا تعداد بسیار زیاد درجات روشنایی در هر باند طیفی، لزوماً به دقت بهتر منجر نمی شود. دکتر Hughes [۲]و [۱]، تحقیقاتی راجع به این موضوع انجام داده که نتایج آن در شکل ۱ نشان داده شده است.



شکل۱ نمایش اثر **Hughes** [۱]

در شکل ۱، افت دقت به وجود آمده به علت عدم وجود نمونه های آموزشی کافی می باشد، به طوریکه اگر نمونه‌های آموزشی به حد کافی موجود می بود، این مشکل به وجود نمی آمد.

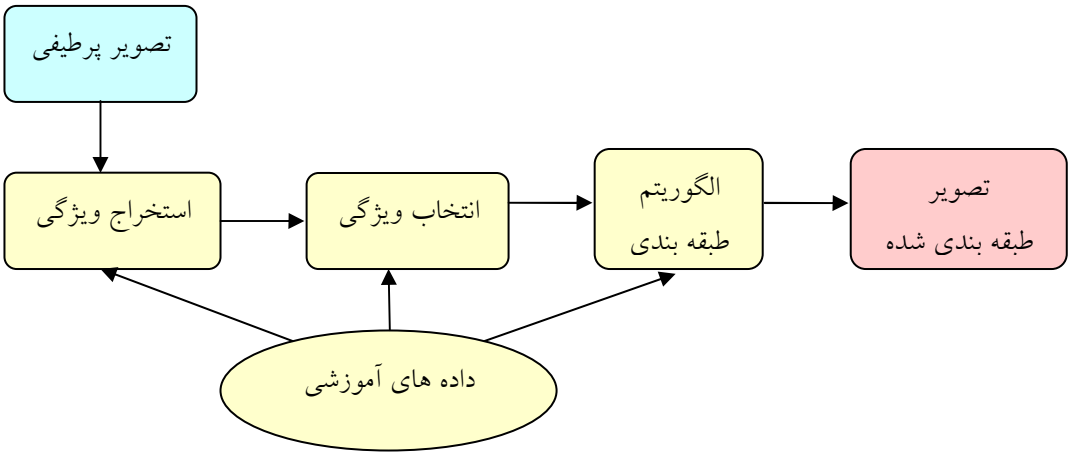


شکل۲ اثراتی که به علت پدیده **Hughes** اتفاق می افتد [۱]

همانطور که در شکل ۳ (الف) دیده می شود، با افزایش ابعاد، جدایی بین کلاس ها افزایش می یابد. یعنی در تعداد نمونه های آموزشی نامحدود، با افزایش میزان جزئیات، میزان جدایی بین کلاس ها کاهش مـی یابـد. شـکل ۳ (ب) نشان می دهد که با افزایش تعداد باندهای طیفی و افزایش پیچیدگی اندازه گیری، بـا یـک تعـداد محـدود داده آموزشی، بعد از یک تعداد ثابت باند ، دقت شروع به کاهش می کند، که این تعداد ثابت باند بـه تعـداد نمونـه هـای آموزشی بستگی دارد. در این شکل، Nها نشان دهنده تعداد نمونه های آموزشی بوده و نشان می دهد، هرچـه تعـداد نمونه های آموزشی بیشتر باشد، دقت در تعداد ابعاد بیشتری شروع به کاهش می‌کند . همانطور کـه در شـکل ۳ (ج) نشان داده شده، در هر محدوده کاربردی یک مقدار بهینه پیچیدگی داده وجود دارد و ویژگی هـای بیـشتر لزومـاً بـه نتایج بهتر منجر نمی شود. این موضوع، همان نتیجه ای است که دکتر Hughes بدان رسیده است.

۲- طبقه بندی تصاویر پرطیفی

طبقه بندی، تقسیم بندی فضای ویژگی به یک دسته از نواحی بدون پوشش می باشد، که مجموع آن‌ها فضای اولیه را تشکیل دهد. به نحوی که هر نقطه در فضا به صورت یک به یک با یکی از کـلاس هـای برچـسب گـذاری شده مرتبط باشد که در آن لازم است محدوده تصمیم گیری تشکیل شود. برای این منظور به یک دسته از نقاط که از قبل برچسب گذاری شده و کلاس هرکدام مشخص است، احتیاج است. این نمونه ها، نمونه های طراحی یا نمونـه های آموزشی نامیده می شود. مثلاً در مورد طبقه بندی کننـده حـداقل فاصـله، تعریـف محـدوده تـصمیم گیـری بـا مشخص کردن مکان هندسی نقاطی که از میانگین کلاس ها فاصله یکسانی دارند، انجام می شود، یـا در مـورد طبقـه بندی کننده نزدیکترین همسایگی، این کار با مشخص نمودن مکان هندسی نقاطی که از نزدیکترین اعضاء دو کـلاس هم فاصله هستند، انجام می شود. روند انجام طبقه بندی در تصاویر پرطیفی در شکل ۳ نمایش داده شده است.

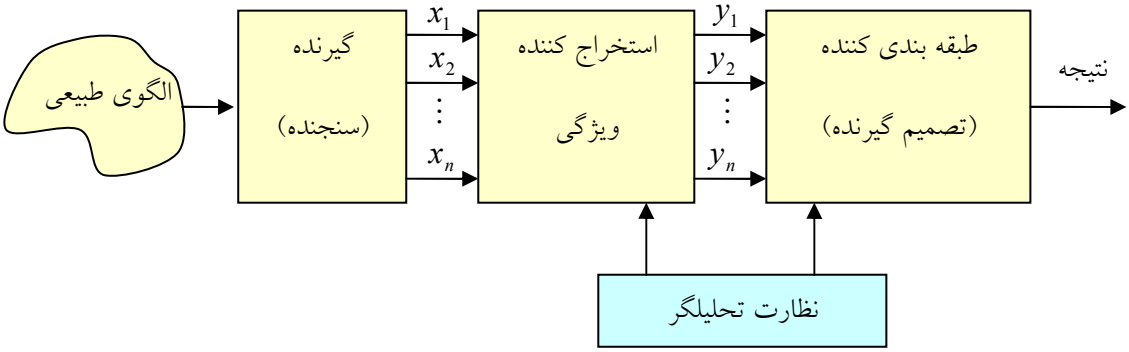


شکل۳ روند طبقه بندی تصاویر پر طیفی

همانطور که در شکل ۳ دیده می شود، برای طبقه بندی تصاویر پرطیفی، ابتدا بایستی روی آن ها یک مرحله استخراج ویژگی انجام داد، سپس بایستی از میان ویژگی های استخراج شده، یک سری ویژگی را برای انجام الگوریتم طبقه بندی انتخاب نمود و سپس الگوریتم طبقه بندی را پیاده سازی نمود. سه مرحله اخیر بایستی تحت نظارت داده های آموزشی انجام شوند و نهایتاً یک تصویر طبقه بندی شده نهایی خواهیم داشت.

۳- روش های استخراج ویژگی

سیستم های سنجنده اغلب برای جمع آوری داده در تعداد باندهای طیفی بیش از میزان موردنیاز، برای یک کاربرد خاص طراحی می شوند. یکی از علل این موضع آن است که در محدوده کاربردی، هر کاربرد ممکن است با یک سری ویژگی متفاوت، به بهترین نحو انجام شود. هرچند همان طور که قبلاً نشان داده شد استفاده از ویژگی های طیفی بسیار زیاد ، در عمل می تواند برای انجام یک تحلیل زیان آور باشد.



شکل۵ استخراج ویژگی به عنوان یک بخش از یک سیستم اطلاعات کلی

ممکن است در تعیین سری ویژگی برای استفاده در یک مورد خاص، اهداف متفاوتی وجود داشته باشد. بعضی از آن ها برای کارهای بارزسازی تصویر، به منظور واضح تر ساختن یک ویژگی خاص برای مشاهده کننده مفید هستند. بقیه به خاطر استفاده از محاسبه ویژگی های بهینه در فضای ویژگی برای استفاده در یک الگوریتم تشخیص الگو مدنظر می‌باشند.

همانطور که قبلا ذکر شد، کلاً برای کاستن ابعاد تصاویر پرطیفی دو روش وجود دارد(شکل ۵) کـه عبارتنـداز: انتخاب ویژگی و استخراج ویژگی. روش اول از میان تمام ویژگی ها، ویژگی هایی را انتخاب می کنـد کـه بیـشترین اطلاعات و کمترین نویز را داشته باشند، ولی روش دوم از اطلاعات تمامی بانـدها اسـتفاده مـی کنـد. در ایـن مقالـه بیشتر روی روش های استخراج ویژگی بحث خواهد شد.

۲-۳- استخراج ویژگی آنالیز تشخیص Discriminant Analysis Feature Extraction (DAFE)

تبدیل مؤلفه های اصلی بر مبنای ماتریس کوواریانس کلی همه سری داده قرار دارد و بنابراین بـه صـراحت بـه ساختار داخل کلاسی حساس نیست. اغلب این روش به عنوان یک وسیله کاهش ویژگی عمـل مـی کنـد. چـون در داده های سنجش از دور، کلاس ها به کرات در جهت بیشترین پراکندگی داده توزیع شده اند. استخراج ویژگی آنالیز تشخیص یک روش برای بارزسازی جدایی است.

در حالت ماتریسی، ماتریس پراکندگی داخل کلاسی Σ_w و ماتریس پراکندگی بین کلاسی Σ_B می توانـد بـه صورت زیر تعریف شود[۳] ، [۹] و[۱۰]:

$$\Sigma_w = \sum p_i(w_i)\Sigma_i \qquad \text{(ماتریس پراکندگی داخل کلاسی)} \qquad \text{(رابطه ۱)}$$

$$\Sigma_B = \sum p(w_i)(\mu_i - \mu_0)(\mu_i - \mu_0)^T \qquad \text{(ماتریس پراکندگی بین کلاسی)} \qquad \text{(رابطه ۲)}$$

$$\mu_0 = \sum p(w_i)\mu_i \qquad \text{(رابطه ۳)}$$

در این جا μ_i، Σ_i و $p(w_i)$ به ترتیب بردار میانگین، ماتریس کوواریانس و احتمال کلاس w_i است. معیـار بهینه‌سازی می تواند با بهینه سازی معیار Fisher، به صورت زیر تعریف شود:

$$J_{DAFE} = tr(\Sigma_W^{-1}\Sigma_B) \qquad \text{(رابطه ۴)}$$

بردارهای ویژگی جدید برای ماکزیمم کردن معیار انتخـاب مـی شـوند. ایـن پایـه ای بـرای آنـالیز تـشخیص استخراج ویژگی (DAFE) است.

این روش همچنین نقایصی دارد، که عبارتند از[۱۰]:

- از آن جاییکه آنالیز تشخیص به طور اصلی اختلاف میانگین کـلاس را در نظـر مـی گیـرد، اگـر بردارهـای میانگین نزدیک به یکدیگر باشند، بردار ویژگی انتخاب شده با آنالیز تشخیص ، قابل اعتماد نخواهـد بـود و با استفاده از جایگذاری کوواریانس در معیار، آنالیز تشخیص ممکن است اطلاعات داخل کوواریانس کلاسی را از دست بدهد.
- یک مشکل دیگر با تابع معیاری که از ماتریس های پراکندگی استفاده می کند، آن است کـه معیـار عمومـا" یک رابطه مستقیم با احتمال خطا ندارد.
- شاید بزرگترین اشکال آن، این باشد که این روش فقط ویژگی های بهینه را تا حدی انتخـاب مـی کنـد کـه یکی کمتر از تعداد کلاس ها باشد.

۳-۳- aPAC Linear Dimension Reduction (aPAC-LDR)

معیار aPAC-LDR[1] به معنی معیار کاهش ابعاد خطی دقت دوتایی برآورد شده می باشد[۱۱]. این روش مانند روش DAFE عمل می کند، با این تفاوت که میزان شرکت وزن دار هر جفت کلاس در نظر گرفته شده در آن، بر اساس فاصله اقلیدسی کلاس های مورد نظر خواهد بود. این وزن به صورت زیر تعریف می شود[۱۲]:

$$w(MD_{ij}) = \frac{1}{2MD_{ij}^2} erf(\frac{MD_{ij}}{2\sqrt{2}}) \quad \text{(رابطه ۵)}$$

که در آن MD_{ij} فاصله Mahalanobis میانگین کلاس های i و j می باشد و erf تابع خطا می باشد، که به صورت زیر تعریف می شود:

$$erf\ (z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \quad \text{(رابطه ۶)}$$

در این معیار ماتریس پراکندگی داخل کلاسی مانند روش DAFE است، ولی ماتریس پراکندگی بین کلاسی مجدداً تعریف می شود.یعنی خواهیم داشت[۱۲] و [۳] و [۱۰]:

$$S_w^{LDR} = \sum_{i=1}^{L} P_i \Sigma_i = S_w^{DA} \quad \text{(رابطه ۷)}$$

$$S_b^{LDR} = \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} P_i P_j w(MD_{ij})(m_i - m_j)(m_i - m_j)^T \quad \text{(رابطه ۸)}$$

یک آزمایش با داده واقعی و نمونه شبیه سازی شده در [۱۱] انجام شده است. این آزمایشات نشان می دهد که مزایای این روش عبارتند از:

- این روش می تواند تأثیر پرتی‌های[2] کلاس ها را در تبدیل LDR نهایی کاهش دهد.
- روش aPAC-LDR برای رسیدن به دقت بهینه ای که روش DAFE بدان می رسد، نیاز به ویژگی های کمتری دارد، ولی بهترین دقت aPAC-LDR تقریباً با DAFE یکسان است.

همچنین به علت این که این روش همچنان به صورت پارامتریک عمل می کند، مشکل بهینه کردن مشابه با DAFE خواهد بود.

۳-۴- استخراج ویژگی محدوده تصمیم گیری Decision Boundary Feature Extraction (DBFE)

در سنجش از دور، به ویژه برای حداقل کردن اثر متغیرهای مشاهداتی غیر قابل کنترل، باید طبقه بندی کننده برای هر سری داده که قرار است تحلیل شود، آموزش داده شود. این موضوع اهمیت استفاده از یک پردازش آموزش

---

[1] The approximated pairwise Accuracy Criterion Linear Dimension Reduction

[2] outlier

صحیح و انتخاب ویژگی را بارزتر می سازد. این مشخصه ما را ملزم می کند که بهترین ویژگی های ابعـاد داده شـده را انتخاب نموده و بهترین ابعاد مورد استفاده قرار گیرد.

این روش به طور مستقیم از نمونه های آموزشی به جای استفاده از آماره هایی که از آن ها استخراج می شـود، استفاده می کند. نمونه های آموزشی برای مشخص کردن موقعیت محدوده تصمیم گیری که با آن ها بـه وجـود مـی آید، استفاده می شود. سپس با کمک این محدوده تصمیم گیری و یافتن بردارهای نرمال بر آن ها می تـوان مـاتریس ویژگی مرز تصمیم گیری را ایجاد نمود و با کمک این ماتریس ویژگی مرز تصمیم گیری و یافتن بردارهای ویژه این مرز تصمیم گیری به ویژگی های استخراجی رسید. در واقع ویژگی های استخراج شده در امتداد این بردارهـای ویـژ می باشند.

قانون Bayes می تواند به صورت زیر بیان شود[۱]:

$$p(w_1)p(x \mid w_1) > p(w_2)p(x \mid w_2) \sim x \in w_1 \qquad \text{(رابطه ۹)}$$

$$p(w_1)p(x \mid w_1) < p(w_2)p(x \mid w_2) \sim x \in w_2 \qquad \text{(رابطه ۱۰)}$$

$$h(x) = -\ln\{p(x \mid w_1)\} + \ln\{p(x \mid w_2)\} \qquad \text{(رابطه ۱۱)}$$

$$t = \ln\left\{\frac{p(w_1)}{p(w_2)}\right\} \qquad \text{(رابطه ۱۲)}$$

سپس قانون تصمیم گیری Bayes به شکل زیر بیان می شود:

$$h(x) < t \sim x \in w_1 \qquad \text{(رابطه ۱۳)}$$

$$h(x) > t \sim x \in w_2 \qquad \text{(رابطه ۱۴)}$$

در این جا بیان بعضی تعاریف لازم به نظر می رسد، که این تعاریف به شرح زیر می باشند:

- ویژگی هایی که از لحاظ تشخیص افزونگی هستند:

اگر $\{\beta_1, \beta_2, \ldots, \beta_n\}$ پایه های یک نمایش فضای ویژگی جدید ممکن باشد، برای انـدازه گیـری N بانـد پاسخ پیکسل ها داریم:

$$x = \sum_{i=1}^{N} a_i \beta_i \qquad \hat{x} = \sum_{i=1}^{N} a_i \beta_i \qquad \text{(رابطه ۱۵ و ۱۶)}$$

که در این جا $\hat{x}$ ، x را در فضای اصلی بدون β_k تخمین می زند.

تعریف: β_k را از لحاظ تشخیص افزونگی می نامیم، اگر برای تمام مشاهدات x داشته باشیم:

$$\{h(x) - t\}\{h(\hat{x}) - t\} > 0 \qquad \text{(رابطه ۱۷)}$$

یعنی بدون حضور و با حضور β_k، کلاس تغییری نکند.

- ویژگی هایی که از لحاظ تشخیص دارای اطلاعات هستند: β_k را از لحاظ تـشخیص دارای اطلاعـات مـی-گوییم، اگر حداقل یک مشاهده x وجود داشته باشد، به نحوی که:

$$\{h(x) - t\}\{h(\hat{x}) - t\} < 0 \qquad \text{(رابطه ۱۸)}$$

یعنی بدون β_k، طبقه بندی تغییر می کند.

- محدوده تصمیم گیری مؤثر: محدوده تصمیم گیری به صورت $\{x \mid h(x) = t\}$ است. این محدوده می تواند یک نقطه، خط، سطح، یا یک رویه فضایی باشد. این محدوده در بسیاری از موارد که در بخشی از آن جدایی بین کلاس ها زیاد نیست، می تواند به سمت بی نهایت برود. ولی برای اهداف عملی ما محدوده تصمیم گیری مؤثر را به صورت زیر تعریف می کنیم:

$$\{x \mid h(x) = t, x \in R_1 \vee x \in R_2\}$$ (رابطه ۲۳)

در حالی که R_1 کوچکترین محدوده ای است که یک سهم خاص $P_{thershold}$ ، از کلاس w_1 را شامل می شود و R_2 کوچکترین محدوده ای است که یک سهم خاص $P_{thershold}$ ، از کلاس w_2 را شامل می شود. بنابراین محدوده تصمیم گیری مؤثر ممکن است یک قطعه از محدوده تصمیم گیری باشد.

- ابعاد تشخیص ذاتی: ابعاد تشخیص ذاتی کوچکترین ابعاد M از یک زیر فضای w است که برای هر مشاهده x داشته باشیم:

$$\{h(x) - t\}\{h(\hat{x}) - t\} > 0$$ (رابطه ۱۹)

در این جا $\hat{x} = \sum_{h=1}^{M} a_i \varphi_i$ یک تقریب از x در یک فضای تبدیل شده با پایه های زیر است:

$$\{\phi_1, \phi_2, \ldots, \phi_M\}, M \le N$$ (رابطه ۲۰)

پس ابعاد تشخیص ذاتی کوچکترین ابعاد زیرفضا است، به نحوی که یک دقت طبقه بندی یکسان با فضای اصلی را ایجاد کند.

- تست افزونگی: اگر بردار V عمود بر نرمال بر محدوده تصمیم گیری در هر نقطه روی محدوده تصمیم-گیری باشد، بردار V هیچ اطلاعاتی را که در تشخیص بین کلاس ها قابل استفاده باشد، ندارد. به این بردار V ، بردار افزونه می گویند.

بردارهای آموزنده: اگر بردار V نرمال بر محدوده تصمیم گیری در حداقل یک نقطه از محدوده تصمیم گیری باشد، بردار V حاوی اطلاعات مفیدی در تشخیص بین کلاس ها خواهد بود. به این بردار V ، بردار آموزنده می گویند.

- ماتریس ویژگی محدوده تصمیم گیری: یک بردار نرمال بر محدوده تصمیم گیری در یک نقطه، بردار آموزنده است و تأثیر بردار تقریباً متناسب با مساحتی از محدوده تصمیم گیری است که بردار نرمال به این شکل دارند. حال می توان ماتریس ویژگی محدوده تصمیم گیری را تعریف کرد، که می توان از آن برای پیش بینی ابعاد محدوده تصمیم گیری و پیدا کردن بردارهای ویژگی مؤثر استفاده کرد.

اگر N(x) بردار نرمال واحد بر محدوده تصمیم گیری در یک نقطه x در محدوده تصمیم گیری باشد، ماتریس ویژگی محدوده تصمیم گیری به صورت زیر تعریف می شود:

$$\Sigma_{DBFM} = \frac{1}{k} \int_s N(x) N^t(x) p(x) dx$$ (رابطه ۲۱)

در حالی که p(x) یک تابع چگالی احتمال است، $k = \int_s p(x)dx$ ، s محدوده تصمیم گیری است و انتگرال در محدوده تصمیم گیری ایجاد شده است. ماتریس ویژگی محدوده تصمیم گیری مؤثر (EDBFM) به عنوان یک حالت خاص که ماتریس ویژگی محدوده تصمیم گیری را از محدوده تصمیم گیری مؤثر ایجاد می کند، در نظر گرفته می شود. قابل توجه است که اگر قطعات متقابلاً جدا و جامع باشند، DBFM همه محدوده می تواند به صورت مجموع DBFM های محاسبه شده از قطعات محدوده تصمیم گیری در نظر گرفته شود.

- تئوری ابعاد تشخیص ذاتی: مرتبه ماتریس ویژگی محدوده تصمیم گیری (DBFM) از یک مسأله تشخیص الگو مساوی ابعاد تشخیص ذاتی است. بردارهای ویژه DBFM منطبق بر مقادیر ویژه غیر صفری است که بردارهای ویژگی لازم را برای به دست آوردن دقت یکسان با فضای اصلی ایجاد می کند.

روش DBFE[۱۳] نه تنها برای پیدا کردن بهترین ویژگی ها خوب عمل می کند، بلکه می تواند به خوبی میزان دقتی را که برای ابعاد مورد نظر به دست می آید، پیش بینی کند

آنالیز استخراج ویژگی محدوده تصمیم گیری محاسبات طولانی دارد، ولی محدودیت اصلی آن، این است که چون تکیه اصلی آن روی نمونه های آموزشی است، وقتی که سری آموزشی کوچک باشد، خوب عمل نمی کند.

۵-۳- استخراج ویژگی غیرپارامتریک وزن دار Nonparametric Weighted Feature Extraction (NWFE)

آزمایش DAFE و DBFE ما را به تعریف یک روش استخراج ویژگی جدید به نام استخراج ویژگی غیر پارامتریک وزن دار (NWFE) راهنمایی کرد. این روش مزایای هر دو DAFE و BDFE را درنظر گرفته و محدودیت های آن ها را کاهش می دهد. از DBFE داریم که تأکید روی نمونه های نزدیک موقعیت محدوده تصمیم گیری نهایی، برای بهبود DAFE می تواند مؤثر باشد. ایده اصلی NWFE قرار دادن وزن های متفاوت برای هر نمونه جهت محاسبه "میانگین محلی" وتعریف ماتریس پراکندگی بین کلاسی و داخل کلاسی غیرپارامتریک جدید برای به دست آوردن ویژگی های بیشتر می باشد. در NWFE ، ماتریس پراکندگی بین کلاسی به صورت زیر تعریف می شود:

$$\lambda_k^{(i,j)} = \frac{dist(x_k^{(i)}, M_j(x_k^{(i)}))^{-1}}{\sum_{l=1}^{n_i} dist(x_l^{(i)}, M_j(x_l^{(i)}))^{-1}}$$

(رابطه ۲۲)

در حالی که $x_k^{(i)}$ نمونه k ام کلاس i می باشد. $\lambda_k^{(i,j)}$ وزن ماتریس پراکندگی است و به صورت زیر می‌باشد:

$$S_b = \sum_{i=1}^{nc} \frac{p_i}{nc-1} \sum_{\substack{j=1 \\ j\neq i}}^{nc} \sum_{k=1}^{n_i} \lambda_k^{(i,j)} (x_k^{(i)} - M_j(x_k^{(i)}))(x_k^{(i)} - M_j(x_k^{(i)}))^T$$

(رابطه ۲۳)

$$M_j(x_k^{(i)}) = \sum_{l=1}^{n_i} W_l^{(i,j)} x_l^{(i)} x_l^{(j)}$$

(رابطه ۲۴)

در حالی که dist(a,b) به معنی فاصله از a تا b است و $M_j(x_k^{(i)})$ میانگین محلی $x_k^{(i)}$ درکلاس j است که به صورت زیر تعریف می شود:

$$W_l^{(i,j)} = \frac{dist(x_k^{(i)}, x_l^{(j)})^{-1}}{\sum_{l=1}^{n_i} dist(x_k^{(i)}, x_l^{(j)})^{-1}}$$

(رابطه ۲۵)

ماتریس پراکندگی داخل کلاسی غیر پارامتریک وزن دار به صورت زیر تعریف می شود:

$$S_w = \sum_{i=1}^{nc} p_i \sum_{k=1}^{n_i} \lambda_k^{(i,j)} (x_k^{(i)} - M_i(x_k^{(i)}))(x_k^{(i)} - M_i(x_k^{(i)}))^T$$

(رابطه ۲۶)

ویژگی های بهینه با بهینه کردن معیار Fisher تعیین می شوند.

برای کاهش اثر ضرب خارجی فاصله های بین کلاسی و جلوگیری از یکتا شدن، می توان S_w را با مورد زیر جایگزین نمود:

$$S_w = 0.5S_w + 0.5diag(S_w)$$

(رابطه ۲۷)

۴- نتیجه گیری

با توجه به مطالب ذکر شده می توان نتیجه گرفت، که در کار با داده های با ابعاد بالا ناگزیر به استفاده از روش‌های کاهش ویژگی هستیم و روش‌های استخراج ویژگی چون اطلاعاتی در مورد جدایی کلاس‌ها از دست نمی دهد، مناسب ترین روش برای استخراج ویژگی است. همچنین می توان به این نتیجه رسید که مناسب‌ترین روش برای استخراج ویژگی جهت دستیابی به یک دقت مناسب، روش NWFE است. چون این روش مزایای روش‌های دیگر را دارد و همزمان معایب آن ها را می کاهد.

با توجه به مطالبی که ذکر شد، در ادامه کار قصد داریم با توجه به معیارهای مورد نظر، استخراج ویژگی، جهت ایجاد طبقه بندی های مؤثرتر انجام دهیم. در واقع معیارهای مورد نظر ما معیارهایی است که به جای استفاده از ماتریس‌های پراکندگی بین کلاسی و داخل کلاسی تمام کلاس‌ها، از معیارهای به دست آمده از تک کلاس استفاده کنیم و بدین ترتیب دقت را بالاتر ببریم. اما روش‌هایی که برای طبقه‌بندی مدنظر ما می‌باشند، به صورت زیر عمل می‌کنند:

منابع و مأخذ

[1] David Landgrebe, *Signal Theory Methods In Multispectral Remote Sensing*, 508 pages plus a CD containing exercises and data. John Wiley & Sons, January 2003, ISBN 0-471-42028-X.

[2] G. F. Hughes, " On the mean accuracy of statistical pattern recognition", *IEEE Trans. Information Theory*, 1968, vol. IT-14, no. 1, pp. 55-63.

[3] B-C. Kuo and D. A. Landgrebe, "Improved Statistics Estimation And Feature

Extraction For Hyperspectral Data Classification", *PhD Thesis and School of Electrical & Computer Engineering Technical Report.* TR-ECE 01-6, December 2001.

[4] J. P. Hoffbeck and D.A. Landgrebe, "Classification of High Dimensional Multispectral Data", *Purdue University, West Lafayette, IN.,* TR-EE 95-14, May, 1995, pp.43-71.

[5] S. Tadjudin and D. A. Landgrebe, "Covariance Estimation With Limited Training Samples" *IEEE Transactions on Geoscience and Remote Sensing* Vol. 37, No. 4, pp. 2113-2118, July 1999.

[6] Saldju Tadjudin and David Landgrebe, "Classification of High Dimensional Data with Limited Training Samples", *PhD Thesis and School of Electrical & Computer Engineering Technical Report* TR-ECE 98-8, May 1998.

[7] J. P. Hoffbeck and D.A. Landgrebe, "Covariance matrix estimation and classification with limited training data" *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol.18, No. 7, 1996, pp. 763-767.

[8] J.H. Friedman, "Regularized Discriminant Analysis", *Journal of the American Statistical Association*, vol. 84, pp. 165-175, March 1989.

[9] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Second Ed., Boston, Academic Press,1990, pp. 445-479.

[10] B-C. Kuo and D. A. Landgrebe, "Nonparametric Weighted Feature Extraction for Classification", *IEEE Transactions on Geoscience and Remote Sensing*, Volume 42 No. 5, pp 1096-1105, May, 2004.

[11] R, P. W. Duin and R. Haeb-Umbach, "Multiclass Linear Dimension Reduction by Weighted Pairwise Fisher Criteria," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, pp. 762-766, 2001.

[12] Marco Loog, R.P.W. Duin, "Multiclass Linear Dimension Reduction by Weighted Pairwise Fisher Criteria", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, NO. 7, July 2001.

[13] Chulhee Lee and David Landgrebe, "Feature Extraction And Classificatiom Algorithms For High Dimensional Data", *PhD Thesis and School of Electrical Engineering Purdue University West* Lafayette, Indiana 47907-1285. TR-EE 93-1, January, 1993.