

نازک‌سازی کاراکترهای تایپ شده فارسی به کمک یک الگوریتم جدید مبتنی بر عملگرهای ریخت شناسی

مجید نیک‌زر^۱، احمدرضا نقش‌نیلچی^۲

۱ عضو هیات علمی دانشگاه یزد و دانشجوی دکتری، گروه هوش مصنوعی، دانشگاه اصفهان

mnikzar@yazd.ac.ir

۲ عضو هیات علمی دانشیار، گروه هوش مصنوعی، دانشگاه اصفهان

nilchi@yahoo.com

چکیده

نازک‌سازی، یک نوع پیش‌پردازش روی تصویر کاراکترها است که طی آن عرض تمام بخش‌های هر کاراکتر به یک پیکسل کاهش می‌یابد. در این مقاله روشی برای نازک‌سازی حروف تایپی فارسی ارائه شده است که از طریق تکرارهای متوالی کار نازک‌سازی را انجام می‌دهد، این کار به کمک حفر سطوح و حذف پیکسل‌های دندان‌های انجام می‌شود. این مقاله قصد دارد با استفاده از الگوریتم‌های پایه ریخت‌شناسی به نازک‌سازی کاراکترهای تایپ شده فارسی به گونه‌ای بپردازد که تشخیص بصری کاراکترهای زبان فارسی را آسان سازد. با توجه به زیاد بودن تعداد نقاط دندان‌های در اطراف کاراکترها، این روش از سرعت بالایی برخوردار است. نتایج حاصل از پیاده‌سازی نشان داد که این روش علاوه بر حفظ پیوستگی، نازک‌سازی در حد یک پیکسل و بهبود سرعت قادر به صاف نمودن منحنی‌های افقی و عمودی در شکل کاراکترها همراه با حفظ دندان‌های حروف می‌باشد. این موضوع کمک مؤثری به تشخیص بصری کاراکترها می‌کند. اگرچه کارایی بصری این روش نسبت به روش ژانگ-سوئن اندکی ضعیف‌تر است، اما زمان مورد نیاز در این روش نسبت به روش ژانگ-سوئن بسیار کمتر است.

کلمات کلیدی

Thinning; Skeleton; Morphological operators; Medial axis; Optical Character Recognition

خطوط یا منحنی‌های باریک قابل تجزیه تبدیل می‌کند. اکثر الگوریتم‌های نازک‌سازی یک سری خطوط یا شاخک‌های^۱ اضافی تولید می‌کنند که باعث تغییر شکل کاراکترها شده و عمل تشخیص را دچار مشکل می‌کنند. میزان و نوع شاخک‌های اضافی بستگی به نوع الگوریتم استفاده شده دارد.

در چند دهه گذشته الگوریتم‌های نازک‌سازی بسیاری توسعه یافته‌اند [1-9]. انواع روش‌های نازک‌سازی را به دو دسته کلی می‌توان تقسیم نمود که عبارتند از جستجوی سطر به سطر^۲ و بر اساس محور میانی^۳. در روش‌های جستجوی سطر به سطر عمدتاً از عملگرهای ریخت‌شناسی^۴ برای نازک‌سازی یا پیدا نمودن الگوهای مورد نظر استفاده می‌شود. روش‌های جستجوی سطر به سطر خود به دو دسته ترتیبی و موازی تقسیم می‌شوند. در روش ترتیبی پیکسل‌های کاراکتر به صورت سطر به سطر بررسی می‌شوند یا مرز کاراکتر پیمایش

۱- مقدمه

نازک‌سازی یک نوع پردازش جهت کاهش تعداد پیکسل‌های تشکیل دهنده اشیاء در تصویر می‌باشد به نحوی که طی آن عرض هر عنصر با حفظ شکل یک پیکسل خواهد شد. با عمل نازک‌سازی بی‌آنکه شکل کاراکتر و اطلاعات با ارزش از بین برود، حجم اطلاعات به نحوی کاسته می‌شود که علاوه بر نیاز به حافظه کمتر، پردازش آن نیز ساده‌تر می‌شود [2].

نازک‌سازی یک پیش‌پردازش اصلی و مهم برای کاربردهایی مانند تشخیص دست خط، تشخیص بصری کاراکترها^۱، تشخیص اثر انگشت، پردازش تصاویر پزشکی و غیره می‌باشد. نازک‌سازی کمک عمده‌ای به سرعت بخشیدن و تسهیل نمودن کاربردهای مذکور نموده و همچنین مصرف حافظه را کاهش می‌دهد. نازک‌سازی، کاراکترها را به یک سری