



شناسایی ابر در تصاویر ماهواره ای اپتیکی با استفاده از طبقه بندی کننده ی ماشین بردار پشتیبان

نفیسه قاسمیان سوربنی

۱- دانشجوی کارشناسی ارشد سنجش از دور دانشکده ی مهندسی نقشه برداری و اطلاعات مکانی دانشگاه تهران

چکیده :

انواع مختلف ابر می توانند با استفاده از توصیف گرهای مختلف در تصاویر ماهواره ای اپتیکی شناسایی شوند. این توصیف گر ها می توانند مثل میانگین دربردارنده ی اطلاعات در مورد روشنایی تصویر باشند یا می توانند مثل خطای میانگین مربعات گرادبان و آنتروپی ترکیبی و همگونی دربردارنده ی اطلاعات بافت تصویر باشند و یا مانند ضرایب فرکانس بالای تبدیل ویولت و کسینوسی شامل اطلاعات لبه باشند. در این مقاله با استفاده از این توصیف گر ها از طبقه بندی کننده ی ماشین بردار پشتیبان (SVM) به منظور شناسایی نواحی ابری و غیر ابری استفاده شده است. باندهای استفاده شده به منظور شناسایی ابر باندهای ۲،۳،۴،۵،۶،۷ و ۹ سنسور تصویربردار عملیاتی زمین (OLI) ماهواره ی لندست ۸ می باشند. این باندها همان باندهای به کار گرفته شده در روش اف ماسک هستند. برخی از توصیف گر ها روی باندهای خاص دقت طبقه بندی بالاتری می دهند که به عنوان نمونه می توان به توصیف گر اختلاف مرتبه اول اشاره کرد که دقت کلی بالاتری روی باند ۴ نسبت به باند ۶ می دهد. سه شاخص دقت کلی، ضریب کاپا و دقت تولید کننده ی ابر روی مجموعه های مختلفی از توصیف گر ها که این مجموعه ها با استفاده از آنالیز کارلیشن بین توصیف گرهای مختلف به دست آمدند، مورد بررسی قرار گرفت. نتایج طبقه بندی به دست آمده نشان می دهد در صورت بررسی کارلیشن به منظور انتخاب توصیف گرهای بهینه و بهینه کردن پارامتر تنظیم (C) طبقه بندی کننده ی SVM با استفاده از روش جستجوی گرید و استفاده از داده های آموزشی مناسب می توان به دقت کلی و دقت تولید کننده ی ابر بالای 90% در طبقه بندی دست یافت. بررسی افزایش تعداد توصیف گر ها روی باند های ۲،۳ و ۴ نشان می دهد که با افزایش تعداد توصیف گر ها، دقت کلی طبقه بندی افزایش می یابد.

واژه های کلیدی : ابر، استخراج توصیف گر، SVM، تصویر لندست ۸.



۱- مقدمه

۱-۱- روش های شناسایی ابر:

روش های شناسایی ابر را می توان به دو دسته ی روش های فیزیکی و مدل های داده کاوی^۱ تقسیم کرد. مدل های فیزیکی شامل یک سری حد آستانه ها مثل انعکاس^۲، دما و... می باشند. پیچیدگی فرآیندهای شیمیایی و فیزیکی آب و هوایی باعث می شود این روش ها برای شناسایی ابر با دقت مطلوب کافی نباشند. برای مثال ابرهای چند لایه^۳ ویژگی های انعکاسی و تشعشع پیچیده ای دارند. به خاطر عدم توانایی در درک کامل ویژگی های ابر، محققین اخیرا به روش های داده کاوی روی آورده اند. [۱]

از طبقه بندی کننده های شبکه عصبی^۴، بیشترین احتمال^۵، درخت تصمیم گیری^۶ و یکی از نزدیک ترین همسایگی ها^۷ برای شناسایی ابر استفاده شده است. در مقایسه با روش های فیزیکی روش های داده کاوی دقت طبقه بندی را به طرز قابل توجهی بهبود دادند. [۱]

عظیمی^۹ در مقاله ی خود از باند مادون قرمز ماهواره ی محیطی- عملیاتی زمین آهنگ^{۱۰} برای طبقه بندی ابر استفاده کرد. از هر بلوک تصویر آمار مرتبه ی اول و دوم استخراج شده و برای آموزش و ارزیابی طبقه بندی کننده به کار گرفته شد. در این مقاله از ماشین بردار پشتیبان سلسله مراتبی^{۱۱} برای طبقه بندی^{۱۰} کلاس ابر و غیر ابر استفاده شد. نتایج اولیه نشان می دهد که این روش برای کاربردهای هواشناسی مناسب است. همچنین عظیمی بیان می کند که استفاده از باند مرئی در کنار باند مادون قرمز مشکل جداسازی انواع ابرهای مشابه مثل آلتوستراتوس^{۱۲} از استراتوس^{۱۳} را برطرف می کند. دقت کلی طبقه بندی در این مقاله تقریبا ۷۸.۵٪ به دست آمد که نسبت به حالتی که از طبقه بندی کننده ی شبکه عصبی احتمالی^{۱۴} استفاده می شود حدود ۳ درصد بهبود می یابد به علاوه داده های آموزشی به کار برده شده بیشتر از داده های آموزشی در حالت استفاده از شبکه ی عصبی بود. انتظار می رود که وقتی از ویژگی های مقدار تجزیه ی منفرد^{۱۵} هر دو باند مادون قرمز و مرئی برای آموزش^{۱۶} طبقه بندی کننده ی ماشین بردار پشتیبان استفاده شود، نتایج بهبود پیدا کند. [۲]

لی در مقاله ی خود کارایی استفاده از اطلاعات طیفی و مکانی چند باند را در شناسایی مناطق ابری و غیر ابری نشان داد. هدف او بهبود ماسک ابر مادیس ۱۷ در مواقعی بود که حد آستانه های مورد استفاده شده در الگوریتم نماینده ی دو

¹ Data mining models

² albedo

³ Multilayer clouds

⁴ classifier

⁵ Neural network

⁶ Maximum likelihood

⁷ Decision tree

⁸ 1-nearest neighbors

⁹ Azimi

¹⁰ Geostationary operational environmental satellite (GOES)

¹¹ Hierarchical support vector machine

¹² Altostratus

¹³ Stratus

¹⁴ Probabilistic neural network

¹⁵ Singular value decomposition features

¹⁶ training

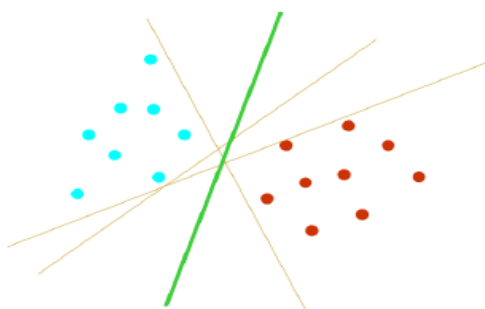
¹⁷ MODIS



کلاس مختلف نباشند. ابتدا از ماسک ابر مادیس برای طبقه‌بندی اولیه استفاده شد سپس روش بیشترین احتمال^{۱۸} به منظور بهبود طبقه‌بندی به کار گرفته شد. از محدودیت‌های این روش وابستگی دقت طبقه‌بندی به طبقه‌بندی اولیه است به علاوه نویز دستگاه و خطاهای کالیبراسیون روی الگوریتم ماسک ابر و دقت طبقه‌بندی نهایی تاثیر می‌گذارد که کالیبراسیون دقیق را ملزم می‌کند. در نهایت لی این طور نتیجه گیری می‌کند که ترکیب روش طبقه‌بندی بیشترین احتمال با الگوریتم ماسک ابر مادیس شناسایی نواحی بدون ابر و همچنین شناسایی انواع ابر را بهبود می‌دهد. [۳]

۱-۲- طبقه بندی بردار پشتیبان^{۱۹}:

هدف از طبقه بندی پیدا کردن تابعی است که داده‌ها را از هم جدا کند، این تابع با استفاده از نمونه های آموزشی موجود به دست می‌آید. در واقع هدف ایجاد یک تابعی است که بتوان آن را به نمونه های دیگر تعمیم داد. همان طور که در شکل ۱ دیده می‌شود طبقه‌بندی کننده‌های خطی زیادی وجود دارد که می‌تواند داده ها را از هم جدا کند اما فقط یکی از این طبقه بندی کننده‌ها وجود دارد که حاشیه^{۲۰} (فاصله ی بین ابر صفحه^{۲۱} و نزدیک ترین داده‌ی هر کلاس) را ماکزیمم می‌کند. به این طبقه بندی کننده‌ی خطی اصطلاحاً ابر صفحه‌ی جدا کننده‌ی بهینه^{۲۲} گفته می‌شود. در نتیجه انتظار می‌رود که این ابر صفحه‌ی بهینه در مقایسه‌ی با سایر ابر صفحه‌ها به خوبی به داده‌های دیگر تعمیم داده شود. [۴]



شکل ۱-۱: ابر صفحه‌ی جدا کننده‌ی بهینه

۱-۳- تعیین صفحه‌ی جدا کننده در حالت تعمیم یافته:

در این قسمت حالتی مورد بررسی قرار می‌گیرد که داده‌ها به صورت خطی از هم قابل جداسازی نباشند (شکل ۱-۲).

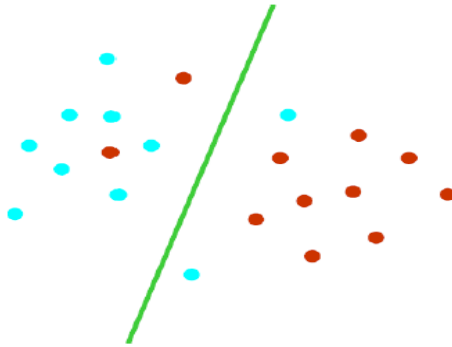
¹⁸ Maximum likelihood

¹⁹ Support vector

²⁰ margin

²¹ Hyper plane

²² Optimal separating hyper plane



شکل ۲-۱: صفحه‌ی جداکننده‌ی بهینه‌ی تعمیم یافته

برای تعمیم مسئله کورتز^{۲۳} و وپنیک^{۲۴} در سال ۱۹۹۵ متغیر غیر منفی $\varepsilon_i \geq 0$ و تابع هزینه‌ی $F_\sigma(\varepsilon)$ را به صورت زیر معرفی کردند:

$$F_\sigma(\varepsilon) = \sum \varepsilon_i^\sigma \quad \sigma > 0 \quad (1)$$

در رابطه‌ی بالا ε_i بیان کننده‌ی خطای طبقه‌بندی است. در صورتی که به مسئله‌ی بهینه‌سازی شرط مینیمم کردن خطای طبقه‌بندی اضافه شود، معادله‌ی قید در صورتی که داده‌ها به صورت خطی قابل جداسازی نباشند به صورت زیر در می‌آید:

$$y^i[\langle w, x^i \rangle + b] \geq 1 - \varepsilon_i \quad i = 1, \dots, l. \quad (2)$$

در رابطه‌ی بالا $\varepsilon_i \geq 0$. صفحه جداکننده‌ی بهینه به وسیله‌ی بردار w ای تعیین می‌شود که مقدار تابع زیر را به شرط برقراری رابطه‌ی (۲) مینیمم کند.

$$\varphi(w, \varepsilon) = \frac{1}{2} \|w\|^2 + C \sum_i \varepsilon_i \quad (3)$$

در رابطه‌ی بالا C مقدار ثابت می‌باشد. جواب این مسئله بهینه‌سازی با پیدا کردن نقطه‌ی اکسترمم تابع لاگرانژ زیر به دست می‌آید.

$$\varphi(w, b, \alpha, \varepsilon, \beta) = \frac{1}{2} \|w\|^2 + C \sum_i \varepsilon_i - \sum_{i=1}^l \alpha_i (y^i[\langle w^T, x^i \rangle + b] - 1 + \varepsilon_i) - \sum_{j=1}^l \varepsilon_j \beta_j \quad (4)$$

در رابطه‌ی (۴) α و β ضرایب لاگرانژ هستند. تابع لاگرانژ باید نسبت به w و b و ε مینیمم و نسبت به ضرایب لاگرانژ α و β ماکزیمم شود، یعنی داریم:

²³ cortes

²⁴ vapanik



$$\varphi(w, b, \alpha, \varepsilon, \beta) = \frac{1}{2} \|w\|^2 + C \sum_i \varepsilon_i - \sum_{i=1}^l \alpha_i (y^i [\langle w^T, x^i \rangle + b] - 1 + \varepsilon_i) - \sum_{j=1}^l \varepsilon_j \beta_j \quad (5)$$

$$\max_{\alpha} W(\alpha, \beta) = \max_{\alpha, \beta} \left(\min_{w, b, \xi} \Phi(w, b, \alpha, \xi, \beta) \right). \quad (6)$$

با مشتق گیری از تابع لاگرانژ نسبت به w و b و ε داریم:

$$\frac{\partial \varphi}{\partial b} = 0 \rightarrow \sum_{i=1}^l \alpha_i y_i = 0 \quad (7)$$

$$\frac{\partial \varphi}{\partial w} = 0 \rightarrow w = \sum_{i=1}^l \alpha_i y_i x_i = 0 \quad (8)$$

$$\frac{\partial \varphi}{\partial \varepsilon} = 0 \rightarrow \alpha_i + \beta_i = C \quad (9)$$

در نتیجه از معادلات (6) و (7) و (8) داریم:

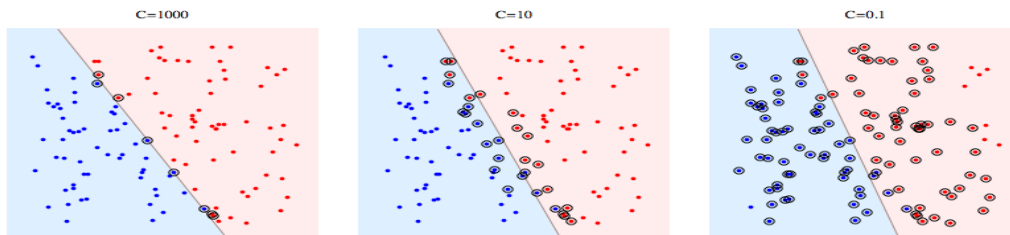
$$\max_{\alpha} W(\alpha) = \max_{\alpha} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle + \sum_{k=1}^l \alpha_k, \quad (10)$$

جواب این مسئله بهینه سازی مشابه حالتی است که داده‌ها به صورت خطی از هم قابل جداسازی باشند تفاوت آن در محدوده‌ی تعریف شده برای ضرایب لاگرانژ است.

$$0 \leq \alpha_i \leq C \quad (11)$$



مقدار C در روش کورتز باید تعیین شود یکی از پارامترهای طبقه‌بندی کننده‌ی SVM پارامتر تنظیم (C) نام دارد که بین کمینه شدن خطا و پیچیدگی مدل تعادل برقرار می‌کند. [۵] وقتی C به سمت بینهایت میل کند (افزایش مقدار C)، مقدار حاشیه کاهش می‌یابد و بالعکس وقتی C به سمت صفر میل کند (کاهش مقدار C)، مقدار حاشیه ماکزیمم می‌شود و خطای طبقه‌بندی افزایش می‌یابد. این مطلب در شکل ۱-۳ نشان داده شده است.



شکل ۱-۳: اثر تغییر پارامتر C روی طبقه بندی

۱-۳- شناسایی ابر بر اساس استخراج توصیف‌گر^{۲۵} در تصاویر سنجش از دور:

توصیف‌گرها به منظور ایجاد یک مدل برای جدا کردن قسمت‌های ابری و غیر ابری دارای اهمیت هستند. در حالت کلی توصیف‌گرها باید بتوانند به طور همزمان شروط زیر را برقرار کنند:

۱. کارایی: توصیف‌گر مورد نظر باید بتواند تا حدی توصیف‌کننده‌ی خصوصیات تارگت باشد.
 ۲. وابستگی پایین: وابستگی بین توصیف‌گرهای مختلف باید پایین باشد تا احتمال طبقه‌بندی اشتباه کاهش یابد.
 ۳. جامعیت^{۲۶}: توصیف‌گر مورد نظر باید تمام توصیف‌گرهای تارگت را در برداشته باشد.
- از مقایسه‌ی تصاویر سنجش از دوری مختلف در می‌یابیم که تفاوت نواحی ابری و غیر ابری در توصیف‌گرهای درجه خاکستری، فرکانس و بافت^{۲۷} است. [۶]

۲- روش انجام کار:

به منظور انتخاب توصیف‌گرهای مناسب برای طبقه‌بندی، از روش آنالیز وابستگی^{۲۸} استفاده شد. ابتدا کارلیشن بین باندهای ۲، ۳، ۴، ۵، ۶، ۷ و ۹ سنسور OLI^{۲۹} به منظور انتخاب باندهای با وابستگی پایین، محاسبه شد. باندهای ۶ و ۹ و باندهای ۷ و ۹ به عنوان باندهای مستقل (کارلیشن کم تر از ۶۰ درصد) در نظر گرفته شدند. بر روی توصیف‌گر^{۳۰} های میانگین درجات خاکستری، وریانس، اختلاف مرتبه اول، ضرایب فرکانس بالای تبدیل کسینوسی و تبدیل ویولت و توصیف‌گرهای بافت به دست آمده از ماتریس رخداد توام گرادیان درجات خاکستری^{۳۱} (GLGCM) شامل میانگین مربعات گرادیان درجه خاکستری، آنتروپی ترکیبی، همگنی و بعد فرکتال محاسبه شده روی باند های مستقل نیز کارلیشن محاسبه شد. بین سه شاخص^{۳۲} $R_{0.86}$ (انعکاس باند ۵)، $R_{1.38}$ (انعکاس باند ۹)، $\frac{R_{0.67}}{R_{0.87}}$ (نسبت انعکاس باند ۴

²⁵ feature

²⁶ integrality

²⁷ texture

²⁸ dependency

²⁹ Operational land imager

³⁰ feature

³¹ Gray level gradient co-occurrence matrix

³² index

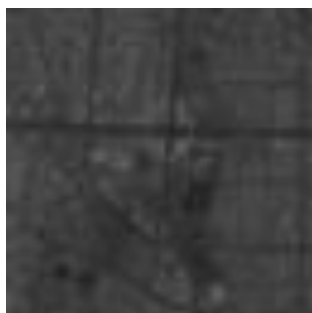


به باند ۵) که در محصول ابر مادیس مورد استفاده قرار گرفته‌اند، و باندهای مستقل و توصیف‌گرهای محاسبه شده از این باندها به منظور انتخاب توصیف‌گرهای نهایی مجدداً آنالیز وابستگی انجام شد. در نهایت ۲۳ توصیف‌گر به عنوان توصیف‌گر نهایی انتخاب شدند. این توصیف‌گرها عبارتند از: بعد فرکتال باند ۳۳^{۳۳}، آنتروپی ترکیبی ۳۴ باند ۶ و ۷ و ۹، میانگین مربعات گرادیان ۳۵ باند ۷، همگنی ۳۶ باندهای ۶ و ۹، وریانس باندهای ۶ و ۹، ضرایب جزئیات قطری باندهای ۷ و ۹ تبدیل ویولت، ضرایب جزئیات افقی باندهای ۶ و ۹ تبدیل ویولت، ضرایب جزئیات عمودی باندهای ۷ و ۹ تبدیل ویولت، میانگین باندهای ۷ و ۹، اختلاف مرتبه اول ۳۷ باندهای ۷ و ۹، ضرایب جزئیات تبدیل کسینوسی باند ۷، باند ۶، باند ۹ و $\frac{b_4}{b_5}$. کارلیشن بین برخی از این توصیف‌گرها در جدول ۱ آمده است.

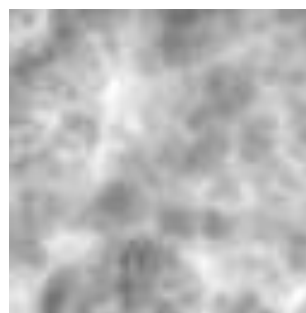
جدول ۱: کارلیشن بین توصیف‌گرهای به دست آمده از باندهای ۶ و ۷ و ۹ سنسور OLI

	B7	B9	Hfdw6 ^{۳۸}	Hfhw7 ^{۳۹}	Hfww6 ^{۴۰}	Hfww9	First difference 6
B7	1	0.519916	-0.00205	0.094793	0.089463	0.0876	0.006734
B9		1	-0.00493	0.001947	0.010092	0.055955	0.009469
Hfdw6			1	-0.00725	-0.00018	-0.0055	0.000269
Hfdw9				0.0092	0.003461	-0.0013	0.000934
Hfwh7				1	0.079669	0.064357	-0.00139
Hfwh9					0.061307	0.099845	-0.00249
Hfww6					1	0.43258	-0.00137
Hfww9						1	-0.00131
First difference 6							1

در این مقاله برای انجام طبقه‌بندی ابتدا داده‌های آموزشی با ابعاد 64×64 برای مناطق ابری و غیر ابری در هر یک از باندهای مناسب برای شناسایی ابر انتخاب شدند. در شکل ۲-۱ نمونه‌ای از داده‌های آموزشی انتخاب شده را نشان می‌دهد.



ب



الف

شکل ۲-۱: نمونه‌ای از نمونه‌های آموزشی انتخاب شده برای طبقه‌بندی. الف: نمونه آموزشی ابر در باند ۵. ب: نمونه آموزشی غیر ابر در باند ۵

³³ Fractal dimension³⁴ Hybrid entropy³⁵ Gradient Mean square error³⁶ homogeneity³⁷ First difference³⁸ High frequency diagonal wavelet coefficients band 6³⁹ High frequency horizontal wavelet coefficients band 7⁴⁰ High frequency vertical wavelet coefficients band 6



برای طبقه‌بندی از طبقه بندی کننده ی SVM استفاده شد. به منظور ارزیابی نتایج طبقه بندی 20% داده‌های آموزشی به صورت تصادفی به عنوان داده‌های تست کنار گذاشته شدند. برای بهبود عملکرد طبقه بندی کننده ی SVM، پارامتر C به روش جستجوی گرید^{۴۱} بهینه شد.

۳- نتایج:

در جدول ۲ نتیجه ی طبقه بندی با هر یک از توصیف‌گرهای اختلاف مرتبه اول، ضرایب فرکانس بالای افقی تبدیل ویولت^{۴۲}، میانگین مربعات گرادیان و ضرایب فرکانس بالای تبدیل کسینوسی بر روی باند های ۴ و ۶ به طور جداگانه آمده است.

جدول ۲: بررسی اثر طبقه بندی با توصیف‌گرهای اختلاف مرتبه اول، ضرایب فرکانس بالای تبدیل ویولت و کسینوسی روی سه

شاخص دقت کلی ۴۳، دقت تولید کننده ی ابر ۴۴، ضریب کاپا ۴۵ و خطای آموزش ۴۶

خطای آموزش	ضریب کاپا	دقت تولید کننده ی ابر	دقت کلی طبقه بندی	شماره باند	توصیف گر
0.32	0.37	0.61	0.68	۴	اختلاف مرتبه اول
0.49	0.04	0.51	0.52	۶	
0.39	0.26	0.57	0.63	۴	ضرایب فرکانس بالای افقی تبدیل ویولت
0.48	0.02	0.5	0.51	۶	
0.5	0.002	0.5	0.5	۴	میانگین مربعات گرادیان
0.5	0	0.5	0.5	۶	
0.5	0.002	0.5	0.5	۴	ضرایب فرکانس بالای تبدیل کسینوسی
0.5	0.02	0.5	0.51	۶	

از جدول ۲ این طور استنباط می شود که نتیجه ی طبقه بندی با توصیف گر اختلاف مرتبه اول و ضرایب فرکانس بالای تبدیل ویولت روی باند ۴ بهتر (دقت کلی بالاتر) از باند ۶ بوده است و در مورد توصیف‌گرهای میانگین مربعات گرادیان و ضرایب فرکانس بالای تبدیل کسینوسی با در نظر نگرفتن ضریب کاپا روی هر دو باند به طور یکسان عمل می کند.

با توجه به شکل ۳ در مورد توصیف گر اختلاف مرتبه اول از نظر بصری نیز این نیز موضوع تایید می شود. همان طور که در شکل ۳-ب دیده می شود این توصیف گر روی باند ۶ برخی از بخش‌های غیر ابری تصویر ۴-الف را به اشتباه ابری شناسایی کرده است.

⁴¹ Grid search

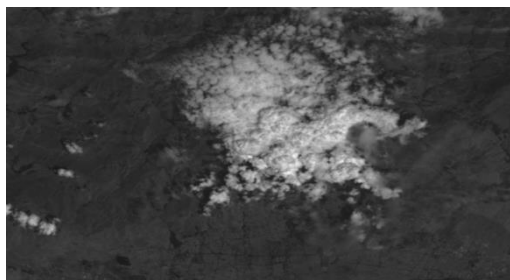
⁴² horizontal High frequency wavelet coefficient

⁴³ Overall accuracy

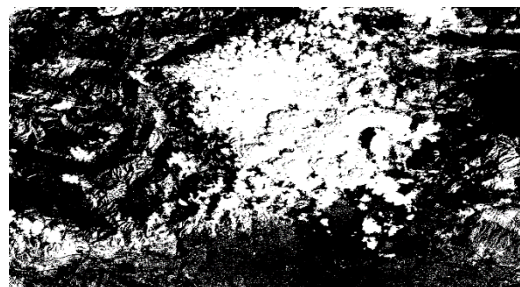
⁴⁴ Cloud producer accuracy

⁴⁵ Kappa coefficient

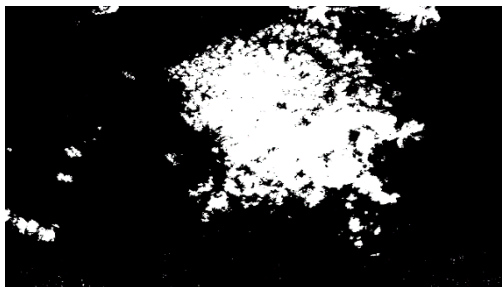
⁴⁶ Training error



الف



ب



ج

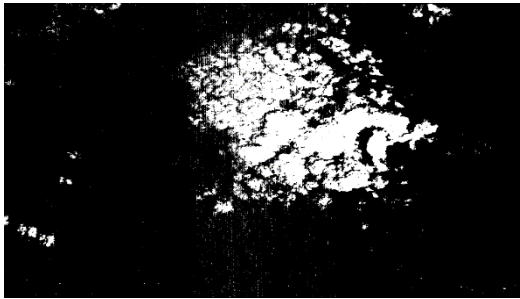
شکل ۳: الف: باند ۴ سنسور OLI ماهواره ی Landsat 8. ب: نتیجه ی طبقه بندی با توصیف گر اختلاف مرتبه اول برای باند ۶. ج: نتیجه ی طبقه بندی برای باند ۴

۴- بحث و نتیجه گیری:

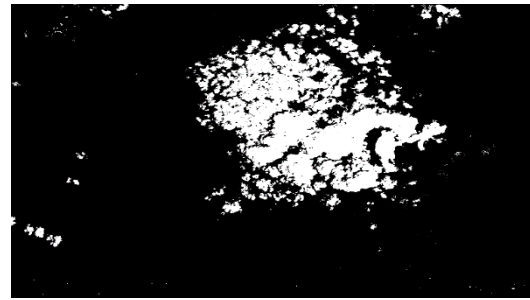
با توجه به جدول ۳ مشاهده می شود که با محاسبه ی کارلیشن بین توصیف گر ها و کم کردن فضای توصیف گر با حذف توصیف گرهایی که دارای کارلیشن بالای 60% بودند دقت کلی طبقه بندی 100% به دست آمد. دقت های بالای به دست آمده در طبقه بندی می تواند ناشی از انتخاب داده های آموزشی مناسب، انتخاب روش مناسب برای کاهش ابعاد فضای توصیف گر، انتخاب توصیف گرهای مناسب برای شناسایی ابر و انتخاب کرنل مناسب برای طبقه بندی باشد.

جدول ۳: تاثیر تغییر نوع توصیف گرهای انتخابی بر روی سه شاخص دقت کلی، ضریب کاپا و دقت تولید کننده ی ابر

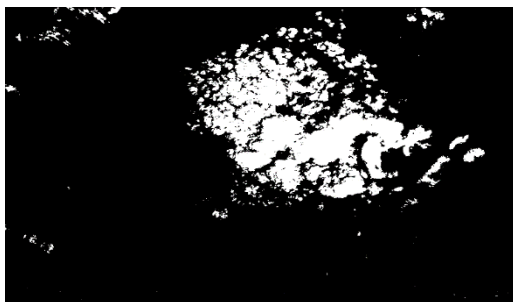
ردیف	توصیف گرهای انتخابی	دقت کلی	ضریب کاپا	دقت تولید کننده ی ابر
۱	۲۳ توصیف گر به دست آمده پس از بررسی کارلیشن بین کل توصیف گرها	۱	۱	۱
۲	حذف توصیف گر اختلاف مرتبه اول باند ۹ از توصیف گرهای ردیف ۱	۰.۹۹۹۸	۰.۹۹۹۵	۰.۹۹۹۵
۳	حذف توصیف گرهای اختلاف مرتبه اول باند های ۷ و ۹ از توصیف گرهای ردیف ۱	۰.۹۹۹۵	۰.۹۹۹۰	۰.۹۹۹۰
۴	حذف توصیف گر اختلاف مرتبه اول باند ۹ و ضرایب فرکانس بالای تبدیل کسینوسی باند ۷ از توصیف گرهای ردیف ۱	۱	۱	۱
۵	حذف توصیف گر اختلاف مرتبه اول باند ۹ از و میانگین باند های ۷ و ۹ از توصیف گرهای ردیف ۱	۱	۱	۱
۶	حذف توصیف گر اختلاف مرتبه اول باند ۹ و توصیف گرهای ضرایب جزئیات تبدیل ویولت از توصیف گرهای ردیف ۱	۰.۹۹۹۵	۰.۹۹۹۰	۰.۹۹۹۰



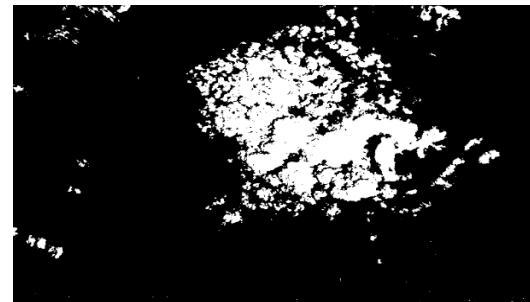
الف



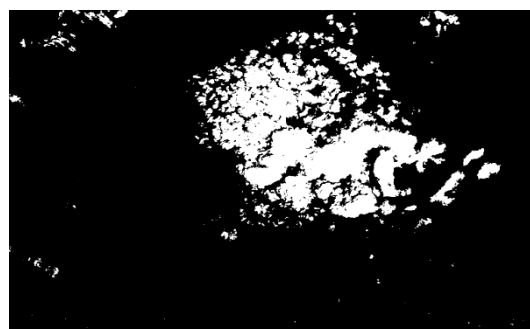
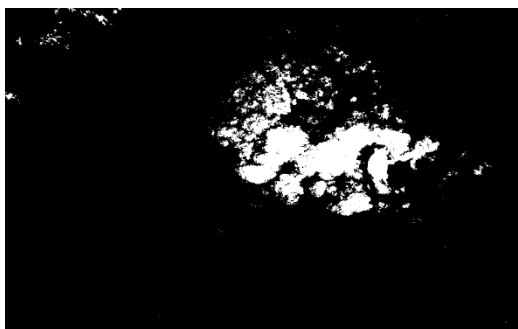
ب



ج



د



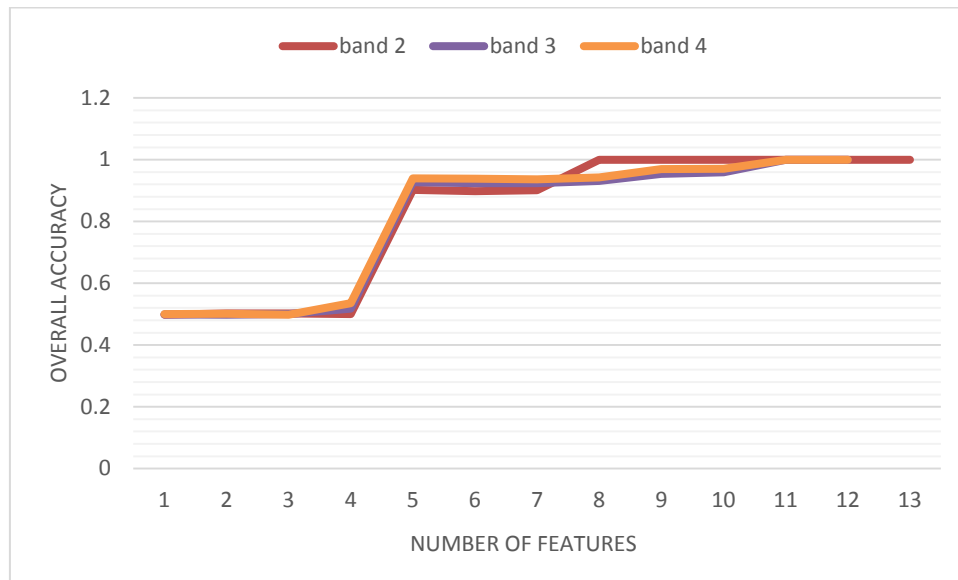
شکل ۴: ماسک های ابر به دست آمده با استفاده از توصیف گرهای جدول ۳. شکل الف ماسک ابر به دست آمده با استفاده از توصیف گر های ردیف ۱ جدول. شکل ب به دست آمده با استفاده از توصیف گرهای ردیف ۲ جدول. شکل ج به دست آمده با استفاده از توصیف گرهای ردیف ۳ جدول. شکل د به دست آمده با استفاده از توصیف گرهای ردیف ۴ جدول. شکل چ به دست آمده با استفاده از توصیف گرهای ردیف ۵ جدول. شکل ح به دست آمده با استفاده از توصیف گرهای ردیف ۶ جدول.

با مقایسه‌ی بصری نتایج طبقه‌بندی در حالت ۱ و ۲ جدول ۳ (شکل ۴-الف و ۴-ب) ملاحظه می‌شود که بعد از حذف توصیف گر اختلاف مرتبه اول باند ۹ از فضای توصیف‌گر نتیجه بصری به دست آمده بهبود پیدا کرده است و آرتیفکت^{۴۷} موجود حذف شده است. با مشاهده‌ی شکل ۴-چ می‌توان دریافت که با حذف توصیف‌گر میانگین درجات خاکستری از فضای توصیف‌گر برخی قسمت های ابری که به صورت پراکنده در تصویر وجود داشتند به خوبی شناسایی نشده اند. بنابراین می‌توان میانگین درجات خاکستری را به عنوان یک توصیف‌گر مناسب در شناسایی ابرهای پراکنده (در صورت داشتن کنتراست بالا با زمینه) معرفی کرد.

⁴⁷ artifact



روی توصیف‌گرهای استخراج شده از باند های ۲،۳ و ۴ آنالیز کارلشن انجام شد برای باند ۲ ۱۳ توصیف‌گر به عنوان توصیف‌گر های مستقل، برای باند ۳ و ۴ ۱۲ توصیف‌گر به عنوان توصیف‌گر مستقل در نظر گرفته شدند. در شکل ۶ اثر افزایش تعداد این توصیف‌گر ها روی دقت کلی طبقه بندی نشان داده شده است. از این نمودار این طور استنباط می‌شود که در مورد هر ۳ باند با افزایش تعداد توصیف‌گر ها از ۴ به ۵ دقت کلی طبقه‌بندی به طور قابل توجهی از حدود ۵۰٪ به ۹۰٪ افزایش می‌یابد.



شکل ۶: اثر افزایش تعداد توصیف‌گر ها روی باند های ۲،۳ و ۴ روی دقت کلی طبقه بندی

مراجع:

- [۱] Bo Han, Lishan Kang, and Huazhu Song, "A Fast Cloud Detection Approach by Integration of Image Segmentation and Support Vector Machine," *Advances in Neural Networks*, vol. 3973, p.1210-1215, 2006.
- [۲] Azimi-Sadjadi, M. R., and S. A. Zekavat. "Cloud classification using support vector machines." *Geoscience and Remote Sensing Symposium*, vol. 2, p. 669-671, 2000.
- [۳] J. Li, W.P Menzel, Z. Yang, R.A Frey, and S.A Ackerman, "High-spatial-resolution Surface and Cloud-type Classification from MODIS Multispectral Band Measurements", *Applied Meteorology*, vol. 42, p. 204-226, 2003.
- [۴] Steve R. Gunn, "Support Vector Classification", *Support Vector Machines for Classification and Regression*, Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science, university of Southampton, 1998, page 5.
- [۵] Farhad Samadzadegan, Hadiseh Sadat Hasani, "Determination of optimum support vector machines in classification of hyper spectral images based on genetic algorithm", *Journal of Iran information and communication technology*, 9-24, 1391.
- [۶] YU Changhui, Yuan Yuan and Miao Minjing, "CLOUD DETECTION METHOD BASED ON FEATURE EXTRACTION IN REMOTE SENSING IMAGES", 8th International Symposium on Spatial Data Quality, Hong Kong, 2013.



Cloud detection in optical satellite images using support vector machine

Nafiseh Ghasemian Sorbonney*

Ms.c student of remote sensing in Department of Geomatics, College of Engineering, University of Tehran

Abstract:

Different types of clouds can be detected in optical satellite images by using different features. These features can include brightness information of image like mean of gray values or can convey textural information of image like gradient mean square error, hybrid entropy and homogeneity or edge information like high frequency wavelet and cosine transformation coefficients. In this paper these features were given to support vector machine (SVM) classifier as inputs for detection of cloudy and non-cloudy regions. Bands 2,3,4,5,6,7,9 of operational land imager (OLI) sensor of Landsat 8 were applied as Fmask cloud detection algorithm, for cloud detection. Some features give higher overall accuracy on specific bands, for example first difference gives higher overall accuracy on band 4 in comparison to band 6. Correlation coefficients were computed between features and overall accuracy, kappa coefficient and cloud producer accuracy were computed between different subsets of features. Classification results shows that by computing correlation coefficients between features, optimal features can be chosen, in addition by applying grid search method for finding optimal penalty parameter (C) in SVM classifier and using suitable training data overall accuracy and cloud producer accuracy higher than 90% can be achieved. Increasing the number of features on bands 2,3 and 4 improve overall accuracy of classification.

Keywords: cloud, feature extraction, SVM, Landsat 8 image.

*Correspondence Address: Remote Sensing Group, School of surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. Tel: +98 9113563208.

Email: n.ghasemian@ut.ac.ir