

بهبود روش استخراج ویژگی گرادیان مبتنی بر تبدیل گسسته کسینوسی جهت بازشناسی ارقام دست‌نوشته فارسی

محمد نحوی^۱، کوروش کیانی^۲ و رضا ابراهیم پور^۳

^۱ دانشجوی کارشناسی ارشد دانشگاه سمنان و پژوهشگر پژوهشکده علوم شناختی، پژوهشگاه دانش‌های بنیادی (IPM)، nahvi@ipm.ir

^۲ استادیار دانشکده برق و کامپیوتر دانشگاه سمنان، Kourosh.kiani@aut.ac.ir

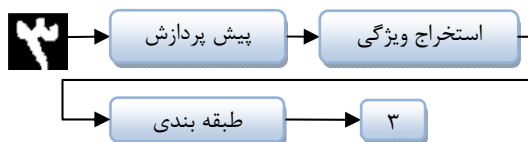
^۳ استادیار دانشکده برق و کامپیوتر دانشگاه شهید رجایی و عضو شورای علمی پژوهشکده علوم شناختی، پژوهشگاه دانش‌های

بنیادی (IPM)، Ebrahimpour@ipm.ir

چکیده - در این مقاله روش جدیدی بر مبنای تبدیل گسسته کسینوسی برای استخراج بردار ویژگی از تصاویر گرادیان ارائه شده است. پس از اعمال گرادیان، ضرایب تبدیل گسسته کسینوسی تصاویر گرادیان محاسبه و ضرایب فرکانس پایین آن به عنوان ویژگی‌های تصویر گرادیان انتخاب گردید. آزمایش‌های انجام شده بر روی مجموعه داده ارقام دست‌نویس فارسی نشان می‌دهد که روش پیشنهادی خطای نرخ بازشناسی را به میزان ۳۰ درصد کاهش می‌دهد. آزمایش‌ها بر روی مجموعه داده هدی انجام شده که شامل ۶۰۰۰۰ نمونه آموزش و ۲۰۰۰۰ نمونه آزمایش است. کلید واژه- ارقام دست‌نویس، استخراج ویژگی، گرادیان روشناپی، تبدیل گسسته کسینوسی (DCT)

باید مرزبندی صحیحی بین بردارهای ویژگی انجام‌شود بنحوی که نمونه‌های هر الگو با مرز مشخصی از بقیه نمونه‌ها تفکیک شوند. در این بخش نیز روش‌های طبقه‌بندی مختلفی از قبیل شبکه‌های عصبی^۱، K همسایگی نزدیک^۲، ماشین بردار پشتیبان^۳، تفکیک کننده خطی فیشر^۴ و ... [۱] و علاوه بر آن، از روش‌های ترکیب طبقه‌بندها برای بهبود نرخ بازشناسی استفاده می‌شود [۲].

در روش‌های که مبتنی بر استخراج ویژگی و استفاده از طبقه‌بند آموزش پذیر هستند انتخاب ویژگی از اهمیت زیادی برخوردار است. در طی سالیان اخیر ویژگی‌های متعددی معرفی شده است که در این میان ویژگی گرادیان نتایج بسیار خوبی روی ارقام لاتین داشته است [۳، ۵].



شکل ۱: سامانه تشخیص ارقام دست‌نویس

همچنین تکنیک‌های مختلف استخراج ویژگی برای ارقام دست‌نویس فارسی (عربی) پیاده‌سازی شده است که مشابه ارقام

۱- مقدمه
مسئله بازشناسی ارقام دست‌نویس^۱ را می‌توان زیر مجموعه‌ای از بازشناسی نوری کاراکتر^۲ قلمداد نمود. تشخیص ارقام دست‌نویس کاربردهای زیادی نظیر خواندن مبلغ چک بانکی، کدپستی و فرم‌های مختلف دارد که با توجه به کاربرد آن نیاز به دقت بالایی دارد [۱].

همانطور که در شکل ۱ نشان داده شده است، سامانه بازشناسی نوری کاراکتر به سه بخش اصلی، پیش‌پردازش، استخراج ویژگی و طبقه‌بندی تقسیم می‌شود. پژوهش‌های مختلفی در هر یک از این زمینه‌ها انجام گرفته است. هدف از بخش پیش‌پردازش، حذف نویز، هموارسازی و نرمالیزه کردن داده‌های ورودی است که می‌تواند نقش بسزایی در تمایز بهتر الگوها در فضای ویژگی بعد باشد.

در بخش استخراج ویژگی به هر نمونه بردار ویژگی‌ای اختصاص داده می‌شود که معرف آن نمونه در فضای ویژگی مورد نظر است و بدین ترتیب آن را از دیگر نمونه‌ها متمایز می‌کند. استخراج ویژگی و تمایز الگوها در فضای ویژگی مورد نظر اثر زیادی در بخش طبقه‌بندی دارد. برخی از این ویژگی‌های شامل ویژگی گرادیان^۳، کریش^۴، کد حلقه مکانی^۵، موجک‌ها^۶، پروجکشن‌های افقی و عمودی^۷ و ... هستند. در بخش طبقه‌بندی

کرده و به اندازه نرمال در آورد.

۲-۱- نرمال کردن اندازه تصاویر

برای نرمال سازی اندازه تصویر به $m \times n$ می توان تصویر را در هر دو راستای طولی و عرضی به اندازه مورد نظر تغییر مقیاس داد. این روش مشکلاتی ایجاد می کند، به عنوان مثال تصاویر نرمال شده رقم های صفر و یک تقریباً تمام فضای $m \times n$ را می پوشانند و تشخیص آنها مشکل می شود. در روش دیگر با نام نرمال متناسب تصویر، برای نرمال کردن تصویر با عرض w و طول l به تصویری با ابعاد مثلاً 40×40 ، اگر $l > w$ بود، l_{new} به 40 نرمال شده و $w_{new} = 40 \times l/w$ و اگر $w > l$ بود w_{new} به 40 نرمال و $l_{new} = 40 \times w/l$ نرمال می شود سپس این تصویر در مرکز یک تصویر 40×40 که تمام درایه های آن هم رنگ پس زمینه تصویر اصلی نرمال شده هستند، قرار می گیرد. استفاده از این روش برای نرمال سازی ابعاد تصویر باعث می شود که ابعاد تصاویر بطور متناسب تغییر اندازه داده شوند و تصویر حاصل، مشابهت بیشتری به تصویر اصلی داشته باشد. این روش نسبت به روش تغییر مقیاس همزمان طولی و عرضی تصویر نرخ بازشناسی را بصورت قابل ملاحظه ای افزایش می دهد. در شکل ۳ مقایسه از این دو روش آورده شده است.



شکل ۳: از چپ به راست، تصویر اصلی، تصویر تغییر مقیاس داده شده در هر دو راستای طولی و عرضی، تصویر نرمال شده به روش بکار رفته در این مقاله

۲-۲- ویژگی گرادیان

ویژگی گرادیان از ویژگی های است که نتایج بسیار خوبی روی ارقام دست نویس انگلیسی و فارسی (عربی) دارد [۴، ۷]. ویژگی گرادیان با استفاده از جهت های محلی، بخوبی مشخصه های نمونه های داخل یک کلاس و تغییرات بین کلاسی را مدل می کند [۷]. نحوه استخراج ویژگی گرادیان به ترتیب زیر است:

ابتدا تصویر ورودی به اندازه ای مثلاً 40×40 نرمال می شود. سپس با اعمال فیلتر، اطلاعات گرادیان برای هر پیکسل بدست می آید. این فیلتر می تواند عملگرهایی مانند سوبل^{۱۳} یا کریش باشد. با توجه به نتایج بدست آمده در [۱] در این مقاله از عملگر

لاتین، ویژگی گرادیان بهترین نتیجه را دربر داشته است [۱]. این ویژگی همانطور که از نامش پیداست مبتنی بر گرادیان سطح خاکستری تصویر است. از این ویژگی در فارسی هم استفاده شده و روش هایی برای بهبود آن ارائه شده است. در [۴] از ۱۶ جهت پایه گرادیان به جای ۸ جهت پایه استفاده شده که باعث بهبود 0.32% نرخ بازشناسی نسبت به روش قبل شده است و نرخ بازشناسی به 99.02% بر روی مجموعه داده هدی رسیده است. در این مقاله نیز از ویژگی گرادیان برای استخراج ویژگی از ارقام دست نویس فارسی استفاده شد و روش جدیدی بر مبنای تبدیل گسسته کسینوسی^{۱۲}، برای گردآوری بردار ویژگی از تصاویر گرادیان پیشنهاد شده است که باعث بهبود نرخ بازشناسی سامانه گردید. طبقه بند مورد استفاده شبکه عصبی پرسپترون ۳ لایه است که با تعداد مختلف نورون لایه میانی پیاده سازی گردید.

در این مقاله از مجموعه داده هدی با ۶۰۰۰۰ نمونه آموزش و ۲۰۰۰۰ نمونه آزمایش برای محک زنی ایده پیشنهادی استفاده شد. این مجموعه داده از ۱۱۹۴۲ فرم پر شده توسط افراد دیپلم شرکت کننده در کنکور سراسری استخراج شده و قدرت تفکیک پذیری نمونه های این مجموعه داده ۲۰۰ نقطه بر اینچ است [۶]. در شکل ۲ نمونه هایی از ارقام آن دیده می شود.



شکل ۲: نمونه هایی از پایگاه داده هدی

ادامه ساختار مقاله به این ترتیب است: بخش ۲ به معرفی ویژگی گرادیان و گرادیان بهبود یافته پیشنهاد شده در [۴] می پردازد. بخش ۳ نحوه بکارگیری تبدیل گسسته کسینوسی در استخراج ویژگی از تصاویر گرادیان مطرح و نتایج آزمون آن بر روی مجموعه داده هدی آورده شده است. بخش ۴ نیز به نتیجه گیری و جمع بندی پرداخته است.

۲- ویژگی گرادیان و گرادیان بهبود یافته

ویژگی گرادیان برای تصاویر خاکستری توسعه یافته و برای استخراج آن، لازم است تصاویر نرمالیزه شوند. از این رو برای استفاده در مورد تصاویر دودویی، ابتدا باید آنها را خاکستری

سوبل برای فیلترینگ تصویر استفاده شد. در شکل ۴ عملگرهای سوبل آمده اند.

-۱	۰	۱
-۲	۰	۲
-۱	۰	۱

۱	۲	۱
۰	۰	۰
-۱	-۲	-۱

شکل ۴: عملگرهای سوبل

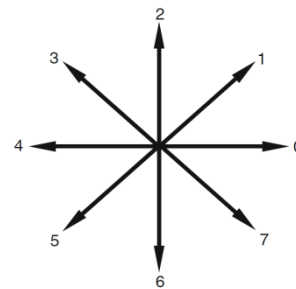
گرادیان $g(x,y)=[g_x, g_y]^T$ در موقعیت (x,y) مطابق با رابطه ۱ محاسبه می شود.

$$\begin{aligned} g_x &= f(x+1,y-1)+2f(x+1,y)+f(x+1,y+1)- \\ &\quad f(x-1,y-1)-2f(x-1,y)-f(x-1,y+1) \quad (1) \\ g_y &= f(x-1,y+1)+2f(x,y+1)+f(x+1,y+1)- \\ &\quad f(x-1,y-1)-2f(x,y-1)-f(x+1,y-1) \end{aligned}$$

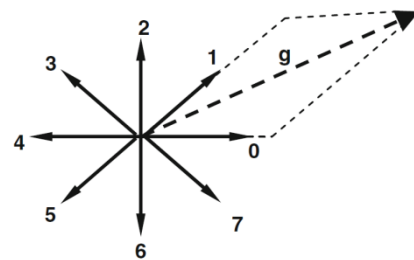
در رابطه ۲ نحوه محاسبه شدت و جهت گرادیان از دو مولفه رابطه ۱ نشان داده شده است.

$$\begin{aligned} |g(x,y)| &= \sqrt{g_x^2(x,y)+g_y^2(x,y)} \\ \angle g(x,y) &= \tan^{-1}\left(\frac{g_y(x,y)}{g_x(x,y)}\right) \quad (2) \end{aligned}$$

برای استخراج ویژگی، جهت های گرادیان به ۴، ۸ یا ۱۶ جهت کوانتیزه شده است. در شکل ۵ جهت ها برای ۸ جهت پایه گرادیان نشان داده شده است. مانند شکل ۶ اگر زاویه ای بین دو جهت پایه باشد، تصویر آن بر دو جهت پایه محاسبه می شود.



شکل ۵: ۸ جهت پایه برای گرادیان



شکل ۶: تصویر کردن یک جهت گرادیان به دو جهت پایه

به ازای هر یک از ۸ جهت پایه یک تصویر ساخته شده

است، مقدار هر پیکسل اطلاعات اندازه گرادیان در آن جهت برای پیکسل مذکور را دارد، به عبارت دیگر، مقدار هر پیکسل برابر با اندازه گرادیان تصویر اصلی در آن جهت است. حال از این ۸ تصویر نمونه برداری می شود تا ویژگی های نهایی بدست آیند. هر چه تعداد نمونه های بیشتری از هر تصویر انتخاب شود طول بردار ویژگی افزایش می یابد، به عنوان مثال اگر از هر تصویر ۱۶ نمونه انتخاب شود، در کل ۱۲۸ ویژگی خواهد شد. برای نمونه برداری از فیلتر گاوسی رابطه ۳ استفاده می شود:

$$h(x,y) = \frac{1}{2\pi\sigma_x^2} \exp\left(-\frac{x^2+y^2}{2\sigma_x^2}\right) \quad (3)$$

پارامتر واریانس σ_x نیز طبق فرمول تجربی رابطه ۴ تعیین می شود:

$$\sigma_x = \frac{\sqrt{2t_x}}{\pi} \quad (4)$$

پارامتر t_x فاصله نمونه برداری از تصویر نرمال شده است که برای ماسک 3×3 برابر با ۳ است [۴]. با فرض (x_0, y_0) به عنوان نقطه مرکزی فیلتر گاوسی و $f(x,y)$ همسایگی متناسب با اندازه ماسک، کانولوشن از رابطه ۵ محاسبه می شود:

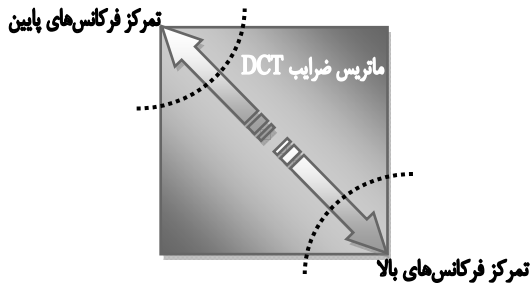
$$F(x_0, y_0) = \sum_x \sum_y f(x,y)h(x-x_0)(y-y_0) \quad (5)$$

۲-۳- ویژگی گرادیان بهبود یافته

ویژگی گرادیان بهبود یافته نتایج خوبی هم از نظر سرعت عملکرد و هم از نظر دقت همراه داشته است. در زیر خلاصه ای از این روش ارائه شده است و برای توضیحات کامل تر می توان به [۴] مراجعه نمود.

در این روش به جای ۸ جهت پایه، ۱۶ جهت پایه گرادیان در نظر گرفته شده و زوایای بین دو جهت پایه به جای آنکه بر بردارهای جهت پایه تصویر شوند، به جهت پایه نزدیک تر انتساب داده می شوند. بنابراین برای هر تصویر، ۱۶ تصویر که هر کدام نمایانگر شدت روشنایی تصویر اصلی در یکی از ۱۶ جهت پایه است، بدست می آید که شدت روشنایی با مقدار قرار گرفته در پیکسل متناظر نشان داده می شود. بمنظور استخراج ویژگی از تصاویر حاصل از اعمال گرادیان، هر تصویر به 16×16 بلوک تقسیم شده و مجموع اندازه گرادیان در هر بلوک محاسبه می شود، بنابراین از هر تصویر یک بردار 16×16 مولفه ای بدست

بعد ۱۶ بدست می‌آید. حال با کنار هم قرار گرفتن بردارهای ویژگی ۱۶ جهات پایه گرادیان، بردار ویژگی ۲۵۶ بعدی به عنوان ویژگی نهایی تصویر اصلی بدست می‌آید. در واقع در قدم اول با اعمال گرادیان که به نوعی مشتق گیری از تصویر اصلی است، فیلتر بالا گذری به تصویر اعمال شده است و سپس با انتخاب ضرایب گوشه سمت چپ بالای ماتریس ضرایب تبدیل گسسته کسینوسی بدست آمده از ماتریس‌های گرادیان، یک فیلتر پایین گذر بر تصاویر گرادیان اعمال شده است.



شکل ۷: توزیع فرکانسی در ماتریس ضرایب DCT

می‌توان گفت بردار ویژگی بدست آمده از هر تصویر گرادیان، اطلاعات موثرتر (به دلیل انتخاب ضرایب فرکانس پایین تبدیل گسسته کسینوسی) و فرکانس بالای (به دلیل اعمال گرادیان در مرحله نخست) تصویر اصلی را داراست. برای اعمال این بردار ویژگی به طبقه‌بند، اندازه هر بردار محاسبه و مولفه‌های هر بردار به اندازه بردار تقسیم شدند.

برای طبقه‌بندی همانند روش استفاده شده در بخش ۲-۳، از یک شبکه عصبی پرسپترون ۳ لایه استفاده شده است که نتایج آن برای تعداد نورون لایه میانی مختلف در جدول ۲ آورده شده است.

جدول ۲: نتایج میانگین ۵ بار تکرار روش پیشنهادی

تعداد نورون لایه میانی	۷۰	۶۰	۵۰
نرخ بازشناسی داده‌های آزمایش	۹۹.۱۵	۹۹.۰۵	۹۸.۹۸

در شکل ۸ نرخ بازشناسی روش پیشنهادی در کنار روش گرادیان بهبود یافته آورده شده است. همانطور که در شکل ۸ مشاهده می‌شود استفاده از تبدیل گسسته کسینوسی بمنظور استخراج ویژگی از تصاویر گرادیان نرخ بازشناسی را بهبود داده است.

می‌آید و با توجه به اینکه ۱۶ تصویر برای جهت‌های مختلف وجود دارد، بردار ویژگی نهایی ۲۵۶ مولفه خواهد داشت. همچنین برای افزایش سرعت، به جای محاسبه ۱۶ تصویر جهت‌های پایه گرادیان، مستقیماً حاصل جمع بلوک‌ها محاسبه شده است [۴]. نتایج حاصل از پیاده سازی این ویژگی به همراه شبکه عصبی پرسپترون ۳ لایه با ساختار ۱۰ نورون خروجی، ۲۵۶ نورون ورودی و تعداد متفاوت نورون لایه میانی به عنوان طبقه‌بند در جدول ۱ آورده شده است.

جدول ۱: نتایج میانگین ۵ بار تکرار برای ویژگی گرادیان بهبود یافته

تعداد نورون لایه میانی	۷۰	۶۰	۵۰
نرخ بازشناسی داده‌های آزمایش	۹۸.۷۸	۹۸.۷۲	۹۸.۶۷

تقسیم هر تصویر گرادیان به ۱۶ بلوک ۴×۴ و استفاده از حاصل جمع بلوک‌های ۴×۴ به عنوان ویژگی تصاویر گرادیان به دلیل نوعی متوسط گیری از تصویر و اعمال فیلتر پایین‌گذر نمی‌تواند اطلاعات جامع و کاملی را جهت طبقه‌بندی از تصاویر گرادیان استخراج کند. با استفاده از تبدیل گسسته کسینوسی می‌توان اطلاعات کاملتری از تصاویر گرادیان بدست آورد. ویژگی مهم تبدیل گسسته کسینوسی تمرکز اطلاعات در ضرایب با فرکانس پایین است که علاوه بر کاهش بعد تصاویر گرادیان، اطلاعات مهم جهت طبقه‌بندی آنرا نیز حفظ می‌کند. در بخش بعد نحوه استفاده از تبدیل گسسته کسینوسی بمنظور استخراج ویژگی از تصاویر گرادیان توضیح داده شده است.

۳- استخراج بردارهای ویژگی از تصاویر گرادیان با استفاده از ضرایب DCT

در روش پیشنهادی این مقاله، به جای روش گفته شده در بخش ۲-۳ بمنظور بدست آوردن بردار ویژگی از تصاویر گرادیان ۱۶ جهت پایه، از تبدیل گسسته کسینوسی برای این منظور استفاده شده است. مانند روش قبل، ابتدا تمام تصاویر به ابعاد ۴۰×۴۰ نرمال می‌شوند. سپس با استفاده از عملگر سوبل ۱۶ ماتریس، معادل ۱۶ جهت گرادیان برای هر تصویر محاسبه می‌شود. حال از هر کدام از این تصاویر تبدیل گسسته کسینوسی دو بعدی گرفته می‌شود و بلوک ۴×۴ گوشه بالا سمت چپ هر ماتریس به عنوان ویژگی آن تصویر انتخاب می‌شود. بنابراین ۱۶ مولفه ویژگی از هر کدام از ۱۶ جهت پایه گرادیان بدست می‌آید که از کنار هم قرار گرفتن آنها یک بردار ویژگی با

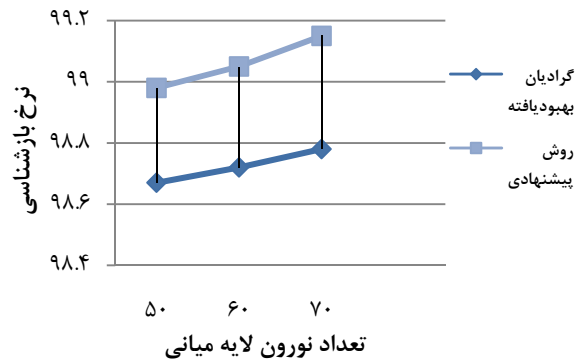
سپاسگزاری

لازم می دانم از راهنمایی های آقایان حسین خسروی و مهدی طبسیان و خانم نینا طاهری برای بهبود نگارش مقاله سپاسگذاری نمایم.

مراجع

- [1] S. Abdleazeem, E.El-Sherif, "Arabic handwrittendigit recognition". IJDAR, 2008. 11: p.127-141.
- [2] A. F. R. Rahman, M. C. Fairhurst, "Multiple classifier decision combination strategies for character recognition: A review". IJDAR, 2003. 5: p.166-194.
- [3] O.D. Trier, A.K. Jain and T. Taxt, "Feature extraction methods for character recognition – a survey", Pattern Recogniton, 1996. 29(4): p. 641-662.
- [4] حسین خسروی، احسان‌اله کبیر، "معرفی دو ویژگی سریع و کارآمد برای بازشناسی ارقام دست‌نویس فارسی"، چهارمین کنفرانس ماشین بینایی و پردازش تصویر ایران، دانشگاه فردوسی مشهد، ۲۵ الی ۲۶ بهمن ماه ۱۳۸۵.
- [5] H. Soltanzadeh, M. Rahmati, "Recognition of Persian handwritten digits using image profile of multiple orientations" Pattern Recognition Letters, 2004, 25: p.1569-1576.
- [6] H. Khosravi, E. Kabir, "Introducing a very large dataset of handwritten Farsi digits and a study on their varieties", Pattern Recognition Lett, 28: p.1133-1141.
- [7] C.-L. Liu, K. Nakashima, H. Sako, H. Fujisawa, "Handwritten digit recognition: benchmarking of state-of-the-art techniques", Pattern Recognition, 2003, 36: p.2271-2285.
- [8] C.-C. Liu, K.-W. Chuang, "An outdoor time simulation schame based on support vector regression with radial basis function on DCT domain", Image and Vision Computing, 2009, doi:10.1016/j.jmavis.2009.04.007.

مقایسه نرخ بازشناسی دو روش پیاده سازی شده



شکل ۸: نمودار مقایسه تغییرات نرخ بازشناسی نسبت به تغییرات نورون لایه میانی برای دو روش پیاده سازی شده در این مقاله.

۴- نتیجه گیری و جمع بندی

در این مقاله از تبدیل گسسته کسینوسی برای استخراج بردارهای ویژگی از تصاویر جهت‌های پایه گرادیان برای افزایش نرخ بازشناسی ارقام دست‌نویس فارسی استفاده شد. این روش با توجه به نتایج بخش ۲-۳ و بخش ۳، نرخ بازشناسی را به میزان ۰.۳۸ درصد افزایش داده است. یادآوری مجدد این مسئله ضروری است که مجموعه داده آزمایش شامل ۲۰۰۰۰ نمونه است و ۰.۳۸ درصد افزایش نرخ بازشناسی بر روی آن مقدار قابل توجهی است و به معنی بازشناسی صحیح ۷۰ نمونه بیشتر است. شایان ذکر است در [۴] نرخ بازشناسی گزارش شده برای روش گردیان بهبود یافته بالاتر از روش مشابه پیاده سازی شده در این مقاله است (جدول ۱) که علت این امر تفاوت در روش پیاده سازی و طبقه‌بندی است. بهمین دلیل برای مقایسه دقیق بین روش گردیان بهبود یافته و روش پیشنهادی این مقاله، هر دو روش با طبقه‌بندهای دقیقاً یکسان پیاده‌سازی شدند، بنابراین بهبود نرخ بازشناسی روش پیشنهادی نسبت به روش گردیان بهبود یافته مستقل از نتیجه گزارش شده در [۴] قابل استناد است.

در عوض این روش نسبت به روش گردیان بهبود یافته کندتر است. در پیاده‌سازی انجام شده توسط نرم افزار MATLAB، تاخیر محاسبه شده برای استخراج ویژگی از هر نمونه نسبت به روش استفاده شده در [۴] حدود ۱ میلی ثانیه بیشتر بود. با توجه به تاخیر سیستم‌های مکانیکی و همچنین امکان بهینه‌سازی نرم‌افزار از نظر کاهش محاسبات، می‌توان مصالحه بین افزایش دقت و کاهش عملکرد از نظر سرعت در این روش را نسبت به روش ارائه شده در [۴] مثبت ارزیابی نمود.

- 1- handwrittend digit
- 2 - Optical Character Recognition(OCR)
- 3 - Gradient
- 4 - Kirsch
- 5 - Local chain code
- 6 - wavelet
- 7 - Vertical and horizontal projections
- 8 - Neural networks
- 9 - K-Nearest neighbor
- 10 - Support Vector Machines(SVM)
- 11 - Fisher linear discriminant
- 12 - Discrete Cosine Transform(DCT)
- 13 - Sobel operator