

باینری کردن نوشتار در تصاویر پیچیده رنگی به روش sliding window

داود زاروی^۱، سعید زاروی^۲

^۱باشگاه پژوهشگران جوان، دانشگاه آزاد اسلامی واحد بوشهر، davodzarsa@gmail.com

^۲دانشگاه آزاد اسلامی واحد دزفول، saeedzaravi@gmail.com

چکیده - فرایند باینری کردن پیکسلها را بطور جداگانه به عنوان متن یا پس زمینه تصاویر دسته بندی می کند. بکارگیری این فرایند برای پر کردن فاصله مرحله های مکان یابی و تشخیص نوشتار به کمک OCR نیاز است. این مقاله شیوه پنجره لغزنده sliding window را برای باینری کردن متن از تصاویر با پس زمینه پیچیده معرفی می کند. در ابتدا تکنیکهای پیش پردازش مناسبی برای افزایش کنتراست تصویر و مات کردن نویزهای پس زمینه بکار می رود، سپس لبه ها توسط آستانه گیری تکراری تشخیص داده می شوند و در ادامه کادریابی که برای تشخیص لبه های نوشته شناسایی شده اند تحلیل می شوند تا لبه های ناخواسته ای که منجر به پیچیدگی تصویر می شوند حذف شده و با الگوریتمی که یکنواخت بودن اندازه کاراکتر را بررسی میکند و بر پایه روش پنجره لغزنده است باینری شوند. شیوه پیشنهاد شده روی نواحی مکان یابی شده ای از تصویرهای بافتی ناهمگن اعمال شده است و با روشهای Otsu و Niblack مقایسه شده و عملکرد موثر شیوه پیشنهادی را نشان می دهد. کلید واژه- باینری کردن، پنجره لغزنده (sliding window)، آستانه گیری تکراری، تجزیه و جداسازی کاراکترها (حرف ها).

می شوند و برای تصاویر پیچیده و تصاویر اسناد و مدارک فرسوده مناسب نمی باشند. روش های سراسری تند هستند و برای متن- های کوچک و ساده مناسب بنظر می رسند. از سوی دیگر، روش های باینری محلی، به طور سازگاری آستانه را مطابق با خصوصیات و ویژگی های محلی تغییر می دهند این شیوه ها برای غلبه بر نقص ها و کاستی های باینری های سراسری پیشنهاد شده اند و میتوانند با محاسبه آستانه های موضعی در محدوده پنجره های جداگانه و یا نواحی بخش بندی شده [۵] [۱۱]، اصلاح شوند.

در روش پنجره های لغزنده، اندازه و شکل پنجره پارامترهای از پیش تعیین شده ای هستند. ضعف در باینری زمانی حاصل می شود که مرزهای پنجره ها از کاراکترها عبور کرده و احتمالاً در تصویر سبب ایجاد کاراکترهای اشتباه شوند و این خود موجب پیامدهای ناخوشایندی در تصویر باینری خواهد شد.

موضوع دیگر در ارتباط با استخراج متن تصویر وجود پس- زمینه پیچیده است. در این مقاله نظریه باینری کردن متن براساس شیوه پنجره لغزشی که برای مدارک رنگی با پس زمینه پیچیده ی بافتی پیشنهاد شده است، روشی برای چیره شدن به این مشکل ارائه می کند.

در ادامه: بخش (۲) مربوط به کارهای مشابه پیشین است. بخش (۳) به معرفی مسئله می پردازد و مدل پیشنهادی ما را معرفی می کند. بخش (۴) بطور ویژه به باینری کردن اختصاص

۱- مقدمه

نوشته ها در تصویرهای دارای متن مانند تابلوهای تجاری و بازرگانی، علامت های ترافیک و بیل بوردهای تبلیغاتی، در بسیاری از زمینه ها همچون سیستم های کمک رسانی برای افراد آسیب دیده، دقت راننده نسبت به علائم ترافیکی و سیستم ترجمه برای افراد خارجی بسیار کارا بوده و در کاربردهای بالقوه- ای مانند تشخیص و شناسایی برگه های مجوز، گرفتن یادداشت های دیجیتالی، بایگانی نمودن اسناد و محاسبات خسته کننده بسیار مفید می باشد.

فرایند باینری، کردن یک یک پیکسلها را جداگانه به دو دسته روزمینه و یا پس زمینه دسته بندی می کند. برای پرکردن شکاف میان مرحله بخش بندی و تشخیص و شناسایی توسط OCR، باینری کردن لازم و ضروری می باشد. خروجی این مرحله تصویری باینری است، از آنجایی که کاراکترهای سیاه متن بر روی یک پس زمینه سفید، ظاهر می شوند تکنیک های رایج به دو گروه طبقه بندی شده اند: باینری کردن سراسری و باینری کردن محلی یا وقتی.

در روش های باینری کردن سراسری [۱۲]، آستانه های سراسری برای تمام پیکسل های صفحه در عکس به کار برده

کادر لبه‌ها، پاک کردن لبه‌های نادرست و الگوریتم پنجره لغزنده (sliding windows) که در شکل (۱) نشان داده شده است.

در درجه اول پیش‌پردازش‌های مناسبی برای افزایش کانتراست تصویر و تار کردن نویزهای ناشی از پس‌زمینه به کار می‌رود، سپس برای لبه‌هایی که با آستانه‌گیری تکراری شناسایی شده‌اند کادر ایجاد می‌گردد. بعد از آن کادرهای اشتباه حذف شده و تصویر به وسیله چک کردن یکسانی اندازه کاراکترها باینری می‌شود.

۳-۱- پیش پردازش

هدف اصلی مرحله پیش پردازش شفاف‌تر کردن اجسام روزمینه نسبت به پس‌زمینه است تا اینکه برای یافتن آشکارتر لبه‌های متنی کمک بیشتری بشود. در این مرحله تصویر رنگی به تصویری با سطوح خاکستری تبدیل شده و با تقویت کانتراست براساس محاسبه آنتروپی، هموارسازی و بسط سطوح خاکستری پی‌گیری می‌شود [۱۷].

۳-۲- بهبود کانتراست

برای تصاویری که در شرایط ضعیف نوری گرفته شده‌اند ممکن است به آنتروپی ضعیف بینجامد که برای پردازش بهتر نیاز به افزایش کانتراست دارند. بنابراین اگر افزایش در کانتراست تصویر لازم باشد، می‌توان آنتروپی را به عنوان یک ویژگی به کاربرد. چنانچه آنتروپی محاسبه شده تصویر بسیار پائین باشد، بهتر است کانتراست افزایش یابد و در غیر این صورت ممکن است لبه‌های تعیین شده کاراکترها پیوسته نشوند. آنتروپی را می‌توان به شکل زیر محاسبه کرد:

$$H = -\sum p_i \times \log p_i = \sum p_i \times \log \frac{1}{p_i} \quad (1)$$

H نشان دهنده آنتروپی تصویر و p_i بیانگر نسبت مقدار سطح خاکستری $i \in [0.255]$ در همه تصویر است. بنا به آزمایش‌های گسترده انجام یافته، مقدار تجربی $H_{thres} = 38$ برای شناسایی تصاویری توصیه می‌شود که کانتراست آنها نیاز به تقویت دارد.

برای این آستانه‌گیری، کانتراست تصویر نیاز به افزایش دارد:

$$c(x, y) = \frac{255}{1 + \exp\left[\frac{avg_T - T(x, y)}{v}\right]} \quad (2)$$

دارد. در بخش (۵) ضمن بحث درباره نتیجه کار به ارائه پیشنهاد برای کارهای بعدی می‌پردازیم.

۲- کارهای مشابه پیشین

شیوهی باینری کردن دو گانه در [۱] ارائه شده است که می‌تواند به آسانی نوشته‌ها را با پلاریته‌های رنگی متفاوت از پس زمینه جداسازی کند. در [۱۰] شیوه باینری کردن و پاک کردن پیکسل‌های پس‌زمینه از لابلای کاراکترها ارائه شده است. شیوه باینری کردن بر پایه یادگیری برای اینگونه مدارک در [۳] معرفی شده. در این مقاله از پهنای ضربه برای ارزیابی باینری کردن استفاده می‌شود. در [۴] مدلی براساس میدان تصادفی مارکف (MRF) برای اسناد ارائه شده است. پارامترهای مدل را می‌توان از طریق آموزش فراگرفت، تصویر باینری نیز در این چهارچوب تخمین زده می‌شوند.

در [۵]، جداسازی نوشتار در ناحیه شناسایی شده دارای متن و با تبدیل همه فضای رنگی به دو مجموعه رنگی قابل تشخیص از هم انجام گرفت. آستانه‌گیری محلی افقی براساس طرح کاوش در آستانه‌های مختلف در [۶] ارائه شده است. [۷] روش بخش بندی متن براساس دسته‌بندی طیفی و هیستوگرام شدت را پیشنهاد داد.

اگر چه، این روشها عمدتاً روی تصویرهایی کار می‌کنند که پس‌زمینه ساده و یکنواختی داشته باشند. برخی کارها [۱۳] و [۸] نیز به پس‌زمینه‌های پیچیده پرداخته‌اند و باینری کردن را روی هر کاراکتر [۹] و هر کلمه [۸] اعمال می‌کنند. اما در این کارها فرض می‌شود که رنگ متن یکسان بوده و روشها برای مدارک رنگی با متن چند رنگ مناسب نیستند.

[۲] با مدارک متن چند رنگی سروکار دارد و بدون توجه به پلاریته سایه‌های پس‌زمینه و روزمینه با تجزیه و تحلیل لبه‌ها به نوشته‌ها پرداخته است. هر چند، اگر پس‌زمینه دارای بافتی پیچیده باشد، اجزای لبه ممکن است به خاطر حاشیه‌های پس-زمینه به طور صحیح قابل شناسایی و تشخیص نباشند. بنابراین، با توجه به موارد بالا روش پنجره لغزنده (sliding window) پیشنهاد می‌شود که براساس اندازه کاراکتر کار می‌کند و الگوریتمی برای کم کردن پیچیدگی پس‌زمینه بدست می‌دهد.

۳- توصیف سیستم

شیوه باینری کردن پیشنهادی شامل مراحل زیر است:
پیش‌پردازش، آستانه‌گیری تکراری برای تعیین لبه، یافتن

تصویر ورودی و تصویر نمایش شده پس از این مراحل در شکل ۲.۲ تا ۲.۵ نشان داده شده است.

۴- آستانه‌گیری تکراری و لبه‌یابی

الگوریتم‌های لبه‌یابی معمولی با فیلترهای بالا گذر کار می‌کنند، که برای تصاویر پس زمینه پیچیده مناسب نیستند. در چنین مواردی، بهتر است از تصویر خاکستری استفاده کرد که در آن تصویر بصورت تابعی مکانی است و با تغییر موقعیت به آهستگی تغییر می‌کند [۱۴].

علاوه بر این، پیوستگی لبه‌ها نیز حفظ شده و از هم پوشانی آن‌ها جلوگیری می‌شود. به این منظور، الگوریتم آستانه‌گیری تکراری و فرسایش هندسی [۱۵] برای لبه‌یابی به کار می‌روند. مراحل لبه‌یابی به روش آستانه‌گیری:

آستانه اولیه: $T^0 = \{T^k | k = 0\}$ که در آن از می‌نیم و ماکزیمم مقدار خاکستری تصویر محاسبه می‌شود:

$$T^0 = \frac{Z_{\min} + Z_{\max}}{2} \quad (5)$$

تصویر به وسیله آستانه T^k به دو بخش (R_1, R_2) تقسیم می‌شود.

$$R_1 = \{f(x, y) | f(x, y) \geq T^k\} \quad (6) \text{ و } (7)$$

$$R_2 = \{f(x, y) | 0 < f(x, y) < T^k\}$$

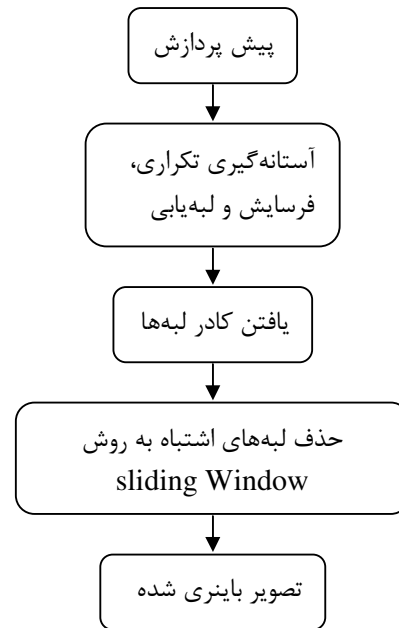
مقدار سطح خاکستری میانگین در R_1, R_2 محاسبه شده و به طور جداگانه به شکل z_1, z_2 نمایش می‌شوند.

$$z_1 = \frac{\sum_{f(i,j) < T^k} f(i, j) \times N(i, j)}{\sum_{f(i,j) < T^k} N(i, j)} \quad (8)$$

$$z_2 = \frac{\sum_{f(i,j) < T^k} f(i, j) \times N(i, j)}{\sum_{f(i,j) < T^k} N(i, j)} \quad (9)$$

که $f(i, j)$ مقدار سطح خاکستری نقطه (i, j) در تصویر بوده، $N(i, j)$ که معمولاً برابر با ۱ است ضریب توان نقطه (i, j) است. آستانه جدید T^{k+1} برابر متوسط z_2, z_1 می‌شود. اگر $T^k = T^{k+1}$ باشد الگوریتم پایان یافته می‌شود و گر نه اگر $k=k+1$ قرار داده و به مرحله (۲) برگردید.

سپس، فرسایش هندسی می‌دهیم تا اطمینان بیابیم لبه‌های



شکل ۱: معماری سیستم پیشنهاد شده

$C(x, y)$ نشان دهنده نتایج تبدیل است. $avgT$ ، متوسط مقادیر خاکستری در تصویر T است. $v=15$ می‌گیریم.

۳-۳ هموارسازی

مرحله هموارسازی مرحله‌ای ساده‌ای برای به حداقل رساندن اختلالات مرحله شناسایی لبه است. گسترش بازه خاکستری تصویر هم جهت افزایش کنتراست بیشتر انجام می‌گیرد.

$$S(x, y) = \frac{1}{10} (c(x, y) \oplus M) \quad (3)$$

که در آن $c(x, y)$ و $s(x, y)$ مقدار خاکستری پیکسل (x, y) تصویر را قبل و بعد از تبدیل نشان می‌دهد. M هم نشان دهنده ماسک بکار رفته است [۱۷].

۳-۴ گسترش بازه خاکستری

اگر $s_{\max} - s_{\min} < 80$ ، کنتراست تصویر خاکستری پایین است. گسترش بازه خاکستری در بازه کامل $\max - \min = 255$ کنتراست را می‌افزاید. معادله تبدیل به صورت زیر است:

$$G(x, y) = s(x, y) + \alpha \times \beta \quad (4)$$

که در آن S ، تصویر خاکستری بعد از هموارسازی و G تصویر بازتر شده را نشان می‌دهد.

$$\alpha = -\min_s, \quad \beta = \frac{255}{\max_s - \min_s}$$

{Reject EB_{int}, Accept EB_{out}}
else
{Reject EB_{out}, Accept EB_{int}}

۴-۲- باینری کردن و زدودن کادرهای نادرست لبه براساس روش sliding window

در ناحیه‌ای که نوشته‌ها مکان‌یابی شده‌اند و کاراکترها تقریباً دارای اندازه یکسان هستند، از این ویژگی برای باینری کردن و زدودن کادرهای نادرست لبه براساس روش sliding window استفاده می‌شود.
مراحل مرتبط با الگوریتم عبارت از:
اندازه کادرهای لبه‌های مختلف محاسبه شده و در آرایه A ذخیره می‌شود.

پنجره اولیه روی درایه‌های آرایه با در نظرگیری A[i] به عنوان نشانگر سمت چپ (LP) و A[i+1] به عنوان نشانگر سمت راست (RP) فرض می‌شود.

اگر اختلاف بین LP و RP بیش از آستانه باشد، فقط RP را در نظر بگیرد و آن را به عنوان LP بشناسید و اندازه پنجره را برای پوشاندن عنصر بعدی بلغزانید و تفاوت بین این دو عنصر را بررسی کنید.

آخرین آرایه سمت چپ را با دیگر آرایه‌های باقیمانده یک به یک مقایسه کنید و اگر تفاوت کمتر از آستانه بود پنجره را با اضافه کردن آن آرایه‌ها به درون پنجره گسترش دهید و هنگامی که این شرایط برقرار نبود پنجره را ببندید. این پنجره در بردارنده کاراکترهای با اندازه یکسان است که در بازه LP و RP شناسایی شده‌اند.

مینیمم آرایه‌هایی را که در اولین پنجره (MW_1) جای گرفته‌اند محاسبه کنید.

بعد از این مرحله پنجره را برای بررسی عناصر باقی مانده و تعیین پنجره‌های معتبر دیگر (w_i) و مقدار می‌نیم آن پنجره‌ها به عنوان mw_1, mw_2, \dots, mw_i بلغزانید.

مقدار مینیمم MW_i را که اندازه کادر را برای آستانه T_s می‌هد محاسبه کنید.

حال تصویر را با آستانه (T_s) باینری کنید. کادرهای با اندازه‌ی کمتر از آستانه T_s حذف می‌شوند. پیکسل‌های متن پیش‌زمینه با رنگ سفید و پس‌زمینه با رنگ سیاه نشان داده شده است.

1=کادر {Ts >= اگر اندازه کادر }
0=کادر {Ts <= اگر اندازه کادر }

یافته شده تنها عرض یک پیکسلی دارند و از هم پوشانی لبه‌ها در اثر افزایش پهنای لبه جلوگیری شود. چون اپراتور فرسایش همه نقاط مرزی را پاک می‌کند لبه‌ها از تصویر بیرون کشیده می‌شوند و تمامی نقاط مرزی دارای پهنای یک پیکسل می‌شوند اگر تصویر فرسایش یافته از تصویر اصلی کم شود، نتیجه لبه‌ای با عرض یک پیکسل خواهد بود.

تصویرهای 'ه.۲' و 'و.۲' نتیجه لبه‌یابی را بعد از آستانه‌گیری تکراری و فرسایش نشان می‌دهند. با کمک مولفه‌های پیوسته، پس از مرحله لبه‌یابی به محاسبه کادر دربرگیرنده نوشته می‌پردازیم. هر مولفه همانطور که در شکل ۳.۳ نشان داده شده، هر جزء حاصل شده را به عنوان یک کادر لبه در نظر می‌گیرد.



شکل ۲: لبه یابی و مراحل مختلف آن

۴-۱- زدودن لبه‌های نادرست و باینری کردن به وسیله نظریه sliding window

هر یک از لبه‌ها از نظر ویژگیهای مختلف و مشخصه‌های متن بطور جداگانه بررسی شده و لبه‌هایی که منطبق با معیارها نباشند حذف می‌گردند. ویژگیهای زیر برای زدودن لبه‌های غیر متنی در نظر گرفته می‌شود:

از جهت کشیده شدگی نسبت ۰/۱ تا ۱۰ برای حذف مناطقی که زیاد کشیده‌اند در نظر گرفته می‌شود.

اگر یک کادر لبه، دارای یک یا دو کادر است که به طور کامل درون آن قرار می‌گیرد، کادرهای داخلی را می‌توان به راحتی نادیده گرفت زیرا با مرزهای داخلی کارکترهای متن منطبق هستند. از طرف دیگر، چنانچه به طور کامل سه یا تعداد بیشتری را محصور کند، فقط کادر داخلی تر نگه داشته می‌شود و کادر بیرونی حذف می‌شود. بنابراین عناصر ناخواسته با بیرون کشیدن لبه با شرایط زیر حذف می‌گردند [۲]:

if ($N_{int} < 3$)



خروجی روش Otsu



خروجی روش Niblack



خروجی روش پیشنهادی: sliding window

شکل ۴: مقایسه عملکرد روش پیشنهادی با دیگر روشها

نتایج نشان می‌دهد که شیوه پیشنهادی شیوه‌ای موثر بوده و نسبت به شیوه‌های دیگر در شرایط پس زمینه بافتی پیچیده به نتیجه بهتری می‌رسد. در شیوه پنجره لغزشی (sliding windows) نسبت به انواع مقایسه شده به خوبی از پیچیدگی-های پس‌زمینه پرهیز شده است، تا اینکه تصویر باینری شده بهتر بتواند توسط OCR تشخیص داده می‌شود.

تصویر پردازش شده پس از این گامها در شکل ۳ نشان داده شده است:



شکل ۳ راست: کادر لبه چپ: تصویر باینری شده

۵- نتیجه گیری

شیوه پنجره لغزشی (sliding windows) برای شناسایی لبه‌های کاراکتری از طریق پاک کردن لبه‌های غیر متنی و ناخواسته ارائه شده است. نتایج آزمایش نشان دهنده عملکرد موثر شیوه پیشنهادی در برابر دیگر الگوریتم‌های باینری است. عملکرد شیوه ارائه شده با دو شیوه آستانه‌گیری معروف مقایسه شد. این دو متد عبارت از الگوریتم‌های [11] Niblack [12] Otsu هستند. بر این باوریم که ارزیابی چشمی برای برآورد کیفی شیوه ما کافی گویا است.



تصویر ورودی

- backgrounds", International Conference on Audio, Language and Image Processing, 2008. ICALIP 2008. pg. 873- 877.
- [11] W. Niblack. "An Introduction to Digital Image Processing". Prentice Hall, Englewood Cliffs, New Jersey, 1986
- [12] N. Otsu. "A threshold selection method from gray level histogram". IEEE Transactions on System, Man, Cybernetics, 19(1):62-66, January 1978.
- [13] Egyul Kim, SeongHun Lee, JinHyung Kim, "Scene Text Extraction using Focus of Mobile Camera 2009 10th International Conference on Document Analysis and Recognition". pg:166-170.
- [14] Kenneth R. Castleman. "Digital Image Processing". Prentice-Hall International, Inc., Bei Jing, BJ, 1998
- [15] Yan Wenzhong, Shen Shuqun, "An Edge Detection Method for Chromosome Images", The 2nd International Conference on Bioinformatics and Biomedical Engineering, 2008, pp. 2390-2392.
- [16] M. Sarifuddin and R. Missaoui. "A New Perceptually Uniform Color Space with Associated Color Similarity Measure for Content-Based Image and Video Retrieval". In Proc. of ACM SIGIR 2005 Workshop on Multimedia Information Retrieval (MMIR 2005), pp 1-8, 2005.
- [17] Yi Zhang and Kok Kiong Tan, "Text extraction from images captured via mobile and digital devices", Int. J. Computational Vision and Robotics, Vol. 1, No. 1, 2009, pp.34-58.

مراجع

- [1] Cheolkon Jung and Joongkyu Kim, "Player Information Extraction for Semantic Annotation in Golf Videos", IEEE Transactions on broadcasting, vol. 55, no. 1, March 2009, pp. 79-83.
- [2] Cheolkon Jung and Joongkyu Kim, "Player Information Extraction for Semantic Annotation in Golf Videos", IEEE Transactions on.
- [3] Yuanping Zhu, "Augment document image binarization by learning", 19th International Conference on Pattern Recognition, ICPR 2008, pg. 1-4.
- [4] Wolf, C. Doermann, D., "Binartization of low quality text using a Markov random field model", 16th International Conference on Pattern Recognition, 2002, vol 3, pp. 160- 163
- [5] Jonghyun Park, Toan Nguyen Dinh, and Gueesang Lee, "Binartization of Text Region based on Fuzzy Clustering and Histogram Distribution in Signboards", Proc. world academy of science, engineering and technology vol. 33 September 2008 pg.85-90.
- [6] Xiaoyi Jiang, Computer Society, and Daniel Mojon, "Adaptive Local Thresholding by Verification-Based Multithreshold Probing with Application to Vessel Detection in Retinal Images", IEEE Transactions on Pattern Analysis And Machine Intelligence, vol. 25, no. 1, January 2003, pg-131-137.
- [7] Rui Wu, Jianhua Huang, Xianglong Tang, Jiafeng Liu, "A Text Image Segmentation Method Based on Spectral Clustering", Computer and Information Science, Vol 1, No.4, November 2008.
- [8] Anh-Nga Lai GueeSang Lee, "Binartization by Local K-means Clustering for Korean Text Extraction", IEEE International symposium on signal processing and information technology 2008, pp. 117- 122
- [9] Huang, Huadong Ma, He Zhang, "A New Video Text Extraction Approach" IEEE International Conference on Multimedia and Expo, 2009. ICME 2009.
- [10] Wuyi Yang Shuwu Zhang Zhi Zeng Haibo Zheng, "Method combination to extract text from images and videos with complex