الگوریتم اصلاح شده جداسازی حروف در متون چاپی با برچسب زدن به کانتور بالایی کلمات

حسین نظام‌آبادی‌پور* ، احسان اله کبیر** و رضا عزمی***
بخش مهندسی برق، هسته پژوهشی پردازش تصویر، دانشگاه شهید باهنر کرمان
بخش مهندسی برق، دانشگاه تربیت مدرس
بخش مهندسی کامپیوتر، دانشگاه الزهرا

(دریافت مقاله: ۸۰/٤/۳۰ – دریافت نسخه نهایی: ۸۲/٦/۲٤)

چکیده – در این مقاله با اصلاح الگوریتم عزمی که مبتنی بر کانتور بالایی کلمات است، الگوریتم جداسازی مناسبی برای متون با کیفیت چاپی پایین ارائه شده است. برای حل مشکل نایکنواختی نوار زمینه خط،روش مناسبی برای تعیین نوار زمینه محلی پیشنهاد شده و با اصلاح روش بر چسب زنی کانتور بالایی وتکمیل قواعد جداسازی، دقت الگوریتم افزایش داده شده است. نرخ جداسازی درست حروف ۹۷٪ است. بر اساس نتایج به دست آمده، بررسی دقیقی درباره علل خطاها ارائه شده است که می‌تواند راهگشای تحقیقات بعدی در این زمینه باشد
.

واژگان کلیدی: جداسازی حروف، متون چاپی فارسی، نوار زمینه محلی، کانتور بالایی، هیستوگرام،کد زنجیره‌ای، برچسب زنی

A Modified Character Segmentation Algorithm for Farsi Printed Text Using Upper Contour Labelling

H. Nezamabadi-Pour , E. Kabir and R. Azmi
Image Processing Center, Departmant of Electrical Engineering, Shahid Bahonar University of Kerman
Department of Electrical Engineering, Tarbiat Modarres University
Department of Computer Engineering, Alzahra University

Abstract: *In this paper, a modified segmentation algorithm for printed Farsi words is presented. This algorithm is based on a previous work by Azmi that uses the conditional labeling of the upper contour to find the segmentation points. The main objective is to improve the segmentation results for low quality prints. To achieve this, various modifications on local baseline detection, contour labeling and segmentation rules have been applied. In an experiment, the correct segmentation rate was 97%. Based on the results obtained, a detailed error analysis is presented which should be useful for furthur research on this topic.*

Keywords: *Character Segmentation, Farsi Printed text, Local baseline, Upper contour, Histogram, Segmental code, Labeling*

این تحقیق در چهارچوب طرح ملی" بازشناسی متون چاپی و حجم محدود کلمات دستنویس" انجام شده است.

فهرست علائم

d برچسب پایین	lbl لبه پایینی نوار زمینه
$h_0(n)$ هیستوگرام ارتفاع نقاط	m برچسب وسط
روی کانتور بیرونی با کد۰	p یک نقطه روی کانتور
$h_{20}(n)$ هیستوگرام هموار شده ارتفاع نقاط روی کانتور	pt پهنای قلم محلی
بیرونی با کد۰	ptg پهنای قلم
$h_4(n)$ هیستوگرام ارتفاع نقاط روی کانتور بیرونی با کد٤	ubl لبه بالایی نوار زمینه
$h_{24}(n)$ هیستوگرام هموار شده ارتفاع نقاط روی کانتور	u برچسب بالا
بیرونی با کد٤	x(p) طول نقطه p
	y(p) عرض نقطه p

۱- مقدمه

بازشناسی حروف، محوریترین بخش در پردازش مستندات است.در متون چاپی انگلیسی، مسئلهٔ جداسازی تنها در چسبیدگیهای ناخواسته حروف مطرح می‌شود که ناشی از کیفیت پایین چاپ یا تصویربرداری است[۱]. در متون دستنویس انگلیسی که حروف ممکن است به شکل پیوسته نوشته شوند، تحقیقات زیادی در مورد جداسازی کلمات به حروف و زیر حروف انجام شده است که بعضی از ایده‌های مطرح در آنها را می‌توان برای جداسازی حروف چاپی فارسی و عربی نیز به‌کار گرفت[۲]. در بازشناسی متون چاپی فارسی و عربی سه رویکرد کلی قابل تصور است[۳]. در رویکرد اول جداسازی و بازشناسی حروف به طور مستقل از یکدیگر انجام می‌شود. در رویکرد دوم که برای مجموعه محدودی از کلمات مناسب است، از جداسازی حروف صرف‌نظر می شود و کل کلمه یا زیر کلمه به عنوان یک الگوی واحد بازشناسی می شود. در رویکرد سوم جداسازی و بازشناسی حروف به شکل ترکیبی انجام می‌شود. در رویکردهای اول وسوم جداسازی حروف یکی از اساسی ترین مراحل است و نتایج بازشناسی نهایی به شدت وابسته به آن است.

در زمینه جداسازی حروف چاپی فارسی و عربی تحقیقات قابل توجهی انجام شده است. در بعضی از این تحقیقات، با استفاده از قواعد حاکم بر یک قلم خاص روشهایی برای جداسازی حروف نوشته شده با آن قلم ارائه شده است[٤ و٥] برای جداسازی حروف از ویژگیهای مختلفی می‌توان استفاده کرد. در بعضی روشها از کانتور کلمات استفاده شده است [٦ و ٧]. در هر ستون تصویر یک کلمه، فاصله میان دورترین نقاط سیاه حساب می‌شود. اگراین مقدار از یک حدآستانه کمتر باشد، آن ستون به عنوان نامزدی برای جداسازی در نظر گرفته می‌شود[٦]. در برخی تحقیقات، جداسازی حروف براساس هیستوگرام عمودی نقاط سیاه انجام شده است[۸-۱۱]. در اینجا فرض شده است که مقدار هیستوگرام عمودی در نقاط اتصال حروف از مقدار متوسط این هیستوگرام در کلمه کمتر است. در الگوریتم دیگری، از پروفیل بالایی کلمات برای جداسازی حروف آنها استفاده شده است [۱۲ و ۱۳]. در این الگوریتم به نقاط پروفیل بالایی کلمه با توجه به موقعیت آنها نسبت به نوار زمینه بر چسبهای مناسبی زده می‌شود. سپس محلهای اتصال حروف با اعمال قواعد خاصی پیدا می‌شوند.در تحقیق دیگری از نازکسازی استفاده شده است[١٤]. در این تحقیق، ابتدا از تصویر نازک شده زیر کلمه استخراج می‌شود. سپس کدهای فریمن مربوط به آن برای تعیین نقاط جداسازی به یک درخت تصمیم دودویی[1] سپرده می‌شوند.

در روش دیگری، از برچسب زدن به کانتور بالایی کلمات استفاده شده است[۳،۱۵،۱٦]. این روش نسبت به سایر روشها مزایایی دارد که در اینجا به مهمترین آنها اشاره می‌شود: الف) به همپوشانی عمودی حساسیت ندارد. ب) به پهنای قلم حساس نیست. ج) برای قلمهای معمول و متداول، به نوع قلم حساس

نیست. د) بعضی از حروف درمرحله جداسازی، بازشناسی هم می شوند.

در این مقاله اصلاح الگوریتم مرجع[۳] برای جداسازی حروف در متون با کیفیت چاپی نامناسب مورد نظر است[۱۶]. *عدم توانایی این الگوریتم در رویارویی با مشکلات متون با کیفیت چاپی نامناسب از جمله یکنواخت نبودن پهنای قلم در یک خط و حتی یک زیر کلمه، بالا و پایین نوشته شدن زیر کلمات و نرم و کوتاه بودن دندانه‌ها از ضعفهای آن محسوب می‌شود. در الگوریتم اصلاح شده، بسیاری از نقاط ضعف الگوریتم عزمی بر طرف شده است که در بخشهای بعدی به آن پرداخته می‌شود. لازم به ذکر است که در این تحقیق از متون ساده استفاده شده است که شامل شکل، جدول یا گراف نیستند.*

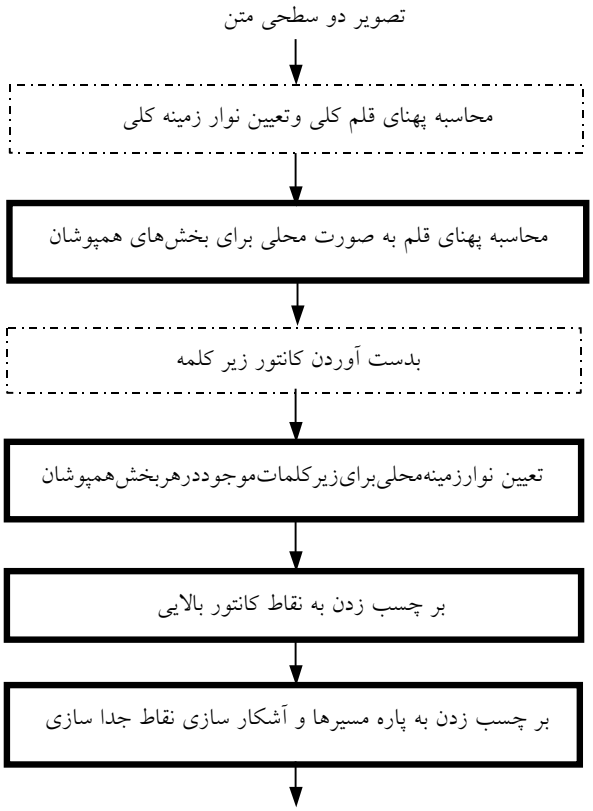
بخش دوم این مقاله به کلیات الگوریتم جداسازی می‌پردازد. بخش سوم به پردازشهای اولیه اختصاص دارد. در بخش چهارم، چگونگی تعیین نوار زمینه محلی بیان می‌شود. بخش پنجم نحوهٔ برچسب زنی به نقاط کانتور بالایی را نشان می‌دهد.در بخش ششم روش برچسب زنی پاره مسیرها وقوانین جداسازی ارائه می‌شود.بخش هفتم به نتایج وبررسی دقیق انواع خطاها می‌پردازد. در بخش هشتم اثرات درجه تفکیک بررسی می‌شود و نتیجه گیری نهایی در بخش نهم مطرح می‌شود.

۲- کلیات الگوریتم جداسازی

نمودار جعبه‌ای عمومی الگوریتم جداسازی، در شکل۱ آمده است. چار چوب اصلی این الگوریتم از مرجع[۳] گرفته شده است. *ورودی الگوریتم، تصویر متن به‌صورت دو سطحی است. در اولین مرحله خطوط متن از یکدیگر جدا می‌شوند. سپس برای هر خط متن پهنای قلم کلی محاسبه شده و موقعیت نوار زمینه تعیین می‌شود. برای محاسبهٔ پهنای قلم، هر خط از متن ورودی در راستاهای عمودی و افقی جاروب می‌شود و قطعات متصل از عناصر سیاه در هر ستون و ردیف بر حسب اندازه شمرده می‌شوند. اندازه‌ای که بیشترین فراوانی را داشته باشد به عنوان پهنای قلم در نظر گرفته می‌شود. نوار زمینه، نواری است با پهنای قلم که بیشترین تعداد عناصر سیاه در تصویر یک خط متن را در خود داشته باشد، شکل(۲). برای تعیین نوار زمینه کلی، ابتدا هیستوگرام افقی عناصر سیاه یک خط متن را محاسبه می‌شود. سپس ردیفهای متوالی این هیستوگرام به اندازه پهنای قلم، با یکدیگر جمع شده و هیستوگرام هموار شده‌ای فراهم می‌شود. محل ماکزیمم این هیستوگرام، نوار زمینه کلی برای آن خط از متن را مشخص می‌کند. به عبارت دیگر می‌توان گفت: نوار زمینه نواری است که کلمات روی آن نوشته می‌شوند. بعد از این تصویر هر خط متن بطور مجزا پردازش می‌شود.*

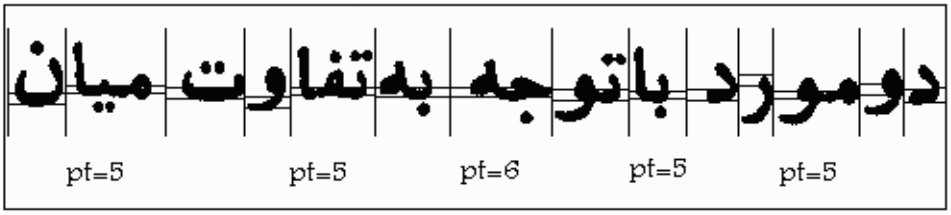
اولین مرحله از پردازش خطوط تصویر متن، پیدا کردن بخشهای همپوشان عمودی است. بخشهای همپوشان به مجموعه‌ای از زیرکلمات گفته می‌شود که هیستوگرام عمودی آنها پیوسته است، شکل(۲). بعد از جداسازی بخشهای همپوشان عمودی، پهنای قلم به‌صورت محلی برای هر بخش محاسبه شده و کانتور زیرکلمات آن تعیین می‌شود. با استفاده از اطلاعات کانتور، نوار زمینه محلی برای هر بخش همپوشان تعیین می‌شود. در ادامه، نقاط کانتور بالایی بر اساس فاصلهٔ آنها با نوار زمینه محلی و شیب منحنی برچسب زده می‌شوند. سپس کانتور بالایی به پاره مسیرهایی تجزیه می‌شود که برچسب خاص خود را می‌گیرند. نقاط جداسازی با استفاده از قواعدی خاص آشکار می‌شود.

برای جداسازی حروف بهتراست مرز بین حروف تک دندانه به عنوان نقطه جداسازی تعیین شود و حروف "س"، "ش"، "ص" و "ض" با وجود داشتن دندانه به زیر حروف شکسته می‌شوند. در متون با کیفیت مناسب، می‌توان در مرحله پس پردازش مشکل شکسته شدن این حروف به زیر حروف را برطرف کرد.این کار برای متون با کیفیت نامناسب، به سادگی میسر نیست.بر خلاف تحقیق قبلی[۳]، در اینجا سعی می‌شود دندانه‌ها بدون توجه به اینکه متعلق به چه حرفی هستند،جدا شوند و کارایی این رویکرد نسبت به رویکرد قبلی بررسی شود.تاکید برتعیین دقیق نوار زمینه در این مقاله، برای رسیدن به دقت بالا در جداسازی حروف به خصوص دندانه‌هاست.



شکل ۱- در نمودار جعبه‌ای عمومی الگوریتم جداسازی، جعبه‌هایی که با خط پر مشخص شده‌اند، نسبت به آنچه در مرجع [۳] آمده است تغییراتی انجام شده است.

(الف)



(ب)

شکل ۲- اثر اصلاح پهنای قلم به صورت محلی

الف) برای هر بخش همپوشان نوارزمینه باتوجه به پهنای قلم کلی محاسبه شده است(بخشهای همپوشان توسط‌خطوط عمودی از یکدیگر جدا شده‌اند)

ب) برای هر بخش همپوشان پهنای قلم محلی محاسبه شده است که در زیر آن نوشته شده و برای بخشهایی که در شکل نوشته نشده،pt =7 است.

۳ - پردازش‌های اولیه

مجموعه تصاویر استفاده شده در این آزمایش، از کتابهای مختلف تهیه شده است. در انتخاب این مجموعه سعی شده است که از متونی با قلمهای متفاوت استفاده شود. از هر صفحه متن بادرجه تفکیک ۳۰۰ dpi به‌صورت دو سطحی تصویر برداری شده است. تصویرها با یک روبشگرScan Jet 6300HP با تنظیم پیش فرض گرفته شده است.

۳-۱- محاسبهٔ پهنای قلم و تعیین نوار زمینه کلی

الگوریتم جداساز، تصویر متن را به عنوان ورودی دریافت می‌کند.در اولین مرحله ،خطوط متن با استفاده از هیستوگرام افقی نقاط سیاه از یکدیگر جدا می‌شوند. سپس پهنای قلم کلی، ptg، برای هر خط متن محاسبه می‌شود.مرحلهٔ بعد، تعیین نوار زمینه کلی براساس هیستوگرام افقی است. سپس بخشهای همپوشان با استفاده از هیستوگرام عمودی نقاط سیاه از یکدیگرجدا می‌شوند [۳].

۳-۲ محاسبهٔ پهنای قلم به صورت محلی برای بخشهای همپوشان

در متون با کیفیت پایین مثل متون قدیمی و متونی که با دستگاه کپی نامناسب تهیه شده اند، مواردی مشاهده می‌شود که پهنای قلم در تمام متن ثابت نیست و در بخشهایی از متن تجمع جوهر و در بخشهایی دیگر خوردگی وجود دارد. با توجه به اینکه الگوریتم جدا ساز وابستگی زیادی به محل دقیق نوار زمینه دارد و پهنای قلم در تعیین عرض نوار زمینه اثر دارد، لازم است پهنای قلم به صورت محلی تعیین شود.

نحوهٔ محاسبه پهنای قلم محلی،pt، مانند محاسبه ptg است، با این تفاوت که pt برای یک بخش همپوشان به‌دست می‌آید. اگر پهنای قلم محلی از ptg ۰/٦٥ بزرگتر و از ptg ۱/۱ کوچکتر باشد، تأیید می شود. در غیر این‌صورت pt برای آن بخش همپوشان همان ptg در نظر گرفته می‌شود. این ضرایب بر اساس تجربه به‌دست آمده‌اند. شکل(۲) پهنای قلم کلی را برای یک عبارت و پهنای قلم محلی را برای بخشهای همپوشان آن نشان می دهد.

۳-۳ به‌دست آوردن کانتور زیر کلمه

برای به‌دست آوردن کانتور زیر کلمه ، سمت راست ترین نقطه در هر زیر کلمه را پیدا کرده و کانتور آن را با پیمایش پاد ساعتگرد به شکل کد فریمن با همسایگی ۸ استخراج می‌کنیم.

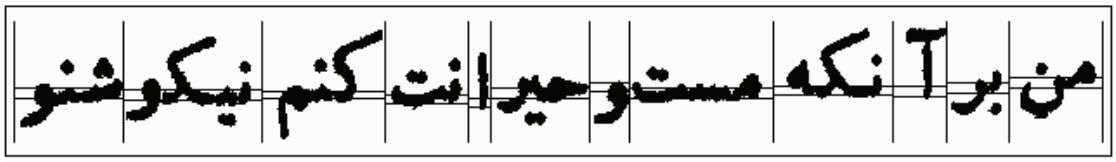
٤- تعیین نوار زمینه محلی

تعیین دقیق نوار زمینه نقش مهمی در الگوریتم جداسازی دارد. اگر به دلیل تصویر برداری با درجهٔ تفکیک پایین، پهنای قلم کم باشد، خطا حتی در حد یک پیکسل نیز می تواند باعث مشکلاتی در جداسازی شود. همچنین با توجه به اینکه در متن‌های قدیمی، نوار زمینه خط ثابت نیست و از کلمه به کلمه ممکن است جابه‌جا شود، به نوار زمینه کلی نمی توان اعتماد زیادی داشت. در این بخش روشهایی برای تعیین دقیقتر نوار زمینه به صورت محلی ارائه می‌شود.

٤-۱- تعیین نوار زمینه محلی با استفاده از کانتور زیر کلمات

توضیحات کامل مربوط به این روش در بخش سوم مرجع [۳] آمده است. به طور خلاصه می‌توان چنین گفت که برای هر زیر کلمه، نواری به پهنای 2pt در حوالی نوار زمینه کلی جستجو می‌شود. *برای هر ردیف تعداد تکرار نقاط کانتور بالایی با کد فریمن ٤ که از فاصله ذکر شده قرار داشته باشند شمرده شده و هیستوگرام این نقاط ،$h_4(n)$ محاسبه می‌شود که n متغیر مربوط به شماره ردیفهاست. ردیفی از تصویر که متناظر با حداکثر این هیستوگرام است به عنوان لبهٔ بالایی نوار زمینه*[۲] *،ubl* تعیین می‌شود. کد ٤ جهت از راست به چپ در راستای افق را نشان می‌دهد. به همین ترتیب با استفاده از هیستوگرام کد فریمن ۰، $h_0(n)$، لبهٔ پایینی نوار زمینه[۳]،lbl، مشخص می‌شود. شکل(۳) نمونه‌ای ازتعیین نوار زمینه محلی را نشان می‌دهد.

لازم به ذکر است که پهنای نوار جستجو در مرجع[۳]، 2pt در نظر گرفته شده است، ولی به دلیل اینکه در متون قدیمی



شکل۳ - تعیین نوار زمینه محلی برای بخشهای همپوشان

جابه‌جایی نوار زمینه کلمات بیشتر است، در این تحقیق پهنای نوار جستجو ۳pt در نظر گرفته شده است. اگر پهنای نوار زمینه محلی، بیش از ۲۵ درصد با ptg اختلاف داشته باشد، با روشی که در[۳] ارائه شده است، اصلاح می‌شود.

٤-۲- تعیین نوار زمینه محلی با استفاده از هیستو گرامهای $h_{24}(n)$,$h_{20}(n)$

در مواردی متعددی مشاهده می‌شود که نوار زمینه محلی به‌دست آمده با استفاده از هیستوگرام‌های $h_0(n)$ *و* $h_4(n)$ *به علت ناهموار بودن کانتور زیرکلمات در اثر نویز، به اشتباه محاسبه می‌شود. برای برطرف کردن این اشتباهات از هیستوگرام هموار شده* $h_0(n)$ *و* $h_4(n)$ *که* $h_{20}(n)$,$h_{24}(n)$ *نامیده شده‌اند، استفاده می‌شود، معادلات (۱ و ۲).*

$$h_{20}(n) = h_0(n) + h_0(n+1) \qquad (۱)$$

$$h_{24}(n) = h_4(n) + h_4(n+1) \qquad (۲)$$

ایده استفاده از هیستوگرامهای $h_{24}(n)$ و$h_{20}(n)$ برای تعیین نوار زمینه محلی، ازهمان ایده هیستو گرامهای $h_4(n)$ و$h_0(n)$ گرفته شده است، با این نگاه تازه که با استفاده از جمع ۲ تایی ردیفهای هیستو گرامهای $h_4(n)$ و$h_0(n)$ سعی شده است تا این هیستوگرامها هموار شده و اثر نویزهای احتمالی از بین برود. محل نوار زمینه در این روش با معادلات زیر تعیین می‌شود.

$$ubl_1 = n_1 \left| h_{24}(n_1) = \max_n \{h_{24}(n)\} \right. \qquad (۳)$$

$$ubl_2 = n_1 + 1 \qquad (٤)$$

$$ubl = \begin{cases} ubl_1 & \text{if} \quad h_4(n_1) \geq h_4(n_1+1) \\ ubl_2 & \text{if} \quad h_4(n_1) < h_4(n_1+1) \end{cases} \qquad (٥)$$

$$lbl_1 = n_2 \left| h_{20}(n_2) = \max_n \{h_{20}(n)\} \right. \qquad (٦)$$

$$lbl_2 = n_2 + 1 \qquad (٧)$$

$$lbl = \begin{cases} lbl_1 & \text{if} \quad h_0(n_2) \geq h_0(n_2+1) \\ lbl_2 & \text{if} \quad h_0(n_2) < h_0(n_2+1) \end{cases} \qquad (٨)$$

همان‌طور که از این معادلات پیداست، از بین دو ردیف مجاوری که ماکزیمم کدهای فریمن با شماره ٤ در آنها اتفاق افتاده است، ردیفی که تعداد این کدها درآن بیشتر است به عنوان ubl برگزیده می شود. به همین ترتیب از بین دو ردیف مجاوری که ماکزیمم کدهای فریمن با شماره صفر درآنها اتفاق افتاده است، ردیفی که تعداد این کدها درآن بیشتر است به عنوان lbl انتخاب می شود.

اگر پهنای نوار زمینه به‌دست آمده بیش از ۲۵ درصد با ptg اختلاف داشته باشد، نوار زمینه محلی مانند روش پیشنهادی در بخش (۳) از مرجع[۳] تعیین می شود.

٤-۳- تصحیح نوار زمینه محلی با استفاده ازماکزیممهای اول و دوم هیستو گرامهای $h_{20}(n)$, $h_{24}(n)$ و موقعیت نوار زمینه کلی

تعیین نوار زمینه براساس ماکزیمم هیستو گرامهای$h_{24}(n)$ و $h_{20}(n)$ درقسمت قبل توضیح داده شد. خطا در تعیین نوار زمینه با این روش بسیار کم است، اما به‌ندرت خطاهایی ایجاد می شود. در بررسی زیر کلماتی که در آنها خطا رخ داده بود، معلوم شد که در بیشتر آنها lbl وubl واقعی ماکزیممهای دوم هیستوگرامهای $h_{20}(n)$ و$h_{24}(n)$ هستند. این موضوع در مورد هیستوگرامهای $h_0(n)$ و$h_4(n)$ نیز اتفاق می افتد. برای رفع این

خطاهـا نـوار زمیـنه محلی با در نظر گرفتن ماکزیممهای اول و دوم هیسـتوگرامها تعیـین مـی‌شـود. اگـر ماکزیمم دوم از ۰/۸۵ ماکزیمم اول بزرگتر و به نوار زمینه کلی نزدیکتر باشد، موقعیت آن بـه عـنوان نـوار زمیـنه محلـی انـتخاب مـی‌شـود. در غیـر ایـن‌صورت موقعیت ماکزیمم اول هیستوگرام، تعیین کننده نوار زمینهٔ محلی است[۱۶]. شکل(٤) نمونه ای از تصحیح نوار زمینه محلی را با استفاده از این روش نشان می‌دهد.

۵- بر چسب زدن به نقاط کانتور بالایی

قوانین جداسازی بر اساس کانتور بالایی زیر کلمات بنا شده است. کانتور بالایی زیر کلمات از کانتور پیرامونی آنها استخراج مـی‌شـود[۳]. سـمت راسـت تـرین نقطـه کانتور بر چسب بالا می‌گیرد و نقاط بعدی کانتور طبق شکل(۵) برچسب می‌خورند. بـه عـنوان نمونه، *شرط عوض شدن برچسب از بالا به پایین که در شـکل(۵) بـا A_2 نمایش داده شده است، آورده می‌شود. اگر نقطه قبلی برچسب بالا داشته باشد و عرض نقطه فعلی بیشتر از ubl-pt/2 باشد و از عرض نقطه قبلی حداقل یکی بیشتر باشد و عرض نقطه بعدی از عرض نقطه قبلی حداقل دو تا بیشتر باشد یـا ایـنکه عـرض نقطه فعلی، برابر یا بزرگتر از ubl باشد، نقطه فعلی برچسب وسط می‌گیرد.*

الگوریتم بر چسب زدن به نقاط کانتور، با آنچه در مرجع[۳] آمـده، تفاوتهایـی دارد. چـون در ایـنجا هـدف پـیدا کـردن و جداسازی دندانه‌هاست وگم کردن دندانه‌ها در مواقعی است که ارتفـاع آنهـا کم باشد، از این رو حد آستانه فاصلهٔ مطلق تا نوار زمیـنه بـرای تـبدیل بـر چسـب وسط به بالا افزایش داده شده اسـت(pt/2 در شرط A1) و شرط محلی ونسبی صعودی بودن کانتور در نقطه مورد نظر اضافه شده است، شکل (۵). برای گم نکـردن دندانه هایی که کمی بالاتر از نوار زمینه نوشته شده اند، شرط تبدیل بر چسب بالا به وسط با کاهـش حد آسـتانه (pt/2 در شـرط A2) و اضـافه کردن شرط نزولی بودن کانتور اصلاح شـده اسـت شـکل (۵). بـه عبارت دیگر سعی شده است که با بهیـنه کردن شرایط تبدیل برچسب ها و نیز استفاده از اطلاعات محلی ونسبی صعودی یا نزولی بودن کانتور در هر نقطه، شرایط بهتـری برای پیدا کردن نقاط جداسازی فراهم شود.همچنین در این الگوریتم برای اینکه دندانه‌های کوتاه تا حد امکان تشخیص داده شـوند، تـبدیل بـر چسـب بـالا به وسط در ubl-pt/2 ولی تبدیل برچسب وسط به بالا در ubl انجام می‌شود، شکل (۵).

این موضوع را نیز باید در نظر داشت که در بعضی متون احتمال بالا رفتن یا پائین آمدن آرام کانتور در دندانه ها وجود دارد و ممکن است این قبیل دندانه ها از دید الگوریتم جا بیفتند. برای جلوگیری از چنین مواردی، شرایط مناسب به‌صورت OR، در A1 و A2 افزوده شده‌است، شکل (۵).

٦- برچسب زدن به پاره مسیرها و آشکارسازی نقاط جداسازی

نقـاط کنار هم در کانتور بالایی که بر چسب یکسانی دارند، پـاره مسـیر ی را ایجاد می‌کنند که برچسب آن نقاط را به خود می گیرد. طول هر پاره مسیر تعداد نقاط تشکیل دهنده آن است. برای برطرف کردن خطاهای احتمالی در بر چسب زدن نقاط به دلـیل وجـود نویـز در کانـتور بالایی، پاره مسیرهای کوتاه تر از pt/3+1 برچسـب پـاره‌مسـیر قبلی را به خود می گیرند. بعد از ادغـام یـک پاره مسیر کوتاه با پاره مسیر قبلی،اگردو پاره مسیر کـه دارای یـک برچسـب هسـتند به یکدیگر برسند، پاره مسیر بزرگتری با همان برچسب تشکیل می دهند که طول آن مجموع طولهای دو پاره مسیر قبلی است.

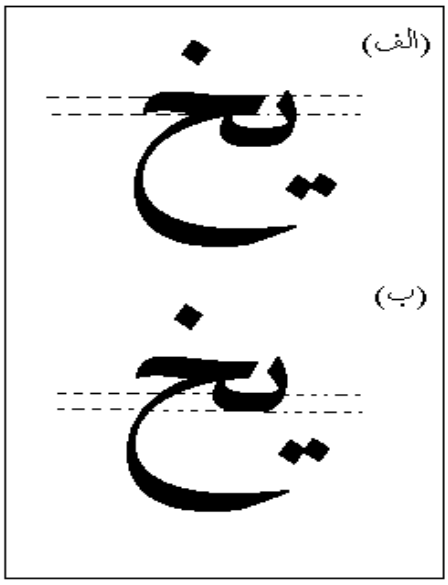
آشکارسازی نقاط جداسازی آخرین مرحله از الگوریتم ارائه شده است. قواعد جداسازی از مرجع[۳] اقتباس شده‌اند ولی با توجه به اهداف خاص این تحقیق تغییراتی در آنها داده شده است.

نقطهٔ جداسازی آخرین نقطه از یک پاره مسیر با برچسب وسط است که در یکی از معادلات زیر صدق کند.

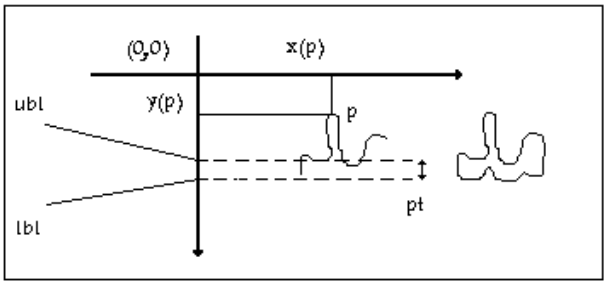
p: PU AND NUNE (۹)

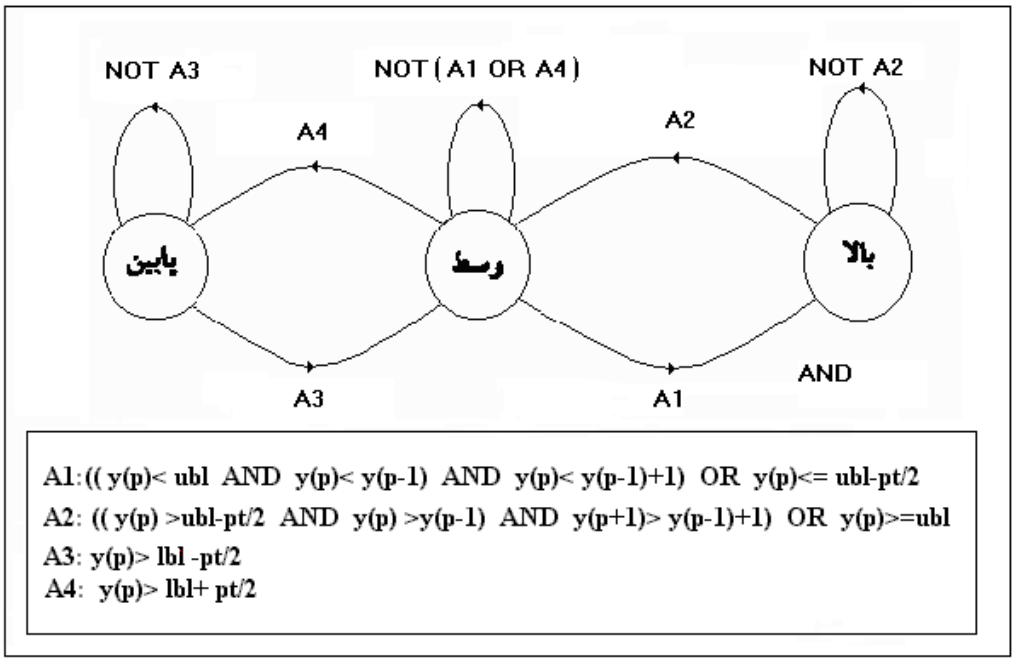
q: PU AND NUELHW (۱۰)

r: CL AND PU AND NDNEL (۱۱)



شکل ٤- تصحیح نوار زمینه محلی با استفاده از ماکزیمم دوم هیستوگرامهای $h_{20}(n)$ و $h_{24}(n)$
الف- نوار زمینه محلی اولیه ب- نوار زمینه محلی اصلاح شده





شکل ٥- نمودار حالت برای تعیین بر چسبهای کانتور بالایی زیر کلمه

s: CL AND PU AND NDEL (۱۲)

t: CL AND PD AND NU (۱۳)

که در آنها

NU,PD,NDEL,NDNEL,CL,NUELHW,NUNE,PU به ترتیب زیر تعریف می‌شوند.

PU: پاره مسیر قبلی برچسب بالا دارد.

NUNE: پاره مسیر بعدی برچسب بالا دارد و آخرین پاره مسیر کانتور بالایی نیست.

NUELHW: پاره مسیر بعدی برچسب بالا دارد ، آخرین پاره مسیر کانتور بالایی است و طول آن از ۲pt بیشتر است. این پاره مسیر یکی از دو شرط زیر را نیز دارد.

۱- ارتفاع آن، اختلاف بین عرض دو نقطه پاره مسیر که دارای بیشترین و کمترین عرض از مبدأ هستند، از۲ptبیشتراست.

۲- پهنای آن، اختلاف بین طول دو نقطه پاره مسیر که دارای بیشترین و کمترین طول از مبدأ هستند ، از ۲/۵pt بیشتر است.

CL:طول پاره مسیر فعلی از pt ۲ بیشتر است.

NDNEL: پاره مسیر بعدی برچسب پایین دارد ، آخرین پاره مسیر کانتور بالایی نیست و طول آن از ۳pt بیشتر است.

NDEL: پاره مسیر بعدی برچسب پایین دارد، آخرین پاره مسیر کانتور بالایی است و طول آن از ٤pt بیشتر است.

PD:پاره مسیر قبلی برچسب پایین دارد.

NU:پاره مسیر بعدی برچسب بالا دارد.

شرط NUELHW از بندq از شکستن شدن حروفی چون "ت، پ، ب، ث، د، ذ" در آخر زیر کلمه جلوگیری می‌کند. شکل(٦) پاره مسیرها و نقاط جداسازی یک کلمه را نشان می‌دهد. نمونه‌ای از کلمات جدا شده با استفاده از قواعد بالا در شکل(۷) آمده است .

برای جلوگیری از شکسته شدن حروف "ی، ن، ل" در هنگامی‌که به تنهایی یک زیر کلمه را تشکیل می‌دهند، از چند قاعده ساده استفاده می شود.

الف) نوار زمینه محلی این تک حرفها حداقل به اندازه pt/2 پائین تر از زمینه کلی قرار می گیرد.

ب) کانتور بالایی این تک حرفها از سه‌پاره مسیر بالا، وسط و بالا تشکیل می شود.

ج) طول پاره مسیر وسط این تک حرفها از ۲/۵pt بیشتر است.

د) ارتفاع دو پاره مسیر بالا از ۲pt بیشتر است.

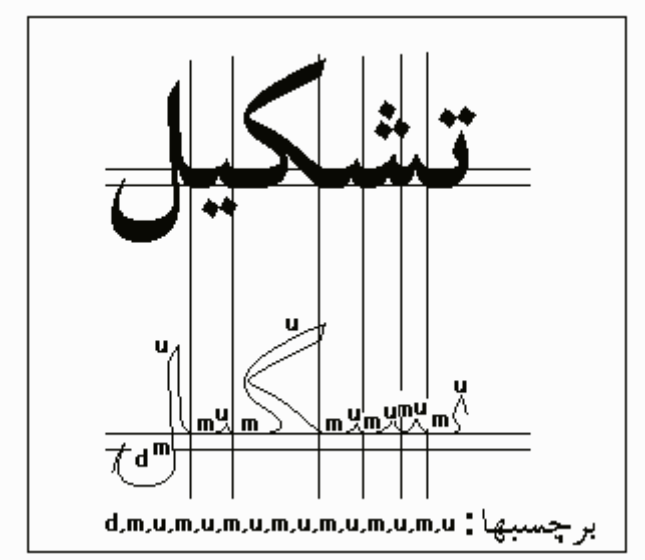
ه) پهنای پاره مسیر بالایی که بعد از پاره مسیر وسط قرار دارد، از ۲pt کمتر است.

۷- نتایج آزمایش، معرفی خطاها و بررسی علل آنها

مجموعهٔ تمرین از ٢٤ نوع قلم متفاوت، شامل ٢٤٢٨ حرف تشکیل می‌شود، جدول(۱) و مجموعهٔ آزمایش از ٦ متن، شامل ٦٤١٥ حرف انتخاب شده است[۱۷](متون شماره ۱ تا ٦ در جدول۱). یک متن برای آزمودن توانایی الگوریتم روی متون با کیفیت چاپ مناسب انتخاب شده است، جدول(۱). نتایج اجرای الگوریتم جداسازی حروف را نشان می دهد. در این جدول خطاهای ٦ گانه که در ادامه تعریف می‌شوند به دو گروه دسته بندی شده‌اند. خطاهایی که از آشکار نشدن نقاط جداسازی ناشی می شوند در گروه اول قرار دارند و خطاهایی که در اثر شکستن حروف در غیراز نقاط جداسازی ایجاد می‌شوند درگروه دوم قرار دارند. در جدول (۱) درصد خطاها با در نظر گرفتن زیر حروف در بدنه‌های "س" و "ص" محاسبه شده است، بخش (۲ مقاله).در ادامه انواع خطاها معرفی و بررسی می‌شوند.

خطای نوع ۱- آشکار نشدن دندانه اول بدنه "س"

یکی از شایعترین خطاهای جداسازی، آشکار نشدن دندانه اول در سین و شین است. این خطا ، از این جا ناشی می شود که در بعضی از قلم ها، مخصوصا درنگارشهای قدیمی، دندانه اول حروف سین و شین کمی بالاتر از سایر دندانه‌های آنها نوشته می شود و کمی بالاتر از نوار زمینه قرار می گیرد .بنابراین نقطه جداسازی آن آشکار نمی شود، شکل (۸).



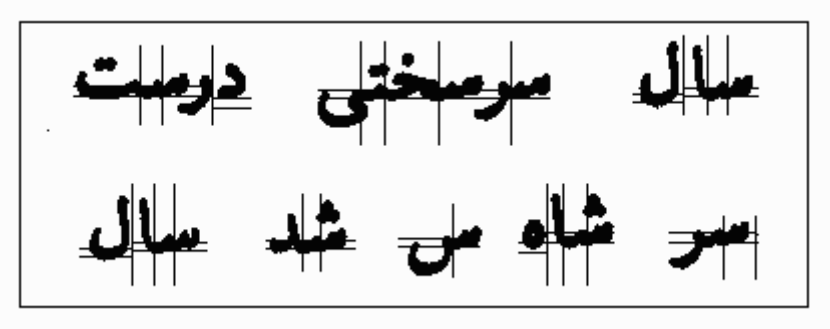
شکل٦- پاره مسیرهای بر چسب خورده و نقاط جداسازی بالا[1](u) ، وسط[2] (m) و پایین[3] (d)

جدول١ - مشخصات مجموعه‌های تمرین و آزمایش وتعداد خطاها بر حسب نوع و گروه آنها

خطای گروه اول: آشکارسازی نقاط جداسازی، خطای گروه دوم: شکستن حروف در غیر از نقاط جداسازی

	مجموعه آزمایش به تفکیک						مجموعه آزمایش	مجموعه تمرین	متن با کیفیت چاپ مناسب
	١	٢	٣	٤	٥	٦			
تعداد کل حروف	١٤٨٧	٦٧٥	٥٤٤	٧١٨	١٣٧٨	١٦١٣	٦٤١٥	٢٤٢٨	١٠٨١
تعدادحروف به هم چسبده	١٠٧٤	٥١٢	٤٤٣	٥٢٥	٩٧٩	١٢٥٧	٤٧٩٠	١٨٨٣	٨٥٦
حروف مجزا	٤١٣	١٦٣	١٠١	١٩٣	٣٩٨	٣٥٦	١٦٢٥	٥٤٥	٢٢٥
تعداد ص و ض	١١	٣	٣	٣	٣	٨	٣١	٢٢	٨
تعداد س و ش	٦٩	٢٤	٣٦	٤٣	٧٧	٧٠	٣١٩	١١٣	٥٩
تعداد نقاط جداسازی	١٢٢٣	٥٦٣	٥١٨	٦١٤	١١٢٦	١٤٠٥	٥٤٤٩	٢١٣١	٩٨٢
خطای نوع ١	٢٤	٧	–	٤	٥	٤	٤٤	١٤	–
خطای نوع ٢	٣	–	–	–	١	–	٤	–	–
خطای نوع ٣	٦	–	–	٥	–	–	١١	–	–
خطای نوع ٤	–	١	٢	–	٢	٩	١٤	٦	١
خطای نوع ٥	٢	١	٢	٥	٦	١٣	٢٩	١٤	٤
خطای نوع ٦	٦	٤	١٣	٧	٤	٢٩	٦٣	٦	–
خطای گروه اول	٣٦	١٢	١٤	١٧	١١	٤٠	١٣٣	٢٢	١
خطای گروه دوم	٢	١	٣	٥	٦	١٥	٣٢	١٨	٤
درصد خطای گروه اول	٣/١٨	٢/١٣	٢/٧	٢/٧٦	٠/٩٧	٢/٨٤	٢/٤٢	١/٠٣	٠/١
درصد خطای گروه دوم	٠/١٦	٠/١٧	٠/٥٨	٠/٨١	٠/٥٣	١/٠٦	٠/٥٨	٠/٨٤	٠/٤١

شکل ۷- نمونه هایی از جداسازی حروف با استفاده از الگوریتم ارائه شده



شکل۸- نمونه‌هایی از خطاهای نوع اول

خطای نوع۲- چسبیدگی ناخواسته حروف به یکدیگر

بعضی اوقات حروف علاوه بر آنکه روی نوار زمینه به یکدیگر می چسبند، در اثر پراکندگی جوهر در بالای نوار زمینه نیز با یکدیگر تماس پیدا می‌کنند. از آنجایی که نقطه جداسازی همواره در پایان یک پاره مسیر با بر چسب وسط قرار دارد، درمواردی که چسبیدگی حروف ، بالای نوار زمینه اتفاق می‌افتد، این قسمت از کانتور بر چسب بالا خواهد گرفت و نقطه جداسازی را می‌پوشاند، شکل (۹). البته این نوع خطا درمتون با کیفیت پایین اتفاق می‌افتد.

خطای نوع ۳- اشتباه در تشخیص نوار زمینه

بعضی از نقاط جداسازی بر اثر اشتباه در به‌دست آوردن نوار زمینه، آشکار نمی شوند. این نوع خطا، ممکن است در اثر عوامل مختلفی بروز کند. بعضی از این خطاها در زیرکلمات همپوشان روی می دهد. بدین ترتیب که زیر کلمات همپوشان، نسبت به یکدیگر جابه‌جایی کرسی دارند ولی در الگوریتم ارائه شده ، برای این زیر کلمات، تنها یک نوار زمینه به‌دست می آید. این موضوع باعث می شود که بعضی از نقاط جداسازی آشکار نشوند. به نظر می آید که راه گریز از این خطا، به‌دست آوردن

شکل ۹- نمونه‌هایی از خطاهای نوع دوم

نوار زمینه محلی برا ی هر زیر کلمه به‌صورت مجزا باشد. اما محدود کردن نوار زمینه محلی به زیر کلمات، می تواند خطاهای زیادتری در تولید نقاط جداسازی اضافی، ایجاد کند، شکل (۱۰، کلمات "جوش" و "سرمست").

یکی دیگر از عوامل خطا در تعیین نوار زمینه این است که نقاط جداسازی حروف در بعضی از زیر کلمات روی نوار زمینه‌ای که بر اساس ماکزیمم هیستوگرامهای h_{20} ,h_{24} به‌دست می آید، قرار نمی‌گیرند(بخش ٤-۳)،(شکل ۱۰، کلمه "خود").

خطای نوع٤- ضعف قواعد جداسازی

بعضی از نقاط جداسازی بر اثر ضعف الگوریتم جداسازی آشکار نمی شوند. این خطاها در اثر کم یا زیاد بودن بعضی از آستانه‌های به‌کار رفته، ایجاد می شوند. اما باید به این نکته توجه داشت که حدود آستانه با استفاده از یک مجموعه متنوع و روش سعی و خطا انتخاب شده اند و تا حد امکان، سعی شده است که کمترین خطا به وجود آید. برطرف کردن خطاهایی که از این دسته هستند، برای یک قلم خاص با تنظیم حدود آستانه برای آن قلم ممکن است، اما باعث می شود که خطاهای دیگری در جداسازی حروف قلمهای دیگر ایجاد شود، شکل(۱۱).

خطاهای نوع ٥- شکستن بعضی از حروف به زیر حروف

یکی دیگر از انواع خطاها که در جداسازی حروف اتفاق می افتد. شکستن ناخواسته حروف"ی، ن، ل، د ،ذ، ب، ت،پ،ث، ع، غ" به زیر حروف است، شکل(۱۲). خوشبختانه این نوع خطاها، به راحتی در مرحله پس‌پردازش قابل شناسایی و تصحیح هستند[۳،۱۲].

خطای نوع ٦- پایین بودن کیفیت چاپ

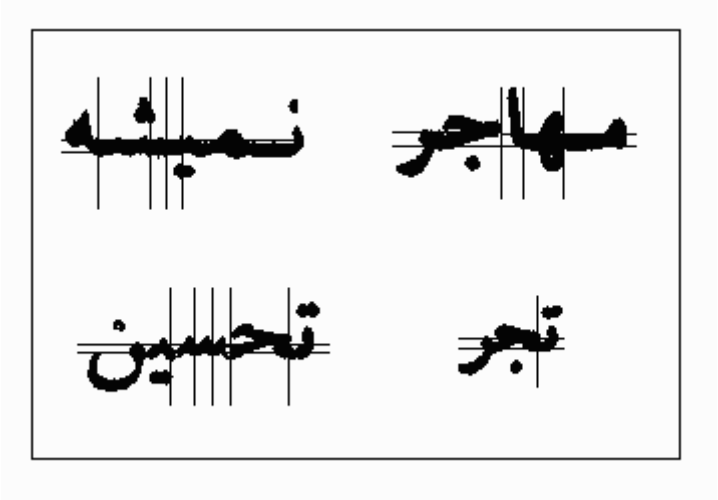
بعضی از خطاها در اثر کیفیت خیلی پایین چاپ ایجاد می‌شوند. درمجموعه آزمایش ما، این نوع خطا بیشترین سهم را به خود اختصاص داده است. دراین موارد کیفیت متن به گونه‌ای است که اجتناب از خطا مشکل به نظر می‌رسد. خطاهایی که در این دسته قرار دارند، ناشی از عوامل مختلفی‌اند که به بررسی آنها می پردازیم، شکل (۱۳).

الف) دندانه های کوتاه و نرم: از مهمترین عوامل تولید خطا دراین دسته، دندانه های کوتاه و نرم‌اند. الگوریتم جداسازی حروف این توانایی را دارد که دندانه های کوتاه را آشکار کند. اما دندانه‌ها در بعضی از متون به حدی کوتاه‌اند که آشکارسازی آنها بسیار مشکل است. اگر بخواهیم پارامترهای الگوریتم را به نحوی تنظیم کنیم که این دندانه‌ها را نیز تشخیص دهد، آن گاه هر نویز پیش آمده روی نوار زمینه را به عنوان دندانه آشکار می‌کند و نرخ تولید خطا افزایش می یابد. دندانه های نرم نیز از عوامل تولید خطا هستند. این دندانه ها از دندانه‌های کوتاه، کمی بلند ترند، اما به حدی آرام بالا می‌روند که الگوریتم را برای تبدیل برچسب کانتور بالایی از وسط به بالا تحریک نمی‌کنند. بنابراین بعضی از این دندانه ها نیز آشکار نمی‌شوند، شکل (۱۳-الف).

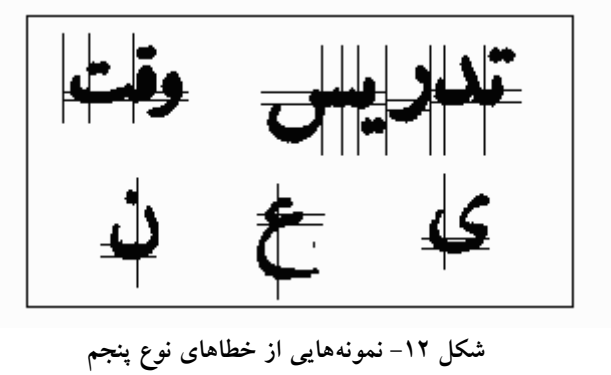
ب) نویزهای موجود در لبهٔ بالایی نوار زمینه: از آنجا که متون با دندانه های کوتاه فراوان‌اند، برای آشکار سازی این دندانه‌ها، پارامترهای الگوریتم را تا حد امکان حساس کرده‌ایم. این حساسیت باعث شده است که بعضی از نویزهای شدید در لبه بالایی نوار زمینه به عنوان دندانه آشکار شوند. به نظر می‌رسد این‌خطاها نیز در قسمت پس پردازش قابل تصحیح‌اند، شکل (۱۳-ب).

ج) قرار نگرفتن حروف زیر کلمه در یک راستا: در بعضی ازمتون چاپی حروف یک زیر کلمه در یک راستای افقی قرار نمی‌گیرند. بعضی از حروف بالاتر و بعضی دیگر پایینتر نوشته می شوند. از آنجایی که الگوریتم ما برای کل زیر کلمه یک نوار

شکل ۱۰- نمونه‌هایی از خطاهای نوع سوم



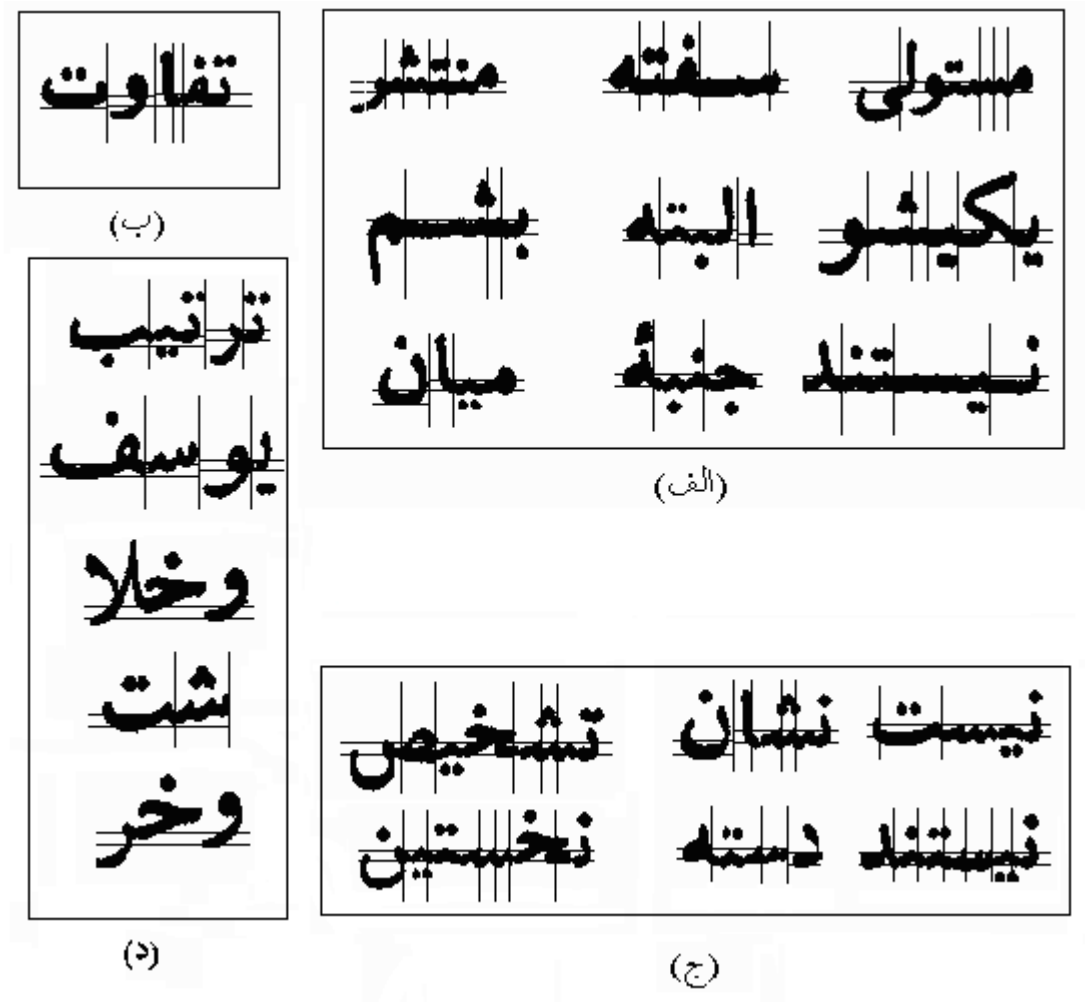
شکل ۱۱- نمونه‌هایی از خطاهای نوع چهارم



شکل ۱۲- نمونه‌هایی از خطاهای نوع پنجم

زمینه پیدا می کند، بعضی از نقاط جداسازی آشکار نمی‌شوند، شکل (۱۳-ج)

د) شیبدار نوشته شدن حروف در یک زیر کلمه: یکی دیگر از عوامل تولید خطا درآشکار سازی نقاط جداسازی، شیبدار نوشته شدن حروف یک زیر کلمه است، شکل(۱۳-د). الگوریتم ارائه شده، نوار زمینه را با این تعریف تعیین می‌کند که حداکثر هیستوگرامهای h_{20} و h_{24} دراین نوار اتفاق می افتد، بخش (٤-٤). اما در بعضی از نگارشها، حروف به حدی زاویه دارند که تعیین یک نوار زمینه افقی برای آنها ممکن نیست. در نهایت در این قبیل کلمات نوار زمینه در هرجا که قرار داده شود، چند نقطه جداسازی آشکار نمی شود. برای حل مشکل این قبیل خطاها باید یک نوار زمینه شیبدار تعریف کرد. در



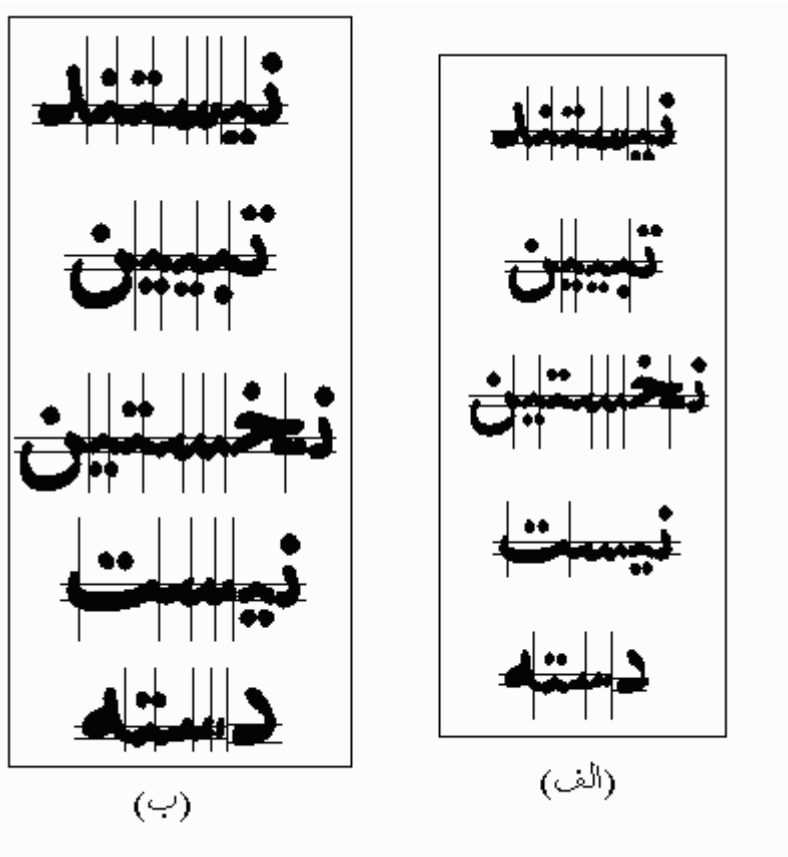
شکل۱۳- الف- مثالهایی از خطاهای نوع ٦- الف
ب- مثالهایی از خطاهای نوع ٦- ب
ج- نمونه‌هایی از خطاهای نوع ٦- ج
د- نمونه‌هایی از خطاهای نوع ٦- د

ضمن قابل ذکر است که سیستم موجود نسبت به کجی خط ناشی از تصویر برداری حساس است. چنانچه اختلاف بین بالاترین و پایینترین نوار زمینه محلی در یک خط متن کمتر از دو برابر پهنای قلم باشد، الگوریتم می‌تواند آن را پردازش کند. معمولاً برای کجی بیشتر، از روشهای رفع کجی متن استفاده می‌شود.

خطاهـای نوع دوم و ششم که مجموعا ٤٠/٦٢٪ از کل خطا را تشـکیل مـی‌دهـند، در اثـر کیفـیت خیلی پایین چاپ پدیدار شـده‌انـد. خطاهای نوع پنجم که ١٧/٥٧٪ ازکل خطا را تشکیل مـی دهـند بـا پس پردازشهای مناسب قابل تصحیح‌اند.خطاهای نـوع اول، سـوم و چهـارم کـه مجموعا ٤١/٨١٪ از کل خطا را تشـکیل می‌دهند، به دلیل ضعف قواعدجداسازی و نحوه تعیین نوار زمینه به وجود آمده‌اند، جدول(٢).

٨- اثر افزایش درجه تفکیک

برای تعیین اثر درجه تفکیک، از متن شماره ٦ با درجه تفکیک ٤٠٠ dpi نیزتصویر برداری شد والگوریتم جداسازی روی تصویر جدید اعمال شد. در مقایسه با نتایج به‌دست آمده برای ٣٠٠dpi (ستون مربوط به متن شماره ٦ در جدول ١)، تعداد خطاهای گروه اول از ٤٠ به ٢٣ و تعداد خطاهای گروه دوم از ١٥ به ١٠ کاهش یافت. به عبارت دیگر حدود نیمی از خطاهای مربوط به آشکار نشدن نقاط جداسازی برطرف شدند،



شکل ١٤- الف) برخی از خطاهای جداسازی در ٣٠٠dpi
ب)تصحیح شدن بعضی از این خطاها در ٤٠٠dpi

جدول٢- درصد خطاهای٦ گانه نسبت به خطای کل در مجموعه آزمایش (متون ١تا٦)

نوع خطا	١	٢	٣	٤	٥	٦
درصدخطا به کل	٢٦/٦٦	٢/٤٢	٦/٦٧	٨/٤٨	١٧/٥٧	٣٨/٢٠

در حالی‌که حدود یک سوم از خطاهای مربوط به شکستن ناخواسته حروف تصحیح شدند، شکل (١٤).

٩- نتیجه‌گیری

هدف از این تحقیق اصلاح الگوریتم جداسازی حروف در مراجع[٣،١٥] برای متون با کیفیت چاپی پایین بود. در بخش چهارم، مراحل مختلف اصلاح الگوریتم آشکارسازی نوار زمینه محلی مطرح شده است. در بخش پنجم مقاله تغییرات انجام شده در برچسب‌زنی نقاط کانتور بالایی نسبت به الگوریتم[٣] ارائه شده است. در این قسمت سطوح آستانه در تبدیل بین برچسبهای مختلف تغییر داده شده اند و با اضافه کردن مفهوم شیب کانتور ،جداسازی دندانه‌های کوچک و نرم با دقت بیشتری صورت گرفته است .در بخش ششم، قوانین جدا سازی برای کاربرد در متون با کیفیت چاپی پایین تکمیل شده‌اند. در بخش هفتم علل بروز انواع مختلف خطا به طور جامع بررسی شده است.

میـزان جداسازی درست در مرجع [٣] قبل از پس پردازش ٩١٪ و بعـد از آن ٩٨/٥٪ اسـت. بـرای الگوریـتم اصلاح شده،

میزان جداسازی درست بدون پس‌پردازش ۹۷٪ است. البته لازم بـه ذکر است که مجموعه های آزمایش در این دو تحقیق کاملا یکسان نیستند و تعریف خطا نیز با توجه به اینکه در این تحقیق جداسازی دندانه‌های "س" و "ص" مورد نظر بوده است، اندکی مـتفاوت اسـت .همچنـین بایـد تـوجه داشت که در این تحقیق سـعی شده است با جامعتر کردن قواعد جداسازی حتی‌الامکان از پس‌پردازشهای مطرح شده در مرجع[۳] اجتناب شود.

واژه‌نامه

1. binary dicision tree
2. upper base line(ubl)
3. lower base line(lbl)
4. upper
5. middle
6. down

مراجع

1. Lu,Y., "Machine Printed Character Segmentation-an Overview," *Pattern Recognition*,vol.28,no.1, pp. 67-80, 1995.
2. 2-Lu,Y. and M. Shridhar, "Character Segmentation in Handwritten Words – an Overview," *pattern Recognition*, vol. 29, no.1 , pp. 77-96,1996.

۳. عزمی، ر. و کبیر، ا.، "معرفی روش جدیدی برای جداسازی حروف در متون چاپی بدون توجه به نوع قلم،" مجله استقلال، سال ۱۸،شماره ۲، صفحات ۱۰-۱، اسفند ۱۳۷۸.

٤. احمد زاده، م. و کبیر،ا.، "شکستن کلمات تایپ شده فارسی بـه حروف،" گزارش اولین کنفرانس بین المللی کامپیوتر در علـوم، فـنون و پزشـکی ایـران، صـفحات٦-۱، ۷-۵ دیماه ۱۳۷۰، اصفهان.

5. Parhami, B. , and Taraghi, M., "Automatic Recognition of Printed Farsi Text," *Pattern Recognition*, vol. 14, pp. 395 – 403, 1981.
6. EL-Sheikh, T.S., and Guindi, R.M. "Computer Recotnition of Arabic Cursive Scripts," *Pattern Recognition*, vol. 24 , no .4, pp. 293 –302, 1988.

۷- شیرعلی شهررضا، م. ح، "تشخیص کلمات و ارقام دستویس فارسـی به وسیله شبکه های عصبی،"رساله دکتری، دانشکده برق، دانشگاه صنعتی امیر کبیر، ۱۳۷٤.

8. Amin, A., and Masini, "Machine Recognition of Multi Font Printed Arabic Texts,", *Proc. 8th Int. Conf . on Patt. Recog*. Paris, pp.392-395 , 1986 .
9. Amin, A., and Mari,J.F., "Machine Recognition and Correction of Printed Arabic Text," *IEEE Trans. Systems, Man and Cybernetic,* vol. 9, pp 1300 – 1306,1989.
10. Amin, A., and Al- Fedaghi, S., "Machine Recognition of Printed Arabic Text Utilizing a Natural Language Morphology," *International Journal of Man _ Machine Studiese*, vol. 35 , pp. 759 –768 , 1991.

۱۱. یکتایی،م.، زحزح، ا. و منارد، م.، "جداسازی کلمات فارسی به حروف وزیر–حروف،" دومین کنفرانس سالانه انجمن کامپیوتر ایران، دانشگاه صنعتی امیر کبیر، صفحات ۲۱۱–۲۱٦، ٤ الی ٦ دیماه ۱۳۷۵.

۱۲. رفیعـی، ش. و کبیر، ا.، "شکستن کلمات تایپ شده فارسی بـه حـروف در رسـم الخطهـای مختلف،" مجموعه مقالات سـومین کنفـرانس الکتـرونیک، دانشـگاه شـیراز، صـفحات ۱۰٤–۹۸، مهرماه ۱۳۷٤.

13. Kurdy, B. M., and Joukhadr,A., "Multifont Recognition System for Arabic Characters," *proc . 3rd Int.Conf. Exhibition of Multi- Lingual Computing* (Arabic and Roman script), U.K, pp. 731 –739,1992.
14. Amin, A., and Al-Sadoun, H.B., "A New Segmentation Technique of Arabic Texts," *Proc. 11th IAPR Int. Conf. on Patt. Recog*. Vol.II. Conf.B: Pattern Recognition Methodology and system, pp.441-5,1992.
15. Azmi, R. and Kabir, E., "A New Segmentation Technique for Omnifont Farsi Text," *Pattern Recognition Letters*, vol. 22, no2, pp. 97-104, 2001.

۱٦.نظـام‌آبادی‌پور،ح.، کبیر، ا. و عزمی، ر.، "جداسازی حروف در متون چاپی قدیمی،" مجموعه مقالات ششمین کنفرانس بـین‌المللـی سـالانه انجمن کامپیوتر ایران،دانشگاه اصفهان، صفحات ۵۹۰– ۵۷۹، اسفندماه ۱۳۷۹.

۱۷.نظـام‌آبـادی‌پـور،ح. "پیش پردازش متون چاپی فارسی برای جداسـازی حـروف،" پایـان نامـه کارشناسـی ارشد، بخش مهندسی برق، دانشگاه تربیت مدرس، ۱۳۷۹.

[DOR: 20.1001.1.22287698.1383.23.1.3.9]