



نخستین از دور & GIS ایران



سنجش از دور و GIS ایران سال پنجم، شماره یکم، بهار ۱۳۹۲
Iranian Remote Sensing & GIS Vol.5, No.1, Spring 2013

۱۲-۲۸

طبقه‌بندی داده‌های ابرطیفی با استفاده از مشتق‌گیری در فضای طیفی و استفاده از روش‌های کدگذاری

مصطفی اوجاقلو^۱، محمدرضا مباحثی^{۲*}، یوسف رضائی^۳

۱. دانشجوی کارشناسی ارشد، دانشکده مهندسی ژئودزی و ژئوماتیک، دانشگاه صنعتی خواجه نصیرالدین طوسی
۲. دانشیار گروه سنجش از دور، دانشکده مهندسی ژئودزی و ژئوماتیک، دانشگاه صنعتی خواجه نصیرالدین طوسی
۳. استادیار گروه عمران دانشکده مهندسی، دانشگاه بوعلی‌سینا

تاریخ پذیرش مقاله: ۱۳۹۱/۸/۳

تاریخ دریافت مقاله: ۱۳۹۰/۹/۶

چکیده

روش‌های متعددی برای استخراج اطلاعات از داده‌های ابرطیفی وجود دارند که از آن جمله می‌توان به روش مشتق‌گیری از طیف بازتابندگی و استفاده از روش‌های کدگذاری اشاره کرد. در این تحقیق با مشتق‌گیری از طیف بازتابندگی، فرکانس‌های پایین طیف - که عمدتاً دارای اطلاعات مفیدی نیستند - حذف گردیده و همپوشانی بین مناطق جذبی از بین رفته و تباین طیفی نیز بالاتر برده شده است. دیگر اینکه استفاده از کدگذاری طیفی باعث افزایش سرعت پردازش داده‌ها گردیده که برای پردازش داده‌های حجیم ابرطیفی بسیار مناسب است. در این تحقیق ابتدا با استفاده از تابع رگرسیون چندگانه نویز موجود در داده‌ها کاهش یافته، و سپس با مشتق‌گیری از طیف داده‌های اصلی و داده‌های کتابخانه طیفی یا داده‌های آموزشی، طیف مشتق به‌دست آمده است. پس از آن با دو روش کدگذاری، طیف مشتق کدگذاری شد، سپس با استفاده از معیار شباهت AVD کدهای داده اصلی و داده‌های آموزشی با همدیگر مقایسه شدند و یک روش برای طبقه‌بندی داده‌های ابرطیفی به نام SDCM معرفی گردید. برای ارزیابی نتایج، دو سری داده شبیه‌سازی شده و یک سری داده واقعی به کار رفت و روش ارائه‌شده با سه روش کدگذاری SPAM و SDFC و SFBC مقایسه شد. نتایج دقت کلی برای داده‌های واقعی روش‌های کدگذاری SPAM، SDFC و SFBC به ترتیب ۷۱/۳۸ و ۷۴/۱۵ و ۶۸/۵۸ درصد و دقت کلی روش SDCM، ۸۶/۳۲ درصد به‌دست آمد، که پس از استفاده از فیلتر مکانی دقت کلی به ۹۰/۶۴ درصد افزایش یافت. این خود حاکی از مؤثر بودن روش ارائه‌شده برای طبقه‌بندی داده‌های ابرطیفی در قیاس با روش‌های کدگذاری ذکر شده است.

کلیدواژه‌ها: داده‌های ابرطیفی، تحلیل مشتق، کدگذاری، مناطق جذبی، طبقه‌بندی.

* نویسنده مکاتبه‌کننده: تهران، خیابان ولیعصر، تقاطع میرداماد، دانشگاه صنعتی خواجه نصیرالدین طوسی، دانشکده مهندسی ژئودزی و ژئوماتیک.

تلفن: ۰۹۱۲۱۲۲۶۶۳۰

Email: mobasheri@kntu.ac.ir

۱- مقدمه

به دست می‌آید. فضای مذکور فضای طیفی نام دارد و به این نمودار برای هر پیکسل «اثر طیفی^۲» آن پیکسل گفته می‌شود. هر عنصر خالص در روی زمین دارای اثر طیفی (بازتابندگی) خاصی است که برای آن عنصر منحصر به فرد است. با استفاده از همین ویژگی می‌توان عناصر روی سطح را در تصویر شناسایی و از هم تفکیک کرد. هر چه تعداد این تأثیرات طیفی (باندهای طیفی) ثبت شده در باندها و طول موج‌های مختلف برای پیکسل بیشتر باشد، نمودار اثر طیفی آن ماده دقیق‌تر خواهد بود. اگر هر پیکسل به صورت نقطه در فضای دکارتی چندبعدی به ابعاد باندها نمایش داده شود، فضای به وجود آمده فضای ویژگی خواهد بود. الگوریتم مورد استفاده برای استخراج اطلاعات طیفی از داده‌های چندطیفی بیشتر در این فضا ارائه می‌شوند ولی فضای مذکور به دلیل اینکه ابعاد فضا برای داده‌های ابرطیفی بسیار زیاد می‌شود و الگوریتم‌ها نمی‌توانند به خوبی در آن کار کنند، برای داده‌های ابرطیفی مطلوب نیست (Landgrebe, 2003).

در میان روش‌هایی که برای استخراج اطلاعات طیفی ارائه شده‌اند، تحلیل مشتق روش مناسبی برای استخراج اطلاعات و طبقه‌بندی داده‌های ابرطیفی است. مشتق‌گیری از طیف بازتابندگی باعث می‌شود که فرکانس‌های پایین طیف حذف شوند، همپوشانی بین مناطق جذبی از بین برود و تباین^۳ طیفی نیز افزایش یابد، و در نتیجه از این راه برآورد اطلاعات از اهداف زمینی بهتر صورت پذیرد (Zhang et al., 2004).

مشتق مرتبه اول و بالاتر نسبت به تغییرات در شدت نور، به خاطر تغییرات توپوگرافی و زاویه تابش خورشید و پوشش ابر، حساسیت کمتری دارد. همچنین در فاصله نمونه‌برداری یا فاصله طیفی بین باندهای سیستم‌های ابرطیفی، مشتق‌گیری حساسیت کمتری

در ابتدا سنجش از دور محدود به سنجنده‌های چندطیفی بود. این سنجنده‌ها اطلاعات نسبتاً مناسبی را از مواد موجود در صحنه تصویربرداری به دست می‌دادند. سنجنده‌های چندطیفی آنهایی هستند که از ۲ تا ۱۰ باند طیفی را هم‌زمان جمع‌آوری می‌کنند. الگوریتم‌های متفاوتی برای استخراج اطلاعات از داده‌های چندطیفی ارائه شده است، که هر یک منجر به دستیابی به نتایج خوبی در استخراج اطلاعات شده است (Duda and Hart, 1973).

در سال‌های اخیر با ورود سنجنده‌های ابرطیفی، همچون سنجنده ۲۲۴ باندهی AVIRIS اطلاعات بسیار دقیقی در حجم‌های زیاد از اهداف زمینی جمع‌آوری شده است. به خاطر ابعاد بسیار بالای داده‌های ابرطیفی، الگوریتم‌های مورد استفاده برای داده‌های چندطیفی نمی‌توانستند برای داده‌های ابرطیفی به کار روند. به همین دلیل برخی از پژوهشگران الگوریتم‌هایی را برای کاهش ابعاد داده‌های ابرطیفی و به کارگیری الگوریتم‌های استخراج اطلاعات از داده‌های چندطیفی ارائه کردند (Hoffbeck and Landgrebe, 1993). سنجنده‌های ابرطیفی، اطلاعات طیفی را هم‌زمان در صدها باند طیفی باریک جمع‌آوری می‌کنند، به گونه‌ای که طیف بازتابندگی^۱ حاصل را به صورتی تقریباً پیوسته به عنوان تابعی از طول موج نمایان می‌سازند (Landgrebe, 2003).

در الگوریتم‌های استخراج اطلاعات طیفی، فضاهای مختلفی برای پردازش داده‌های سنجش از دوری به کار می‌روند. فضاهای پردازش داده‌ها مشتمل اند بر فضای تصویر، فضای طیفی و فضای ویژگی. فضای تصویر آرایشی از پیکسل‌هاست که به صورت مکانی در کنار هم چیده می‌شوند و تصویری را تشکیل می‌دهند. اگر برای داده‌های چندطیفی یا ابرطیفی مقدار بازتابندگی هر پیکسل در هر باند در نمودار دوبعدی نمایش داده شود که محور افقی طول موج و محور عمودی آن بازتابندگی باشد، فضایی دوبعدی برای نمایش پیکسل

1. Spectral Reflectance
2. Spectral Signature
3. Contrast

Tsai and Philpot,) تا دقت طبقه‌بندی را بهبود بخشند (2002).

روش دیگری که در آن برای استخراج اطلاعات از طیف بازتابندگی استفاده شده، کدگذاری طیف‌ها است. در روش‌های کدگذاری هر طیف به صورت یک سری رشته اعداد کدگذاری می‌گردد و سپس این کدها با استفاده از روش فاصله همینگ^۴ و یا فاصله مقادیر مطلق^۵ (AVD)، به عنوان روش‌های اندازه‌گیری شباهت طیفی، با یکدیگر مقایسه می‌شوند. در سال‌های اخیر روش‌های مختلفی برای استخراج اطلاعات از داده‌های ابرطیفی با استفاده از کدگذاری ارائه شده است. از روش‌های معروف در این زمینه می‌توان به اینها اشاره کرد: SPAM^۶ (Mazer et al., 1988)، Qian et al. (1996) (SFBC^۷ al., 2009) SDFC^۸؛ که هر سه برای استخراج اطلاعات از داده‌های ابرطیفی به کار رفته و به نتایج خوبی دست یافته‌اند.

برخی از پژوهشگران نیز با استفاده از تلفیق روش‌های کدگذاری و روش‌های مشتق‌گیری الگوریتم‌هایی را برای طبقه‌بندی داده‌های ابرطیفی ارائه کرده‌اند. Sung در سال ۲۰۱۱ با استفاده از مشتق‌گیری طیف بازتابندگی، طیف مشتق را به‌دست آورده و آن را با روش میانگین‌گیری نرم‌سازی کرده و سپس به کدگذاری آن با روش مرسوم پرداخته است. سپس نمودار فراوانی^۹ کدها را ترسیم و آن را نیز کدگذاری کرده است. وی برای محاسبه شباهت طیفی طیف‌های بازتابندگی تصویر و مرجع، کدهای به‌دست‌آمده را با استفاده از فاصله همینگ مورد مقایسه قرار داده است (Sung, 2011).

نسبت به تغییرات تابش آفتاب و نور آسمانی دارد. (Bulter and Hopkins, 1970; Penuelas et al., 1994; Phil Pot, 1991).

در سنجش از دور چندطیفی و ابرطیفی مشتق اول و دوم و بالاتر برای آنالیز و شناسایی پوشش گیاهی، خاک، مواد معدنی و مانند اینها به کار رفته است. Esaias و Campbell در سال ۱۹۸۳ از مشتق مرتبه دوم منحنی طیف بازتابندگی برای تخمین غلظت کلروفیل اقیانوس استفاده کردند (Campbell and Esaias, 1983). Demetriades-Shah و همکاران در سال ۱۹۹۱ مشتق اول را برای شناسایی «لبه قرمز»^۱ در منحنی‌های طیف بازتابندگی به کار گرفتند (Demetriades-Shah et al., 1990). آنها از مشتق دوم برای تخمین میزان کلروفیل برگ گیاهان استفاده کردند. Blackburn در سال ۱۹۹۸ از مشتق برای کمی‌سازی کلروفیل گیاهان استفاده کرد (Blackburn, 1998). در این تحقیق نشان داده شد که مشتق اول و دوم از طیف بازتابندگی در مناطق جذبی همبستگی خوبی با مقدار کلروفیل دارند. Gong و همکاران در سال ۱۹۹۷ مشتق اول داده‌های ابرطیفی را به عنوان ورودی شبکه عصبی به کار بردند (Gong et al., 1997). از این روش برای طبقه‌بندی شش نوع درخت کاج استفاده شد. Philpot در سال ۱۹۹۱ الگوریتم نسبت مشتق (DRA^۲) را ایجاد کرد (Philpot, 1991). DRA روشی است که در آن از نسبت n امین مشتق در یک باند مشخص (λ_1) به n امین مشتق در یک باند مشخص دیگر (λ_2) در منحنی طیف بازتابندگی اندازه‌گیری شده استفاده می‌شود. او نشان داد که در شرایط معین این نسبت، معادل مشتقات n ام در طیف تابندگی^۳ واقعی ماده در همان طول موج‌هاست. پس تحت شرایط معین، DRA می‌تواند بدون نیاز به تصحیح جوی کامل، برای آنالیز منحنی‌های بازتابندگی طیفی به کار رود (Bruce et al., 2001). Tsai و Philpot در سال ۲۰۰۲ از طریق مشتق‌گیری چند باند جدید تولید کردند و آن را به فرایند طبقه‌بندی افزودند

1. Rea edge
2. Derivative ratio algorithm
3. Radiance
4. Hamming Distance
5. Absolute value distance
6. Spectral analysis manager
7. Spectral feature-based binary coding
8. Spectral derivative feature coding
9. Histogram

۱-۱- مشتق‌گیری متناهی

با توجه به طبیعت داده‌های ابرطیفی که ناپیوسته‌اند، مشتق‌گیری از طیف باید به صورت متناهی باشد. اگر فرض گردد که داده‌های ابرطیفی دارای n باند طیفی‌اند و مقادیر بازتابندگی در طیف i ام با طول موج λ_i با $s(\lambda_i)$ نشان داده می‌شود، آن‌گاه می‌توان پیکسل ابرطیفی را به صورت رابطه (۱) نمایش داد.

رابطه (۱)

$$S = [s(\lambda_1), s(\lambda_2), s(\lambda_3), \dots, s(\lambda_n)]^T$$

در این صورت مشتق اول آن به صورت رابطه (۲)

نوشته می‌شود.

رابطه (۲)

$$\left. \frac{dS}{d\lambda} \right|_i = \frac{s(\lambda_i) - s(\lambda_{i-1})}{\Delta\lambda}$$

که در آن $\left. \frac{dS}{d\lambda} \right|_i$ مشتق اول در طول موج λ_i است و

$\Delta\lambda$ فاصله بین باندهای مجاور، به گونه‌ای که $\Delta\lambda = \lambda_j - \lambda_i$ و $\lambda_j > \lambda_i$ است. با این مشتق‌گیری اکستریم‌های منحنی بازتابندگی طیف به دست می‌آید. با مشتق‌گیری بعدی (مشتق مرتبه دوم) جهت تقعرهای منحنی بازتابندگی طیفی به دست می‌آید. روابط مورد استفاده برای مشتق مرتبه دوم بدین صورت است:

رابطه (۳)

$$\left. \frac{d^2S}{d\lambda^2} \right|_i = \frac{s(\lambda_i) - 2s(\lambda_j) + s(\lambda_k)}{(\Delta\lambda)^2}$$

که در آن $\left. \frac{d^2S}{d\lambda^2} \right|_i$ مشتق دوم در طول موج λ_i بوده

و $\Delta\lambda$ فاصله بین باندهای مجاور i و j و k است، و $\lambda_k > \lambda_j > \lambda_i$ و $\lambda_k - \lambda_j = \lambda_j - \lambda_i = \Delta\lambda$ می‌باشد (Tsai and Philpot, 1998).

به دلیل اینکه مشتق طیفی به نویزهای با فرکانس بالا بسیار حساس است، برای استفاده از مشتق باید فرایند حذف فرکانس‌های بالا انجام گیرد؛ و معمولاً حذف فرکانس‌های بالا با فیلترهای نرم‌سازی صورت می‌پذیرد. به همین دلیل فرایند مذکور فرایند نرم‌سازی

نام دارد. روش‌های نرم‌سازی Savitzky-Golay (Savitzky and Golay, 1964) Minami-Kavata, (Kavata and Minami, 1984) Tsai (Tsai and Philpot, 1998)، و فیلتر میانگین (Tsai and Philpot, 1998)، از روش‌های معروف نرم‌سازی به شمار می‌آیند. در روش Savitzky مشخصات نویز در کل طیف ثابت و نامتغیر فرض می‌گردد. همچنین فرض می‌شود که می‌توان با یک فیلتر ثابت، نویز را حذف کرد. به همین خاطر اساس روش Savitzky بر برآزش چندجمله‌ای به چند باند همسایه به صورت قطعه‌ای^۱ و سپس بر برآورد مقدار هر طیف از معادله چندجمله‌ای استوار است. معادله عمومی برای محاسبه طیف نرم شده در این روش با رابطه ۴ بیان می‌شود (Savitzky and Golay, 1964).

رابطه (۴)

$$Y_j^* = \frac{\sum_{i=-m}^m C_i Y_{j+i}}{N}$$

در این رابطه Y طیف اصلی، و Y^* طیف نرم شده، C_i ضرایب ثابت است؛ و N تعداد طیف‌های مورد استفاده در یک پنجره همسایگی، m نصف عرض پنجره طیفی و مقدار N برابر $2m+1$ است. مقادیر C_i برای تعداد محدودی از چندجمله‌ای‌ها و برای پنجره همسایگی فرد ۱ تا ۲۵ ($m=1, 3, \dots, 12$) محاسبه شده و در جدول‌ها موجود است. در این روش رابطه (۵) برای مشتق‌گیری و نرم‌سازی همزمان ارائه شده است.

رابطه (۵)

$$\frac{d^q \bar{Y}_j}{dx^q} = \sum_{i=-m}^m P_i^{(q)} Y_{j+i}$$

در این رابطه P ضرایب معادله، و q مرتبه مشتق است. اگر q مساوی صفر باشد، آن‌گاه ضریب P مساوی $\frac{C_i}{N}$ خواهد بود که برابر رابطه (۴) است (Tsai and Philpot, 1998). فیلتر میانگین در واقع میانگین مقادیر موجود در یک پنجره همسایگی هر پیکسل را به

1. Piece wise

در رابطه (۶)، μ نشان‌دهنده میانگین طیف بازتابندگی است و x_i مقدار طیف بازتابندگی را در باند نام نشان می‌دهد. در روش کدگذاری چهارگانه از سه آستانه استفاده شده است. برای این کار طیف بازتابندگی به چهار ناحیه تقسیم می‌شود؛ به این صورت که پس از تقسیم طیف به دو ناحیه به‌وسیله میانگین هر قسمت دوباره با استفاده از میانگین مقادیر موجود در هر ناحیه به دو قسمت تقسیم می‌گردد و در کل چهار ناحیه به‌دست می‌آید. به هر کدام یک کد اختصاص می‌یابد که معمولاً به صورت رابطه (۷) است.

رابطه (۷)

$$\text{Quat} = \begin{cases} 3 & \text{if } \mu_{Up} \leq x_i \\ 2 & \text{if } \mu \leq x_i < \mu_{Up} \\ 1 & \text{if } \mu_{Low} \leq x_i < \mu \\ 0 & \text{if } x_i < \mu_{Low} \end{cases}$$

در رابطه (۷) μ_{Up} نشان‌دهنده میانگین مقادیر بالاتر از میانگین طیف بازتابندگی و μ_{Low} حاکی از میانگین مقادیر کمتر از میانگین طیف بازتابندگی است. در روش کدگذاری اختلاف باندهای مجاور بدون استفاده از مقادیر بازتابندگی طیف و فقط با استفاده از اختلاف بین باندهای مجاور کدگذاری می‌شود. از نمونه‌های این روش کدگذاری دودویی است که به صورت رابطه (۸) انجام می‌گیرد (Chakravarty, 2008).

رابطه (۸)

$$\text{DerivativeBinary} = \begin{cases} 1 & \text{if } x_i \geq x_{i-1} \\ 0 & \text{if } x_i < x_{i-1} \end{cases}, i = 2, 3, \dots, L$$

برمبنای تعریف مشتق متناهی می‌توان رابطه مذکور را برمبنای علامت مشتق تعریف کرد. در این روش اگر منحنی طیف بازتابندگی صعودی باشد که یک است و اگر نزولی باشد کد صفر (Chakravarty, 2008).

پیکسل مرکزی نسبت می‌دهد (Tsai and Philpot, 1998).

۲-۱- روش‌های کدگذاری

هنگام کار با حجم بسیار بزرگ داده‌های ابرطیفی سرعت پردازش بسیار اهمیت دارد. به عنوان مثال، اگر تصویر ۱۰۰ در ۱۰۰ پیکسل در ۲۲۴ باند باشد، مقدار ۲/۲۴ میلیون خواهد بود. این مقدار برای قطعه‌ای کوچک از تصویر است و اگر تصویر بزرگ‌تر شود بسته به الگوریتم مورد استفاده زمان پردازش داده‌ها نیز افزایش می‌یابد، و پردازش در زمان بیشتری انجام می‌گیرد. برای اینکه پردازش داده‌ها در زمان کمتری انجام شود، استفاده از روش سریع اجتناب‌ناپذیر است. روش‌های کدگذاری آنهایی هستند که به دلیل کاهش عملیات پردازشی، شدیداً بر سرعت پردازش داده‌ها می‌افزایند و برای پردازش داده‌های ابرطیفی مناسب‌اند. در کل دو روش برای کدگذاری وجود دارد: یکی برمبنای آستانه‌گذاری و دیگری برمبنای اختلاف بین باندهای مجاور است که می‌توان آن را به مشتق تعبیر کرد (Chakravarty, 2008). در روش‌های کدگذاری آستانه، یک یا چند حد آستانه در نظر گرفته و طیف به نواحی مختلفی تقسیم می‌شود و سپس به هر ناحیه کد ویژه‌ای اختصاص می‌یابد (Chakravarty, 2008). از روش‌های معروف در این زمینه می‌توان به روش کدگذاری دودویی^۱ (Du and Chen, 2004)، کدگذاری چهارگانه^۲ (Du and Chen, 2004)، و کدگذاری دهگانه^۳ (Du and Chen, 2004) اشاره کرد. در روش کدگذاری دودویی فقط از یک آستانه استفاده می‌شود که معمولاً میانگین طیف است. مقادیر بالاتر از میانگین کد ۱ می‌گیرد، و مقادیر پایین‌تر از میانگین کد صفر. رابطه (۶) برای کدگذاری دودویی آستانه ارائه شده است.

رابطه (۶)

$$\text{TreshBinary} = \begin{cases} 1 & \text{if } x_i \geq \mu \\ 0 & \text{if } x_i < \mu \end{cases}$$

1. Binary encoding
2. Quaternary encoding
3. Decimal encoding

داده شده‌اند (Chang et al., 2009).



شکل ۱. نمایش گرافیکی ۴ نوع متفاوت در روش SDFC

به دلیل اینکه مقدار صفر در مشتق به ندرت اتفاق می‌افتد، در روش SDFC یک تolerانس^۱ تعریف شده است تا مقادیر مشتق کوچک‌تر از آن صفر در نظر گرفته شود. این تolerانس در روش SDFC با رابطه (۱۵) بیان می‌گردد (Chang et al., 2009).
رابطه (۱۵)

$$\Delta = \frac{1}{L-1} \sum_{i=2}^L |r_i - r_{i-1}|$$

۲- مواد و روش‌ها

همان‌گونه که در بخش قبل ذکر شد، روش‌های کدگذاری به دو گونه تقسیم می‌شوند: کدگذاری آستانه و کدگذاری مشتق. در این تحقیق هر دوی این روش‌ها به کار رفته و تا حدی توسعه داده شده‌اند. در ادامه به شرح هر یک از این روش‌ها پرداخته خواهد شد.

۲-۱- کدگذاری آستانه

در این قسمت، کدگذاری با تعریف هشت ناحیه در طیف بازتابندگی انجام گرفته است. هر یک از این چهار ناحیه در روش کدگذاری چهارگانه با استفاده از میانگین همان ناحیه‌ها به دو قسمت بالاتر از میانگین و پایین‌تر از میانگین تقسیم شده است. بدین ترتیب هشت ناحیه شکل گرفته که به صورت رابطه (۱۶) کدهایی به طیف بازتابندگی اختصاص داده شده است:

1. Tolerance

روش‌هایی نیز از تلفیق روش کدگذاری آستانه دودویی و روش‌های کدگذاری مشتق ارائه شده‌اند که از آن جمله می‌توان به اینها اشاره کرد. در روش SPAM از تلفیق روش کدگذاری دودویی آستانه و روش مشتق استفاده می‌شود که روابط آن با روابط (۹) و (۱۰) بیان می‌گردد (Mazer et al., 1988).
رابطه (۹)

$$s_1^a = \begin{cases} 1 & \text{if } s_1 \geq \mu \\ 0 & \text{otherwise} \end{cases}, \text{ for } 1 \leq i \leq L$$

و

رابطه (۱۰)

$$s_1^{\text{SPAM}_b} = \begin{cases} 1 & \text{if } s_{i+1} \geq s_{i-1} \\ 0 & \text{otherwise} \end{cases}, \text{ for } 2 \leq i \leq L-1$$

و در صورتی که این دو کد با هم ترکیب شوند، رابطه (۱۱) به دست می‌آید (Mazer et al., 1988).
رابطه (۱۱)

$$S_1^{\text{SPAM}} = (s_1^a s_1^{\text{SPAM}_b})$$

در روش SFBC از کدگذاری به صورت رابطه (۱۲) استفاده شده است (Qian et al., 1996).

رابطه (۱۲)

$$S_1^{\text{SFBC}_b} = \begin{cases} 0 & \text{if } s_{i+1} < s_{i-1} \text{ and } |s_i - \mu| < MD \\ 1 & \text{if } s_{i+1} < s_{i-1} \text{ and } |s_i - \mu| \geq MD \\ 2 & \text{if } s_{i+1} \geq s_{i-1} \text{ and } |s_i - \mu| < MD \\ 3 & \text{if } s_{i+1} \geq s_{i-1} \text{ and } |s_i - \mu| \geq MD \end{cases}$$

که MD را می‌توان با رابطه ۱۳ محاسبه کرد (Qian et al., 1996).

رابطه (۱۳)

$$MD = \frac{1}{L} \sum_{i=1}^L |r_i - \mu|$$

کدهایی این روش در نهایت با ترکیب رابطه (۹) و (۱۲) به صورت رابطه (۱۴) محاسبه می‌شود (Qian et al., 1996).

رابطه (۱۴)

$$S_1^{\text{SFBC}} = (s_1^a s_1^{\text{SFBC}_b})$$

در روش SDFC علامت مشتق دو باند مجاور طیف بازتابندگی به کار می‌رود و ۹ رفتار کل بازتابندگی طیف به چهار گروه تقسیم می‌شود، در شکل ۱ نمایش

رابطه (۱۶)

$$T = \begin{cases} 8 & \text{if } \mu_{uu} \leq x_i \\ 7 & \text{if } \mu_u \leq x_i < \mu_{uu} \\ 6 & \text{if } \mu_{ul} \leq x_i < \mu_u \\ 5 & \text{if } \mu \leq x_i < \mu_{ul} \\ 4 & \text{if } \mu_{ll} \leq x_i < \mu \\ 3 & \text{if } \mu_l \leq x_i < \mu_{ll} \\ 2 & \text{if } \mu_{ll} \leq x_i < \mu_l \\ 1 & \text{if } x_i < \mu_{ll} \end{cases}$$

طیف دارای ۲۵۰ باند باشد، به ۱۰۰۰ کد برای کدگذاری این طیف نیاز است. برای رفع این ایرادها، پژوهشگران از روش AVD استفاده کردند، که در آن کدگذاری می‌تواند با هر عددی انجام شود. تعداد رشته کد به تعداد طیف‌هاست و اختلاف بین دو کد می‌تواند به اندازه یکی کمتر از طول رشته کد ضرب در اختلاف عددی بین کدها باشد. یعنی اگر اختلاف بین کدها یک واحد باشد و کدهای یک تا n به طیف واگذار شود بیشترین اختلاف n-1 خواهد بود.

در این رابطه $\mu_{ll}, \mu_l, \mu, \mu_u, \mu_{ul}, \mu_{uu}$ به ترتیب میانگین مقادیر ناحیه اول تا چهارم تشکیل شده در کدگذاری چهارگانه‌اند؛ و بقیه پارامترها مانند کدگذاری چهارگانه هستند. حال اگر از روش AVD برای مقایسه کدها استفاده شود، رابطه نهایی برای طبقه‌بندی داده‌ها به صورت رابطه (۱۷) خواهد بود.

رابطه (۱۷)

$$\text{Tresh} = \frac{1}{L-1} \sum_{i=1}^{L-1} |T_i^s - T_i^t|$$

۲-۲- کدگذاری مشتق

در قسمت کدگذاری مشتق طیفی، از علامت مشتق طیفی برای کدگذاری استفاده شده است. اگر از طیف مشتق گرفته شود و علامت طیف مشتق به دست آید، سه حالت مختلف به وجود خواهد آمد. علامت مشتق یا مثبت است، و یا منفی و یا بدون علامت (مقدار مشتق طیف بازتابندگی در آن نقطه صفر است). اگر هم‌زمان از دو باند مجاور برای کدگذاری استفاده شود، ۹ حالت متفاوت وجود خواهد داشت که در جدول ۱ نمایش داده شده‌اند.

به عبارت دیگر، چهار حالت موجود در روش SDFC به ۹ حالت گسترش داده شده است که با رابطه (۱۸) بیان می‌شود.

رابطه (۱۸)

$$D = \begin{cases} 9 & \text{if } s_1 > \Delta \ \& \ s_2 > \Delta \\ 8 & \text{if } s_1 > \Delta \ \& \ s_2 \leq \Delta \\ 7 & \text{if } s_1 > \Delta \ \& \ s_2 < -\Delta \\ 6 & \text{if } s_1 \leq \Delta \ \& \ s_2 > \Delta \\ 5 & \text{if } s_1 \leq \Delta \ \& \ s_2 \leq \Delta \\ 4 & \text{if } s_1 \leq \Delta \ \& \ s_2 < -\Delta \\ 3 & \text{if } s_1 < -\Delta \ \& \ s_2 > \Delta \\ 2 & \text{if } s_1 < -\Delta \ \& \ s_2 \leq \Delta \\ 1 & \text{if } s_1 < -\Delta \ \& \ s_2 < -\Delta \end{cases}$$

در این رابطه Δ تلرانس برای مقادیر نزدیک به صفر در طیف مشتق است که با رابطه (۱۹) بیان شده است.

رابطه (۱۹)

$$\Delta = \left| \frac{1}{L-1} \sum_{i=1}^{L-1} (x_i - x_{i+1}) \right|$$

در این رابطه T_i^s کد نام از رشته کد مربوط به طیف کتابخانه طیفی و T_i^t کد نام از رشته کد مربوط به طیف داده اصلی و L تعداد کدها است. در روش AVD پس از کدگذاری دو طیف بازتابندگی مرجع و پیکسل، میانگین قدر مطلق اختلافات دو رشته کد به عنوان فاصله مقادیر مطلق به دست می‌آید. روش فاصله همینگ نیز به همین صورت است، با این تفاوت که کدهای واگذار شده به دو طیف بازتابندگی به صورت دودویی‌اند و به اصطلاح فاصله همینگ بین دو رشته کد به دست می‌آید. روش مقایسه کد از طریق فاصله همینگ در بردارنده کاستی‌هایی است. نخست اینکه بیشترین فاصله بین این کدها به تعداد بیت‌های استفاده شده برای کدگذاری خواهد بود. به عنوان مثال برای کدگذاری با سه بیت ۸ کد متفاوت وجود دارد ولی بیشترین فاصله سه واحد است. در اکثر کدگذاری‌ها بهتر است بیشترین فاصله یکی کمتر از تعداد کدهای مورد استفاده باشد. دوم اینکه در این روش رشته‌های کد با طول بسیار بزرگ ایجاد می‌شود. به عنوان مثال اگر کدگذاری ۴ بیتی استفاده شود و

که اگر قرارگیری کدها مرتب شود، خواهیم داشت:

رابطه (۲۴)

$$TD = T_1 T_2 T_3 \dots T_{L-1} T_L D_2 D_3 D_4 \dots D_{L-2} D_{L-1}$$

پس می‌توان دو کد را به صورت جداگانه با هم مقایسه و نتایج را با هم جمع کرد. برای مقایسه کدها، دو روش Hamming و AVD وجود دارد، که در اینجا از روش AVD استفاده شده است.

رابطه (۲۵)

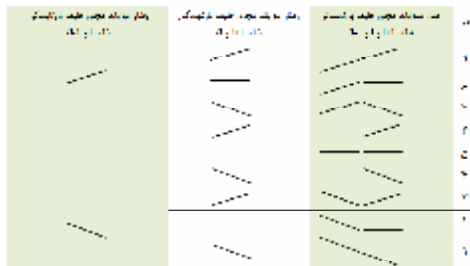
$$SDCM = \frac{1}{L-1} \sum_{i=1}^{L-1} |T_i^s - T_i^t| + \frac{1}{L-2} \sum_{i=1}^{L-2} |D_i^s - D_i^t|$$

در این رابطه D^s, D^t کدهای مربوط به طیف کتابخانه طیفی‌اند و D^t و T^t کدهای مربوط به پیکسل تصویر هستند. برای دستیابی به نتایج بهتر این کدگذاری بر طیف مشتق اعمال گردید و این روش $SDCM^1$ نام‌گذاری شد. اگر فاصله مقادیر مطلق (AVD) یک پیکسل نسبت به کتابخانه طیفی به دست آید، مقادیر متفاوتی بر مبنای شباهت هر کدام از طیف‌ها به طیف پیکسل حاصل می‌شود، که در آن‌ها طیف کتابخانه طیفی دارای کمترین مقدار SDCM به عنوان کلاس پیکسل در نظر گرفته خواهد شد.

۳- تجزیه و تحلیل نتایج

برای ارزیابی روش SDCM و مقایسه آن با روش‌های SPAM، SDFC و SFBC از دو داده شبیه‌سازی و یک داده واقعی استفاده شده است. در روش نخست با استفاده از کتابخانه طیفی USGS 1995 که دارای ۵۰۰ ماده معدنی است شبیه‌سازی انجام شد. در این روش طیف‌های لازم برای شبیه‌سازی، تصادفی انتخاب گردید. فراوانی هر ماده نیز در طیف بازتابندگی پیکسل به صورت تصادفی با توزیع درخنده^۲ تعیین شد. تابع توزیع درخنده با استفاده از رابطه (۲۶) به دست می‌آید. در این رابطه Γ تابع گاما، P تعداد کلاس‌ها، S_i فراوانی

جدول ۱. نه حالت متفاوت تغییر طیف بازتابندگی در سه باند مجاور



S_1 و S_2 اختلاف دو باند مجاور هستند که با رابطه (۲۰) به دست می‌آیند و x_i ها مقدار بازتابندگی در باند نام هستند.

$$S_1 = x_i - x_{i-1} \quad \text{رابطه (۲۰)}$$

$$S_2 = x_{i+1} - x_i$$

در این بخش نیز اگر از معیار شباهت AVD استفاده شود رابطه نهایی کدگذاری مشتق به صورت رابطه (۲۱) خواهد بود.

$$\text{Der} = \frac{1}{L-2} \sum_{i=1}^{L-2} |D_i^s - D_i^t| \quad \text{رابطه (۲۱)}$$

در این رابطه D_i^s کد نام از رشته کد مربوط به طیف کتابخانه طیفی و D_i^t کد نام از رشته کد مربوط به طیف داده اصلی و L تعداد کدها است.

برای اینکه بتوان از قابلیت‌های هر دو نوع کدگذاری استفاده کرد، کدگذاری‌های ذکر شده با روابط (۱۸) و (۱۹) با هم ترکیب شدند. از ترکیب آنها برای یک طیف بازتابندگی، خواهیم داشت.

$$T = T_1 T_2 T_3 \dots T_{L-1} T_L \quad \text{رابطه (۲۲)}$$

$$D = D_2 D_3 D_4 \dots D_{L-2} D_{L-1}$$

در این رابطه‌ها D کدگذاری مشتق و T کدگذاری آستانه است. همچنین D_i بیانگر کدهای واگذار شده در کدگذاری مشتق و T_i بیانگر کدهای واگذار شده در کدگذاری آستانه است. اگر به صورت ترکیبی از هر دو کدگذاری استفاده شود، به صورت رابطه (۲۳) درخواهد آمد.

رابطه (۲۳)

$$TD = T_1 D_2 T_2 D_3 T_3 D_4 \dots D_{L-1} T_{L-1} T_L$$

1. Spectral Deraivative Coding Method
2. Dirichlet

هر طیف و θ_i طیف i ام از کتابخانه طیفی است.

رابطه (۲۶)

$$D(s_1, s_2, \dots, s_p | \theta_1, \theta_2, \dots, \theta_p) = \frac{\Gamma(\sum_{j=1}^p \theta_j)}{\prod_{j=1}^p \Gamma(\theta_j)} \prod_{j=1}^p s_j^{\theta_j - 1}$$

فراوانی‌ها نیز با اعمال دو شرط منفی نبودن و برابر یک بودن مجموع فراوانی‌ها به دست آمده‌اند. برای اینکه داده‌های شبیه‌سازی شده به داده‌های واقعی شباهت بیشتری داشته باشند از مقداری نویز تصادفی جمع شونده با واریانس σ^2 که با استفاده از رابطه (۲۷) محاسبه شده، استفاده گردید. در این رابطه η نقش واریانس را در تابع گوسین نویز خواهد داشت. اگر η بی‌نهایت باشد، نویز فقط در یک باند است و اگر صفر باشد حاکی از نویز سفید است.

رابطه (۲۷)

$$\sigma_i^2 = \sigma^2 \frac{e^{-\frac{(i-L/2)^2}{2\eta^2}}}{\sum_{j=1}^L e^{-\frac{(j-L/2)^2}{2\eta^2}}}$$

پس از اینکه داده‌های شبیه‌سازی شده ساخته شدند، با استفاده از روش رگرسیون چندگانه نویز موجود در داده‌ها کاهش یافت (Bioucas-D and Nascimento, 2008). سپس روش پیشنهادی بر آن اعمال گردید و داده‌ها طبقه‌بندی شدند. در جدول ۲ چند تکرار شبیه‌سازی و نتایج طبقه‌بندی برای چند روش ارائه شده است. در این جدول ستون اول (P) نشان دهنده تعداد عناصر خالص شرکت داده شده در شبیه‌سازی، ستون دوم (N) تعداد پیکسل‌ها، ستون سوم (Noise) نوع نویز اضافه شده، ستون چهارم (SNR) نسبت سیگنال به نویز افزوده شده به تصویر شبیه‌سازی شده به دسی‌بل^۱، ستون پنجم، ششم و هفتم به ترتیب دقت کلی روش‌های SPAM، SFBC و SDFC است؛ و نیز ستون هشتم (Der) دقت کلی روش مشتق، ستون نهم (Tresh) دقت روش آستانه‌گذاری و ستون دهم (SDCM) دقت کلی تلفیق روش مشتق و آستانه‌گذاری است.

جدول ۲. نتایج طبقه‌بندی روش‌های SPAM، SFBC و SDFC و همچنین روش‌های پیشنهادی برای طبقه‌بندی داده‌های ابرطیفی شبیه‌سازی شده

p	N	Noise	SNR	SPAM	SFBC	SDFC	Der	Tresh	SDCM
۲۰	۱۰۰۰۰	Additive	۲۲/۴۹	۵۰/۲۱	۴۶/۰۰	۷۶/۷۲	۷۹/۳۵	۷۱/۲۳	۷۹/۳۸
۲۰	۱۰۰۰۰	Poisson	۲۰/۶۶	۵۵/۲۶	۵۰/۸۴	۵۴/۵۱	۷۲/۸۴	۶۹/۵۰	۷۴/۱۹
۱۵	۱۰۰۰۰	Additive	۲۲/۸۲	۴۹/۴۶	۴۸/۰۷	۵۳/۳۰	۷۷/۴۷	۶۹/۲۴	۷۶/۶۲
۱۵	۱۰۰۰۰	Additive	۲۲/۷۷	۵۲/۰۹	۵۰/۲۰	۵۷/۰۲	۷۷/۲۷	۷۰/۳۶	۷۵/۵۳
۱۵	۱۰۰۰۰	Poisson	۲۱/۶۷	۵۱/۷۶	۴۷/۸۹	۵۶/۳۷	۸۴/۱۲	۸۱/۳۰	۸۳/۹۳
۱۰	۱۰۰۰۰	Additive	۲۲/۵۸	۴۹/۲۸	۴۲/۳۱	۵۷/۵۹	۸۷/۳۷	۸۷/۹۵	۸۸/۳۲
۱۰	۱۰۰۰۰	Poisson	۲۱/۷۴	۶۱/۵۴	۵۵/۹۹	۶۵/۶۴	۷۸/۸۴	۷۶/۷۶	۷۸/۳۷
۷	۱۰۰۰۰	Additive	۲۲/۵۵	۶۴/۷۶	۶۲/۴۶	۶۷/۶۱	۸۵/۵۰	۸۳/۰۳	۸۵/۰۲
۷	۱۰۰۰۰	Poisson	۲۱/۸۵۲	۶۶/۰۸	۶۷/۹۶	۶۷/۴۶	۸۶/۲۲	۸۹/۶۴	۹۱/۱۶
۵	۱۰۰۰۰	Additive	۲۲/۴۶	۶۵/۳۶	۵۲/۸۷	۷۳/۳۵	۸۷/۳۳	۸۹/۵۰	۸۹/۴۰
۵	۱۰۰۰۰	Poisson	۲۱/۶۱	۶۷/۲۴	۶۰/۸۱	۵۹/۱۹	۷۸/۲۶	۷۹/۶۹	۸۱/۶۰
۳	۱۰۰۰۰	Additive	۲۲/۵۴	۵۴/۲۸	۵۳/۴۷	۶۸/۱۱	۹۰/۱۸	۹۰/۷۶	۹۱/۷۳
۳	۱۰۰۰۰	Poisson	۲۱/۷۹	۷۷/۴۵	۶۸/۰۱	۸۰/۳۳	۹۱/۵۶	۸۸/۰۲	۹۱/۰۶

1. Decibel

شماره آم با رابطه (۲۸) به دست می آید.

$$b_i = m_1(1-i/100) + m_2(i/100) \quad (28) \text{ رابطه}$$

در این رابطه b_i پیکسل آم تصویر خطی b است. قسمت c نیز مانند قسمت b است، با این تفاوت که در هر تکرار یک تصویر خطی اضافه می شود و در نهایت یک تصویر ۱۰۰ در ۱۰۰ پیکسل تولید می گردد.

برای شبیه سازی از سه سری مجموعه چهارتایی از طیف ها استفاده شد که در شکل ۳ نمایش داده شده است. هر یک از دوسری طیف مربوط به یک خانواده مواد است و اینها با هم متفاوت اند.

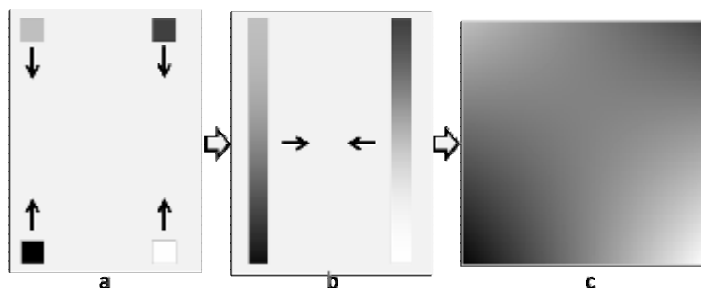
جدول ۳ نتایج (دقت کلی) طبقه بندی داده های شبیه سازی شده را با روش های متفاوت نشان می دهد. در این جدول رنگ سیاه نشانگر مناطقی است که اشتباه طبقه بندی شده اند. رنگ های دیگر طبق راهنمای نقشه کنار جدول است. در زیر هر تصویر دقت کلی و ضریب کاپای مربوط به آن درج شده است. ستون اول جدول نیز مشخص کننده سری مواد اختصاص داده شده در شکل ۳ و نوع نویز است. اسامی روش های ذکر شده مانند جدول ۲ هستند و منظور از روش Dersl در ستون پنجم ترکیب روش مشتق (Der) با رابطه ۹ بوده است. به خاطر اینکه در تمامی روش های کدگذاری معمول، رابطه (۹) در روش اصلی به کار رفته است، روش Dersl ارائه شده تا مقایسه ای را بین روش ها به دست دهد. رابطه (۹) روش کدگذاری آستانه است. روش SDCM نیز دارای کدگذاری آستانه است، که در آن کارایی کدگذاری آستانه پیشنهادی نیز ارزیابی می شود.

نتایج به دست آمده در جدول ۲ گویای کارایی بسیار بهتر روش های ارائه شده در قیاس با روش های معمول (SDFC, SPAM و SFBC) است. در قسمت A که طیف ها از چند خانواده مواد هستند، روش های کدگذاری معمول نتایج بس ضعیف تری در مقایسه با روش های ارائه شده دارند. ولی در قسمت B و C که از یک خانواده هستند روش های کدگذاری به خوبی توانسته اند طیف ماده اصلی را تشخیص دهند.

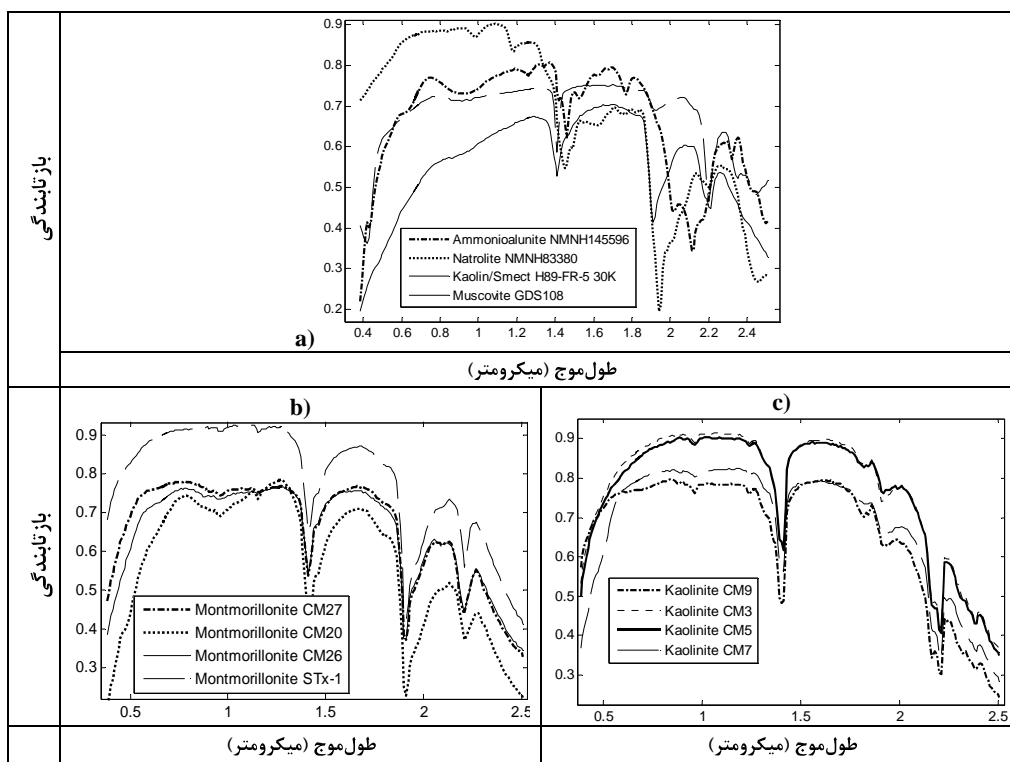
در جدول ۲ نتایج دیگر روش ها بر روی طیف بازتابندگی اصلی درج گردیده اند. همان گونه که مشاهده می شود، نتایج این روش ها در مقایسه با مشتق بسیار ضعیف اند. همچنین نتایج جدول ۲ بیانگر دقت بسیار بالای روش مشتق در قیاس با روش های کدگذاری معمول است. با افزایش تعداد اعضای خالص دقت تمامی روش ها کاهش می یابد، که این می تواند به خاطر تأثیر گذاشتن تعداد اعضای خالص بر روند کاهش نویز تصادفی با استفاده از تابع رگرسیون چندگانه باشد. نتایج جدول ۲ بدون استفاده از فرایند نرم سازی طیف های بازتابندگی است. در کل تأثیر فرایند نرم سازی بر داده های شبیه سازی شده منفی است و دقت طبقه بندی را کاهش می دهد و این خود می تواند به دلیل استفاده از طیف کتابخانه طیفی برای شبیه سازی و همچنین استفاده از همان طیف برای شناسایی ماده باشد. این بدان معنی است که فرکانس های بالای طیف های بازتابندگی تصویر و کتابخانه طیفی تقریباً یکسان اند ولی در تصاویر واقعی تقریباً طیف های بازتابندگی، هر پیکسل دارای فرکانس های بالای متفاوتی است؛ و اینها فرکانس های بالای طیف های بازتابندگی کتابخانه طیفی نیز متفاوت اند.

در روش شبیه سازی دوم چهار ماده معدنی کتابخانه طیفی به نسبت های متفاوت با هم ترکیب شدند و سپس طبقه بندی با روش پیشنهادی و روش های دیگر به عمل آمد. شکل ۲ فرایند شبیه سازی تصویر را نمایش می دهد. ابتدا ۴ طیف بازتابندگی خالص در نظر گرفته شد (قسمت a) و سپس دو طیف بازتابندگی با هم ترکیب شدند، به صورتی که در هر پیکسل یک درصد از یک طیف بازتابندگی کم شد و به طیف بازتابندگی دیگر افزوده گشت و تصویر خطی قسمت b ساخته شد و دوباره به همین صورت تصویرهای خطی با هم ترکیب شدند و تصویر نهایی شبیه سازی شده ۱۰۰ در ۱۰۰ پیکسل به دست آمد. در قسمت b اگر ماده خالص m_1 باشد و ماده خالص دوم m_2 باشد، طیف

طبقه‌بندی داده‌های ابرطیفی با استفاده از مشتق‌گیری در فضای طیفی و استفاده از روش‌های کدگذاری



شکل ۲. فرایند شبیه‌سازی تصویر



شکل ۳. چهار طیف بازتابندگی چهار ماده معدنی مورد استفاده برای شبیه‌سازی دوم

شده‌اند. این داده‌ها دارای ۱۶ کلاس هستند که قرارگیری داده‌های زمینی در شکل ۴ قسمت آخر نشان داده شده است. ۲ درصد از این داده‌ها به عنوان داده‌های آموزشی یا کتابخانه طیفی در نظر گرفته شده‌اند. نتایج طبقه‌بندی این داده‌ها در شکل ۴ نمایش داده شده‌اند.

با استفاده از داده‌های واقعی نیز نتیجه روش ارزیابی گردیده است. داده‌های واقعی مربوط به منطقه Salinas در ایالت کالیفرنیا، امریکا است. این داده‌ها به وسیله سنجنده هوآبرد AVIRIS گرفته شده است. داده‌های مذکور به صورت رادیانس^۱ هستند و به بازتابندگی تبدیل نشده‌اند. به خاطر جذب جو در برخی از طول موج‌ها، باندهای ۱۰۸ تا ۱۱۲، ۱۵۴ تا ۱۶۷ و باند ۲۲۴ دارای اطلاعات مطلوب نیست و اینها همه حذف

1. Radiance

مصطفی اوجاقلو و همکاران

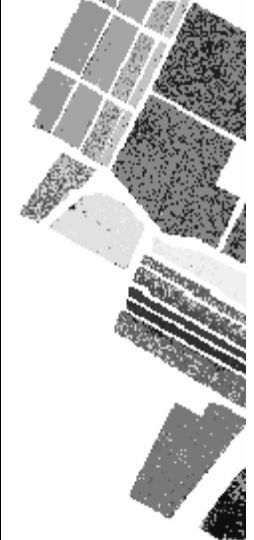
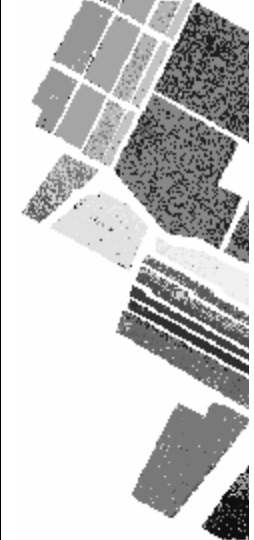
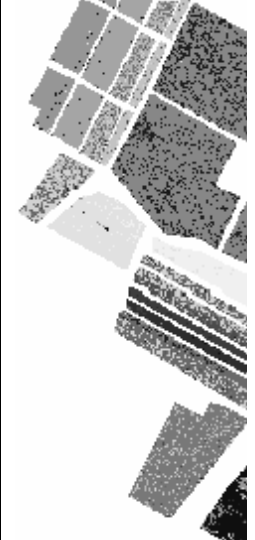
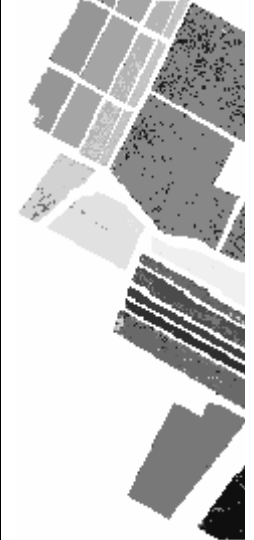
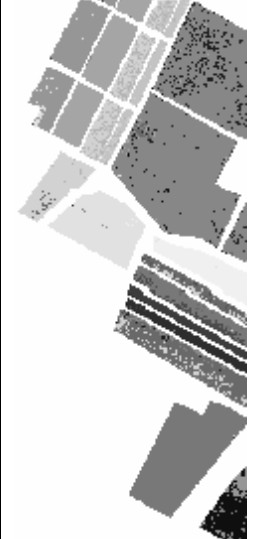
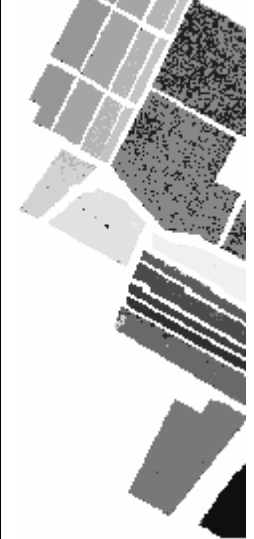
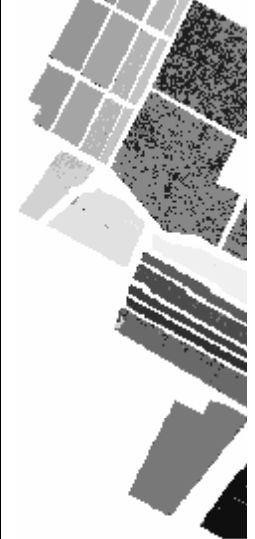

جدول ۳. نتایج طبقه‌بندی داده‌های شبیه‌سازی شده

	SPAM	SFBC	SDFC	Dersl	Der	Tresh	SDCM	راهنمای نقشه
C Poisson								
دقت کلی	۶۴/۷	۶۳/۴۹	۷۳/۵۹	۸۱/۲۷	۸۱/۰۴	۸۶/۱۴	۸۷/۱۱	
کاپا	-/۵۳۰۰	-/۵۴۹۰	-/۶۴۸۳	-/۷۵۰۳۷	-/۷۴۸۵	-/۸۱۵۲	-/۸۲۸۱	
C Additive								
دقت کلی	۶۴/۰۵	۶۴/۹۹	۷۴/۸۲	۸۴/۱۲	۸۳/۵۹	۸۵/۳۸	۸۸/۲۴	
کاپا	-/۵۲۰۸	-/۵۶۸۵	-/۶۶۴۴	-/۷۸۸۳	-/۷۸۱۷	-/۸۰۵۱	-/۸۴۳۲	
B Poisson								
دقت کلی	۶۵/۷۷	۶۸/۶	۶۰/۴۶	۷۳/۵۸	۶۶/۲۲	۷۸/۷۴	۷۸/۰۵	
کاپا	-/۵۴۳۶	-/۵۸۱۷	-/۴۷۲۸	-/۶۵۴۸	-/۵۸۷۳	-/۷۳۱۵	-/۷۱۵۷	
B Additive								
دقت کلی	۶۸/۵۸	۷۱/۵۷	۶۱/۸۸	۷۳/۵۶	۶۶/۲	۸۱/۷۷	۷۹/۶۵	
کاپا	-/۵۸۱۱	-/۶۲۱۱	-/۴۹۱۷	-/۶۵۴۶	-/۵۸۵۶	-/۷۵۹۱	-/۷۳۴۴	
A Poisson								
دقت کلی	۴۱/۱۰	۴۸/۵۸	۳۱/۳۱	۶۶/۵۱	۷۷/۸۱	۷۵/۰۸	۸۲/۰۵	
کاپا	-/۳۶۹۰	-/۴۳۸۷	-/۲۸۰۶	-/۵۷۹۸	-/۷۱۴۹	-/۶۹۸۱	-/۷۷۳۵	
A Additive								
دقت کلی	۴۲/۸۷	۴۹/۶۱	۳۱/۴۸	۶۷/۷۳	۷۹/۷۱	۷۴/۴۵	۸۳/۲۶	
کاپا	-/۳۸۴۶	-/۴۴۷۳	-/۲۸۲۴	-/۶۰۲۸	-/۷۳۵۴	-/۷۱۲۳	-/۷۸۷۴	

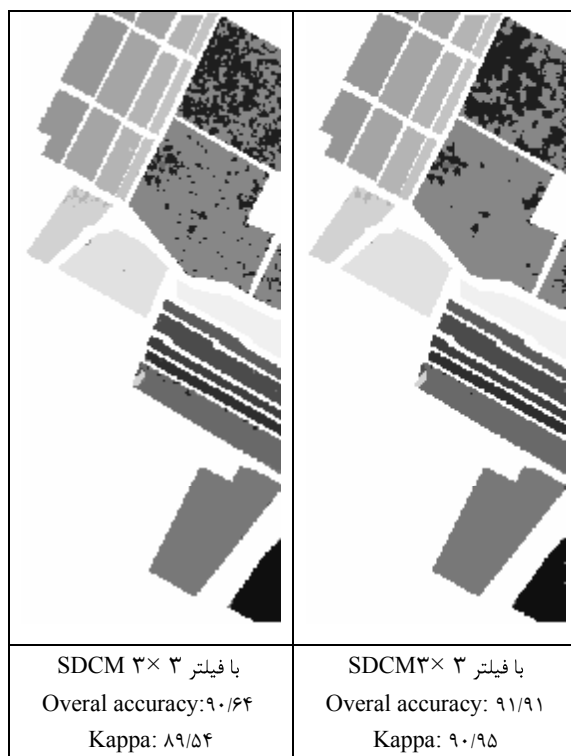
فیلتر مکانی می‌توان این مشکل را تا حدودی برطرف کرد و دقت طبقه‌بندی را تا حدودی افزایش داد. در این قسمت از تحقیق با به‌کارگیری فیلتر مکانی مدد دقت طبقه‌بندی افزایش می‌یابد. برای این فیلتر پنجره 3×3 و 5×5 روی تصویر حرکت داده می‌شود و شماره کلاس با بیشترین تکرار داخل پنجره به عنوان کلاس پیکسل مد نظر قرار می‌گیرد. در آخر نتایج روش به صورت شکل ۵ خواهد بود.

همان‌گونه که از نتایج پیداست، دقت روش SDCM بهتر از روش‌های دیگر است. این می‌تواند به دلیل استفاده از روش کدگذاری با تعداد کد مناسب‌تر، در مقایسه با روش‌های کدگذاری SPAM، SFBC و SDFC باشد. یکی از اشکالات روش‌های طیفی این است که روابط مکانی پیکسل‌های ابرطیفی در آنها در نظر گرفته نمی‌شود و این باعث می‌گردد نتایج طبقه‌بندی به صورت نقطه نقطه باشد. با استفاده از

طبقه‌بندی داده‌های ابرطیفی با استفاده از مشتق‌گیری در فضای طیفی و استفاده از روش‌های کدگذاری

			
SPAM Overall accuracy: ۷۱/۳۸ Kappa: ۰/۶۷۹۹	SFBC Overall accuracy: ۷۴/۱۵ Kappa: ۰/۷۱۱۲	SDFC Overall accuracy: ۶۸/۵۸ Kappa: ۰/۶۴۸۱	Dersl Overall accuracy: ۷۷/۹۴ Kappa: ۸۰/۳۸
			
Der Overall accuracy: ۷۴/۰۵ Kappa: ۷۱/۸۹	Tresh Overall accuracy: ۸۶/۳۵ Kappa: ۸۴/۷۷	SDCM Overall accuracy: ۸۶/۳۲ Kappa: ۸۴/۷۴	داده‌های صحیح زمینی

شکل ۴. نتایج اعمال روش‌های متفاوت طبقه‌بندی بر روی داده‌های زمینی Salinas، نقشه آخر مربوط به داده‌های صحیح زمینی است.



شکل ۵. نتایج اعمال فیلتر مکانی مد بر روی نتیجه طبقه‌بندی روش SDCM بر روی داده‌های زمینی Salinas، با پنجره فیلتر ۳×۳ و ۵×۵

کدگذاری نیز باعث افزایش شدید سرعت پردازش داده‌ها می‌شود و برای پردازش داده‌های حجیم ابرطیفی بسیار مناسب است.

در این تحقیق با مشتق‌گیری و سپس کدگذاری طیف‌های داده‌های ابرطیفی و طیف‌های کتابخانه طیفی، روشی برای طبقه‌بندی داده‌های ابرطیفی ارائه شده است. نتایج نشان می‌دهند که مشتق‌گیری از طیف باعث می‌شود تغییرات طیف نمایان‌تر گردند و کدگذاری بسیار دقیق‌تر انجام گیرد که حاصل آن طبقه‌بندی دقیق‌تر طیف است. نتایج شبیه‌سازی دوم نشان می‌دهد که روش‌های کدگذاری طیف با خلوص بیشتر را بهتر تشخیص می‌دهند. همچنین اگر طیف مختلط از یک خانواده مانند Kaolinite باشد، تشخیص طیف اصلی با روش‌های کدگذاری بهتر از تشخیص

۴- نتیجه‌گیری

داده‌های ابرطیفی اطلاعات بسیار زیادی را از اهداف زمینی جمع‌آوری می‌کنند که با آنها می‌توان اهداف مذکور را به‌خوبی شناسایی کرد. روش‌های متعددی برای استخراج اطلاعات از داده‌های ابرطیفی وجود دارند که از آن جمله می‌توان به تحلیل مشتق، استفاده از روش‌های کدگذاری و همچنین استفاده از مناطق جذبی اشاره کرد.

مشتق‌گیری از طیف بازتابندگی باعث می‌شود که فرکانس‌های پایین طیف بازتابندگی حذف شود، همپوشانی بین مناطق جذبی از بین برود و کنتراست طیفی بالاتر رود، و از این راه برآورد اطلاعات از اهداف زمینی بهتر صورت پذیرد. در نتیجه استفاده از تحلیل مشتق دقت استخراج اطلاعات را افزایش می‌دهد.

طیف را تخریب می‌کند، از کوچک‌ترین پنجره که پنجره 3×3 استفاده شد.

۵- منابع

- Bioucas-D. J. M., Nascimento J. M. P., 2008, **Hyperspectral Subspace Identification**, IEEE Trans. Geosci. Remote Sensing, vol. 46(8): 2435-2445.
- Blackburn, G. A., 1998, **Quantifying Chlorophylls and Carotenoids at Leaf and Canopy Scales: An Evaluation of Some Hyperspectral Approaches**, Remote Sens. Environ, 66: 273-285.
- Bruce, L.M., Li, J., 2001, **Wavelets for Computationally Efficient Hyperspectral Derivative Analysis**, IEEE Trans. Geosci. Remote Sensing, 39(7): 1540-1546.
- Butler, W.L., Hopkins, D.W., 1970, **Higher Derivative Analysis of Complex Absorption Spectra**, Photochem, Photobiol, 12: 439-450.
- Campbell, J. W., & Esaias, W. E., 1983, **Basis for Spectral Curvature Algorithms in Remote Sensing of Chlorophyll**, 22(7): 1084-1093.
- Chakravarty, S., 2008, **Analysis on Hyperspectral Signature Coding**, Ph.D. Thesis, Faculty of the Graduate School of the University of Maryland, Maryland.
- Chang C.I., Chakravarty S., Chen H.M., Ouyang Y.C., 2009, **Spectral Derivative Feature Coding for Hyperspectral Signature Analysis**, Pattern Recognition 42: 395 - 408.

طیف مختلط از خانواده‌های متفاوت است. در روش‌های کدگذاری SPAM، SFBC و SDFC از تعداد کد کمتر و کدهای باینری استفاده می‌شود. به دلیل اینکه هدف فشرده‌سازی تصویر نیست. افزایش تعداد کد عامل منفی به شمار نمی‌آید و باعث ارتقای دقت طبقه‌بندی هم می‌گردد. روش‌های کدگذاری معمول ذکر شده همگی دارای یک بخش کد هستند که از رابطه (۹) به دست می‌آید و از نوع کدگذاری آستانه است. روش ارائه شده Dersl نیز ترکیب کدگذاری مشتق با رابطه (۹) است. نتایج نشان می‌دهند که اگرچه ترکیب کدگذاری مشتق با رابطه (۹) دقت طبقه‌بندی را افزایش می‌دهد ولی می‌توان تعداد آستانه‌ها را افزایش داد و از کدگذاری آستانه بهتری استفاده کرد. در این تحقیق یک کدگذاری آستانه با ۷ آستانه ارائه گردید و سپس با کدگذاری مشتق ترکیب شد که حاکی از دقت بیشتر روش ترکیب شده SDCM در قیاس با روش Dersl است.

نتایج طبقه‌بندی داده‌های واقعی نیز نشانگر کارایی بسیار خوب روش‌های ارائه شده‌اند. در این داده‌ها دو کلاس دارای طیف بسیار شبیه هم هستند که این خود باعث کاهش دقت می‌شود. طبقه‌بندی این کلاس‌ها با روش‌های طیفی بسیار مشکل است. استفاده از روش‌های مکانی برای جداسازی این نوع کلاس‌ها می‌تواند نتایج بهتری داشته باشد.

به دلیل اینکه مشتق‌گیری به فرکانس‌های مکانی کوچک یا تغییرات ناگهانی طیف حساس است، استفاده از روش‌های کاهش نویز ضروری است. به همین دلیل در روش ارائه شده ابتدا با استفاده از تابع رگرسیون چندگانه نویز داده‌ها کاهش یافت، پس از مشتق‌گیری نیز با روش‌های نرم‌سازی طیف، طیف مواد نرم‌تر شدند. دو روش مشهور برای نرم‌سازی طیف روش‌های Savitzky و میانگین‌گیری هستند. به دلیل اینکه میانگین‌گیری ساده است و از هیچ فرض اولیه‌ای استفاده نمی‌کند برای داده‌های ابرطیفی مناسب است؛ ولی در اینجا به دلیل اینکه استفاده از پنجره با ابعاد بالا

- Demetriades-Shah, T.H., Steven, M.D., Clark, J.A., 1990, **High Resolution Derivatives Spectra in Remote Sensing**, Remote Sens. Environ, 33: 55–64.
- Du, P., Chen, Y., 2004, **Study on Spectral Similarity Measure in Hyperspectral Remote Sensing Data**, IEEE, 268-271.
- Duda, R.O., Hart, P.E., 1973, **Pattern Classification and Scene Analysis**, Wiley, New York.
- Gong, P., Pu, R., Yu, B., 1997, **Conifer Species Recognition: An Exploratory Analysis of in Situ Hyperspectral Data**, Remote Sens. Environ, 62: 189–200.
- Hoffbeck, J.P., & Landgrebe, D.A., 1993, **Classification of High Dimensional Multispectral Image Data**, Washington D. C.: In The Fourth Annual JPL Airborne Geoscience Workshop.
- Kawata, S., and Minami, S., 1984, **Adaptive Smoothing of Spectroscopic Data by a Linear Mean-square Estimation**. Appl. Spectrosc. 38(1): 49–58.
- Landgrebe, D.A., 2003, **Signal Theory Methods in Multispectral Remote Sensing**, John Wiley & Sons, Inc., Hoboken, New Jersey.
- Mazer, A.S., Martin, M., Lee, M., Solomon, J.E., 1988, **Image Processing Software for Imaging Spectrometry Data Analysis**, Remote Sensing Environ, 24: 201-210.
- Qian, S., Hollinger, A.B., Williams, D., Manak, D., 1996, **Fast Three-dimensional Data Compression of Hyperspectral Imagery Using Vector Quantization with Spectral-feature-Based Binary Coding**, Opt. Eng., 35(11): 3242-3249.
- Philpot, W.D., 1991, **The Derivative Ratio Algorithm: Avoiding Atmospheric Effects in Remote Sensing**, IEEE Trans. Geosci. Remote Sens. 29(3): 350–357.
- Peñuelas, J., Gamon, J.A., Fredeen, A. L., Merino, J., Field, C. B., 1994, **Reflectance Indices Associated with Physiological Changes in Nitrogen- and Water-limited Sunflower Leaves**, Remote Sens. Environ, 48: 135–146.
- Rhae Sung, K., 2011, **Spectral Matching using Bitmap Indices of Spectral Derivatives for the Analysis of Hyperspectral Imagery**, Graduate Program in Geodetic Science and Surveying, The Ohio State University.
- Tsai F., Philpot W., 1998, **Derivative Analysis of Hyperspectral Data**, REMOTE SENS. ENVIRON, 66: 41–51.
- Tsai, F., Philpot, W.D., 2002, **A Derivative-Aided Hyperspectral Image Analysis System for Land-Cover Classification**, IEEE Trans. Geosci. Remote Sensing, 40(2): 416-425.
- Savitzky, A., Golay, M. J. E., 1964, **Smoothing and Differentiation of Data by Simplified Least Squares Procedures**, Anal. Chem. 36:1627–1639.
- Zhang, J., Rivard, B., Sanchez A.A., 2004, **Derivative Spectral Unmixing of Hyperspectral Data Applied to Mixtures of Lichen and Rock**, IEEE Transactions on geoscience and remote sensing, 42: 1934-1940.