

## طبقه‌بندی معنایی تصاویر متنی مبتنی بر الگوی ارزشمندی متن

حسین پورقاسم<sup>(۱)</sup> - محمد صادق هل فروش<sup>(۲)</sup> - سیلان دانشور<sup>(۳)</sup>

(۱) دانشکده برق - دانشگاه آزاد اسلامی واحد نجف‌آباد

(۲) دانشکده مهندسی برق و الکترونیک - دانشگاه صنعتی شیراز

(۳) گروه مهندسی پزشکی - دانشگاه صنعتی سهند

**خلاصه:** استخراج اطلاعات از تصاویر متنی شناسایی شده از مجرای اینترنت یکی از مهمترین معضلات سیستم‌های جمع‌آوری اطلاعاتی در حوزه تکنولوژی اطلاعات می‌باشد. این مسأله وقتی حادث می‌گردد که بدانیم در بین انبوه تصاویر متنی تنها درصد ناچیزی از تصاویر متنی شناسایی شده، دارای ارزش اطلاعاتی می‌باشند. در این مقاله، جهت تجزیه و تحلیل تصاویر متنی و دسترسی به محتوای آنها، از یک روش طبقه‌بندی مبتنی بر ناحیه‌بندی تصویر استفاده شده است. در این الگوریتم به کمک یک روش ناحیه‌بندی دومرحله‌ای، نواحی تصویر شناسایی شده، سپس به کمک یک ساختار سلسله‌مراتبی طبقه‌بندی، نوع ناحیه از لحاظ متنی یا عکسی (غیرمتنی) بودن مشخص می‌گردد. در ادامه با تعریف بارزش بودن متن یک تصویر متنی، سعی می‌شود تصویر متنی در یکی از دو گروه معنایی بارزش و بی‌ارزش دسته‌بندی گردد. الگوریتم پیشنهادی بر روی یک پایگاه داده تصاویر متنی و غیرمتنی که از تصاویر موجود در اینترنت فراهم شده است، مورد ارزیابی قرار می‌گیرد. نتایج آزمایشها بیانگر کارایی روش پیشنهادی در طبقه‌بندی معنایی تصاویر براساس تعریف کاربر از بارزش و بی‌ارزش بودن تصاویر متنی است. الگوریتم پیشنهادی، صحت طبقه‌بندی ۹۸/۸٪ برای طبقه‌بندی تصاویر متنی بارزش از بی‌ارزش را فراهم کرده است.

**کلمات کلیدی:** طبقه‌بندی محتوایی، تصاویر متنی و غیرمتنی، ارزشمندی اطلاعاتی، سیستم‌های جمع‌آوری اطلاعات.

### ۱- مقدمه

رشد فزاینده اسناد کاغذی دیجیتال (اسناد کاغذی اسکن شده) در مجاری اینترنتی و سامانه‌های خودکارسازی اداری، نیاز به بایگانی و دسترسی آنها جهت استخراج و داده‌کاوی اطلاعات را به یک نیاز مهم تبدیل کرده است. با توجه به زمان‌بر بودن این فرآیند به صورت دستی (توسط کاربر)، ارائه ابزار خودکاری که بتواند این نیاز را مرتفع کند، ضروری به نظر می‌رسد. بنابراین در یک سامانه خودکار جستجو و بازیابی اطلاعات متنی، تجزیه و تحلیل اسناد متنی<sup>(۱)</sup> جهت دسترسی به اطلاعات این اسناد امری اجتناب‌ناپذیر تلقی می‌شود [۱]. تجزیه و تحلیل اسناد اسکن شده، مقدمه‌ای بر دسترسی به اطلاعات معنایی و مفهومی نهفته در تصویر متنی است. این اطلاعات سطح بالا جهت طبقه‌بندی معنایی تصاویر می‌تواند ابزار نیرومندی در اختیار طراحان سیستم‌های جمع‌آوری اطلاعات قرار دهد. اسناد به سه روش عمده تجزیه و تحلیل می‌شوند [۲]:

۱- تجزیه و تحلیل بر اساس محتویات سند: در این دسته روش صرفنظر از اینکه چطور اسناد ایجاد شده‌اند، براساس محتویاتشان به دو دسته تقسیم می‌شوند: طبیعی (مانند عکس صورت و تصاویر

ماهواره‌ای) و نمادین (مانند کارهای چاپی، فرم‌های اداری و

طراحی‌های مهندسی) [۲].

۲- تجزیه و تحلیل بر اساس لایه‌بندی سند: در این دسته روش، لایه‌بندی سند مورد بررسی قرار می‌گیرد. این دسته روش نیز به نوبه خود به دو دسته تقسیم می‌شود: تجزیه و تحلیل هندسی (فیزیکی) و تجزیه و تحلیل منطقی. تجزیه و تحلیل هندسی، خصوصیات هندسی است که مناطق همگون مانند متن، جدول، تصویر، نقاشی و ... را دسته‌بندی می‌کند. در مقابل، تجزیه و تحلیل منطقی، به نوع صفحه‌بندی می‌پردازد و برچسب‌های موجود در صفحه مانند عنوان، علامت تجاری، پاورقی، حروف بزرگ و کوچک و ... را مشخص می‌کند [۱، ۳].

۳- تجزیه و تحلیل بر اساس متن: در این تجزیه و تحلیل، سند به دو قسمت متنی و غیرمتنی (گرافیکی) تقسیم می‌شود. برای دسته‌بندی و جداسازی نواحی متنی از نواحی گرافیکی، از فیلترهای موجود در پردازش تصویر استفاده می‌شود [۴-۵].

اکثر کارهایی که اخیراً انجام شده است، تجزیه و تحلیل سند را براساس ناحیه‌بندی تصاویر سندی به نواحی متنی و گرافیکی موجود

انتها نیز برچسب نواحی ناحیه‌بندی شده توسط ناحیه‌بند دوم توسط دو طبقه‌بند (MLP و ماشین بردار پشتیبان (SVM)) مشخص می‌گردد. در هر سه طبقه‌بند، از ویژگیهای بافت جهت طبقه‌بندی نواحی استفاده می‌شود. ویژگیهایی همچون Tamura (شامل زبری، جهت‌دار بودن و کنتراست)، همبستگی، انرژی، همگنی و آنتروپی مستخرج از ماتریس هم‌خدادی و همچنین دو ویژگی از زیرباند‌های تبدیل موجک استخراج می‌گردد. در ادامه با تعریف الگوی ارزشمندی متن و استخراج ویژگی از این الگو، یک طبقه‌بند شبکه عصبی آموزش داده می‌شود تا در مرحله آزمایش تصاویر متنی با ارزش از بقیه تصاویر مجزا گردد.

در ادامه این مقاله، در بخش (۲) چارچوب پیشنهادی طبقه‌بندی معنایی تصاویر معرفی می‌گردد. در این بخش، جزئیات ناحیه‌بندی دومرحله‌ای، استخراج ویژگی، تعریف الگوی ارزشمندی متن و طبقه‌بندی معنایی تصویر تشریح می‌شود. در بخش (۳) آزمایشهای انجام شده جهت ارزیابی الگوریتم پیشنهادی آورده شده است. بخش (۴) مروری بر عملکرد الگوریتم پیشنهادی ارائه می‌دهد. در انتها نیز در بخش (۵) نتیجه‌گیری از کار انجام شده در این مقاله ارائه می‌گردد.

## ۲- چارچوب پیشنهادی طبقه‌بندی معنایی تصاویر

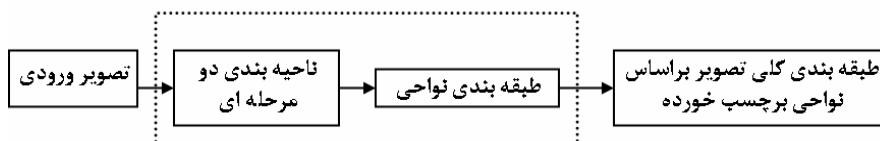
شکل (۱) روندنمای کلی چارچوب پیشنهادی برای طبقه‌بندی معنایی تصاویر متنی مبتنی بر طبقه‌بندی نواحی تصویر را نشان می‌دهد. در چارچوب پیشنهادی تصویر ورودی، طی دو مرحله ناحیه‌بندی، به یک سری نواحی با برچسبهای متنی و گرافیکی ناحیه‌بندی می‌شود. در مرحله آخر نیز، با آموزش الگوهای تصاویر متنی با ارزش و بی‌ارزش تصاویر مطلوب و نامطلوب از یکدیگر جدا می‌شوند. در این مقاله جهت طبقه‌بندی نواحی مستخرج از تصویر از الگوریتم معرفی شده در [۱] استفاده شده است. روندنمای طبقه‌بند سلسله‌مراتبی مبتنی بر ناحیه‌بندی با جزئیات بیشتر در شکل (۲) نشان داده شده است.

در [۱]، از یک الگوریتم ناحیه‌بندی دومرحله‌ای برای ناحیه‌بندی تصویر و به دنبال آن طبقه‌بندی نواحی استفاده شده است. در مرحله اول این الگوریتم، تصویر ورودی با یک روش ناحیه‌بندی مبتنی بر تبدیل موجک ناحیه‌بندی می‌شود. سپس این نواحی با طبقه‌بند اول، طبقه‌بندی می‌شود. نواحی متنی شناسایی شده از تصویر ورودی حذف شده و نواحی گرافیکی در تصویر باقی می‌مانند تا دوباره و با دقت بیشتر توسط ناحیه‌بند دوم (ناحیه‌بند مبتنی بر آستانه‌گذاری) ناحیه‌بندی شده و به کمک دو طبقه‌بند (طبقه‌بند دوم و سوم)، متنی یا گرافیکی بودن نواحی مشخص می‌گردد. برای استفاده بهتر از الگوریتم طبقه‌بندی مبتنی بر ناحیه‌بندی ارائه شده در [۱]، مرحله استخراج ویژگی از نواحی را تغییر دادیم. بدین صورت که با اضافه کردن یک ویژگی بافت دیگر (ویژگی Tamura) به بردار ویژگی سعی شده است که بردار ویژگی تقویت شود. در ادامه تنها به تشریح ویژگی جدید Tamura پرداخته می‌شود.

در تصویر انجام داده‌اند (دسته روش سوم). در این تحقیقات ناحیه‌بندی معمولاً به یکی از سه روش بالا به پایین (Top-Down)، پایین به بالا (Bottom-Up) و مبتنی بر بافت (Texture) انجام شده است [۷،۶]. روش‌هایی که به صورت پایین به بالا عمل می‌کنند، از اجزای کوچک تصویر (پیکسل) استفاده کرده و با اتصال این اجزا به هم و تشکیل اجزای بزرگتر برای شناسایی قطعات اصلی موجود در تصویر استفاده می‌کنند. به عبارت دیگر این روشها، ناحیه‌بندی را از پیکسل شروع کرده و به ناحیه یا شیئی می‌رسند. اشکال این روش در حساس بودن به اندازه حروف، وضوح سند اسکن شده و فاصله میان حروف است [۴]. از جمله این روش‌ها می‌توان به الگوریتم Docstrum اشاره کرد که براساس K- نزدیکترین همسایگی از اجزا صفحه عمل می‌کند [۸]. در روش‌هایی که به صورت بالا به پایین عمل می‌کنند، عملیات ناحیه‌بندی از بزرگترین جزء، یعنی تمام صفحه شروع شده تا به کوچکترین قطعه در صفحه می‌رسد. اشکال این روش در تشخیص زاویه و انحنا متن است. زیرا محدود به بلوکهای مستطیلی است [۴]. از جمله این روش‌ها می‌توان به الگوریتم بازگشتی X-Y Cut [۹] اشاره کرد. در هر مرحله مقادیر واریانس در جهت‌های افقی و عمودی محاسبه شده و با مقدار آستانه‌ای که از پیش تعریف شده است، مقایسه می‌گردد و در انتها نیز ناحیه‌بندی انجام می‌گیرد. روش‌هایی که بافت را برای ناحیه‌بندی بکار می‌برند، از ویژگیهای بافت تصویر برای ناحیه‌بندی استفاده می‌کنند [۱۰]. در دو روش اول (پایین به بالا، بالا به پایین) نیاز به پیش‌اطلاعاتی درباره فاصله‌های افقی و عمودی، مجزا بودن متن‌ها، عکس‌ها، گرافیک‌ها و حتی فرض‌هایی درباره متن، سبک و اندازه می‌باشد. با توجه به اشکالات و موارد مطرح شده در مورد روشهای ناحیه‌بندی، در این مقاله از روش ناحیه‌بندی مبتنی بر بافت استفاده می‌شود.

در بسیاری از روشهای ارائه شده قبلی در زمینه طبقه‌بندی، برای آنکه ویژگیها با اطلاعات محلی در اختیار قرار گیرند تا عملکرد طبقه‌بندی بهبود یابد، ویژگیها بصورت محلی استخراج می‌شوند [۱۱]. برای این منظور، تصویر به چندین ناحیه ثابت مثلاً بلوکهای مستطیلی هم‌اندازه تقسیم شده و ویژگیها از هر بلوک استخراج می‌شوند. در این مقاله، با استفاده از یک ساختار طبقه‌بندی سلسله‌مراتبی مبتنی بر ناحیه‌بندی تصاویر سندی ارائه شده در [۱]، نواحی متنی و گرافیکی تصویر ورودی مشخص می‌شود. در این نقشه، از یک الگوریتم ناحیه‌بندی دومرحله‌ای مبتنی بر تبدیل موجک و آستانه‌گذاری استفاده می‌شود. در اولین مرحله از الگوریتم ناحیه‌بندی، از تبدیل موجک [۴-۷،۱۲، ۵] برای ناحیه‌بندی استفاده می‌شود. سپس با استفاده از یک طبقه‌بند شبکه عصبی پرسپترون چندلایه (MLP)، برچسب نواحی از نظر متنی بودن یا گرافیکی (غیرمتنی) بودن مشخص می‌گردد و نواحی متنی شناسایی شده از تصویر کنار گذاشته می‌شوند. در مرحله دوم با اعمال یک ناحیه‌بندی ثانویه مبتنی بر آستانه‌گذاری بر روی باقی‌مانده تصویر حاصل از ناحیه‌بندی اولیه، تصویر بصورت کامل ناحیه‌بندی می‌شود. در

طبقه‌بندی تصویر مبتنی بر ناحیه‌بندی



شکل (۱): روندنمای کلی چارچوب پیشنهادی طبقه‌بندی معنایی تصاویر.

افقی و عمودی برای هر پیکسل محاسبه می‌شود:

$$E_{k,h}(x,y) = |A_k(x+2^{k-1},y) - A_k(x-2^{k-1},y)| \quad (2)$$

$$E_{k,v}(x,y) = |A_k(x,y+2^{k-1}) - A_k(x,y-2^{k-1})| \quad (3)$$

بعد از این، مقدار  $k$ ی که  $E$  را در یک جهت حداکثر می‌کند، جهت مشخص کردن اندازه پنجره فیلتر برای هر پیکسل استفاده می‌شود.

$$S_{best}(x,y) = 2^k \quad (4)$$

در نهایت زبری از میانگین  $S_{best}$  کل تصویر محاسبه می‌شود یعنی:

$$F_{crs} = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n S_{best}(i,j) \quad (5)$$

ب- کنتراست

رابطه‌ای که از روی آن کنتراست محاسبه می‌شود، به صورت زیر است:

$$F_{con} = \frac{\sigma}{\alpha_4^{1/4}} \quad (6)$$

که  $\alpha_4 = \mu_4 / \sigma^4$  ممان چهارم یا Kurtosis است.  $\mu_4$  ممان چهارم حول میانگین و  $\sigma^2$  واریانس است. این رابطه هم برای کل تصویر و هم برای یک ناحیه از تصویر قابل استفاده است.

ج- جهت‌دار بودن

جهت محاسبه جهت دار بودن، تصویر با دو ماتریس  $3 \times 3$  یعنی  $M_1$  و  $M_2$  کانوالو می‌شود و یک بردار گرادیان در هر پیکسل محاسبه می‌شود.

$$M_1 = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}, M_2 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \quad (7)$$

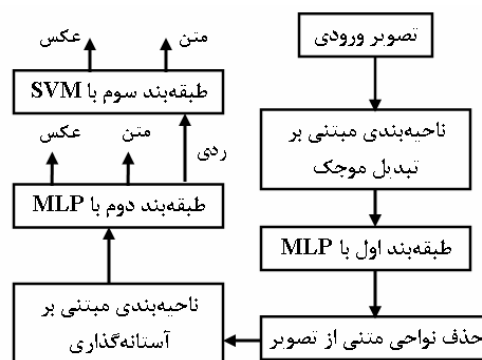
دامنه و فاز این بردار به صورت زیر تعریف می‌شود:

$$|\Delta G| = (|\Delta_H| + |\Delta_V|) / 2 \quad (8)$$

$$\theta = \tan^{-1}(\Delta_V / \Delta_H) + \pi / 2 \quad (9)$$

که  $\Delta_V, \Delta_H$  تفاضلات کانولوشن در دو جهت افقی و عمودی است.

سپس با عددی کردن  $\theta$  و شمارش تعداد پیکسلهایی که دامنه،  $|\Delta G|$ ، بزرگتر از یک حد آستانه دارند، یک هیستوگرام از  $\theta$  و  $H_D$  ساخته می‌شود. این هیستوگرام در مورد تصاویر جهتی دارای پیکهای بزرگی است و در مورد تصاویر غیرجهتی به صورت مسطح است. در نهایت از کل هیستوگرام جهت استخراج یک معیار جهتی براساس بلندی پیکها استفاده می‌شود.



شکل (۲): روند نمای طبقه‌بند سلسله‌مراتبی مبتنی بر ناحیه‌بندی [۱].

۱-۲- استخراج ویژگی

مهمترین مرحله در حل یک مسأله شناسایی الگو، مرحله انتخاب و استخراج ویژگی است. در این مقاله، از یک مجموعه کامل ویژگی بافت جهت طبقه‌بندی نواحی تصاویر در ساختار پیشنهادی [۱] استفاده شده است. این مجموعه شامل ویژگیهای Tamura (شامل زبری، جهت‌دار بودن و کنتراست)، همبستگی، انرژی، همگنی و آنتروپی مستخرج از ماتریس هم‌خدادی [۱۶] و همچنین دو ویژگی مستخرج از زبرباندهای تبدیل موجک می‌باشد [۱]. در ادامه جزئیات استخراج این ویژگیها آورده می‌شود.

۱-۱-۲- ویژگیهای Tamura

ویژگیهای تامورا، شامل زبری<sup>۲</sup>، کنتراست<sup>۳</sup>، جهت‌دار بودن<sup>۴</sup>، شباهت خطی<sup>۵</sup>، منظم بودن<sup>۶</sup> و سختی<sup>۷</sup> می‌شوند [۱۳]. این ویژگیها بواسطه مطالعات روان‌شناسی بر روی ادراک انسان از بافت طراحی شده است. در این مقاله، سه جزء اول این ویژگی استفاده می‌شود که در اینجا نیز جزئیات آنها آورده شده است.

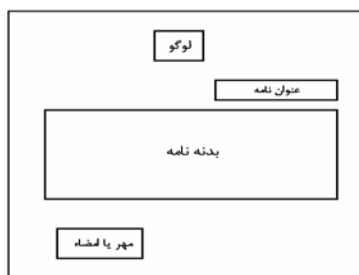
الف- زبری

زبری معیاری از دانه دانه شدن و زبر شدن بافت است. جهت محاسبه زبری، ابتدا با یک پنجره  $2^k \times 2^k$  ( $k=0,1,\dots,5$ ) بر روی هر پیکسل  $(x,y)$ ، فیلتر (Moving Average) MA، به صورت زیر اعمال می‌شود:

$$A_k(x,y) = \sum_{i=x-2^{k-1}}^{x+2^{k+1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k+1}-1} \frac{g(i,j)}{2^{2k}} \quad (1)$$

در رابطه بالا  $g(i,j)$  شدت روشنایی در پیکسل  $(i,j)$  است.

سپس تفاضلات بین جفت  $A_k(x,y)$ های غیرهمپوشان در جهت‌های



شکل (۳): الگوی نوعی یک متن بارز.

در بخش قبل الگوی کلی ارزشمند بودن متن تشریح شد. طبق تعریف فوق، ویژگیهایی از الگوی ارزشمندی تعریف و استخراج می‌گردد تا به کمک آن بتوان یک شبکه عصبی پرسپترون چند لایه را آموزش داد. پس از انجام آموزش، از شبکه جهت تعیین مقدار بارز بودن تصویر ورودی استفاده می‌گردد. در ادامه با یک آستانه‌گذاری ساده بر روی این مقدار فازی می‌توان دسته بارز بودن (کلاس مطلوب) یا بی‌ارزش بودن (کلاس نامطلوب) تصاویر متنی را تعیین کرد.

ویژگیهایی که از الگوی ارزشمندی استخراج می‌گردد به صورت زیر است:

الف- نسبت مساحت نواحی متنی و عکسی به کل تصویر.

ب- نسبت مساحت نواحی عکسی به نواحی متنی.

ج- مساحت هر ناحیه.

د- موقعیت هر ناحیه (مرکز جرم هر ناحیه).

ه- موقعیت نسبی مراکز جرم هر دو ناحیه (یعنی بالا، پایین، راست، چپ بودن).

و- نوع ناحیه (متنی یا گرافیکی بودن).

با بردارهای ویژگی شکل گرفته به صورت بالا، جهت آموزش یک شبکه عصبی استفاده می‌گردد. طبقه‌بند مذکور در سه لایه و با الگوریتم پس‌انتشار خطا آموزش می‌بیند تا در مورد نمونه‌های ناشناخته ورودی تصمیم‌گیری کند. شکل (۴) روند نامی طبقه‌بندی معنایی تصاویر متنی مبتنی بر ویژگیهای مستخرج از الگوی ارزشمندی به دو گروه بارز و بی‌ارزش را نشان می‌دهد.

### ۳- نتایج آزمایشها

#### ۳-۱- پایگاه داده

در این مقاله، جهت ارزیابی و آزمایش الگوریتم‌های پیشنهادی ناحیه‌بندی و طبقه‌بندی، از یک پایگاه داده تصاویر که از تصاویر موجود در شبکه اینترنت فراهم شده است، استفاده می‌شود. این مجموعه شامل ۹۷۷ سند تصویری است که به ترتیب شامل ۵۷۸ سند محض (یعنی تصویری که فقط شامل متن است)، ۸۱ تصویر محض (تصویر بدون متن) و ۳۱۸ سند شامل متن و تصویر (مانند روزنامه) می‌باشد. ۵۷۸ تصویر متنی خالص را می‌توان به اسناد دست‌نوشته، اسناد اداری، کارت‌های تبلیغاتی، روزنامه (چند ستونی)، اسناد تایپی، اسناد جدول‌بندی‌شده، فرم‌های ویندوزی و اینترنتی تقسیم نمود. نمونه‌ای از انواع تصاویر پایگاه داده در شکل (۵) نشان داده شده است.

$$F_{dir} = \sum_p \sum_{\phi \in W_p}^{n_p} (\phi - \phi_p)^2 H_D(\phi) \quad (10)$$

در رابطه بالا P پیک،  $n_p$  تعداد پیکها و  $W_p$  تعداد بیت‌هایی است که پیک p را شکل داده است.  $\phi_p$  بیتی است که پیک p بر روی آن قرار گرفته است.

### ۲-۲- تعریف ارزشمندی متن

در این مقاله، جهت انجام طبقه‌بندی معنایی تصاویر متنی شناسایی شده، از یک ساختار قابل تعریف ارزشمندی متن توسط کاربر استفاده می‌شود. بدین معنی که پس از انجام طبقه‌بندی نواحی استخراج شده از تصویر و مشخص شدن متنی یا عکسی بودن هر ناحیه از تصویر توسط الگوریتم طبقه‌بندی مبتنی بر ناحیه‌بندی، با مقایسه فازی این اطلاعات با الگوی از پیش تعریف شده ارزشمندی متن، میزان ارزشمند بودن متن مشخص می‌شود. بنابراین در تعریف الگوی ارزشمندی متن و فرآیند مقایسه فازی (به عبارت دیگر طبقه‌بندی معنایی تصویر) آن با یک تصویر، دو نکته نهفته است که در ادامه به آن پرداخته می‌شود. در این مقاله، تصاویر متنی مشابه به نامه‌های رسمی و اداری بارز تلقی می‌گردد. به عبارت دیگر، متن با ارزش باید حداقل شامل نواحی متنی و گرافیکی زیر باشد:

الف- یک ناحیه گرافیکی در بالای صفحه (در وسط یا سمت راست صفحه) که بیانگر علامت تجاری یا لوگوی یک شرکت یا سازمانی باشد.

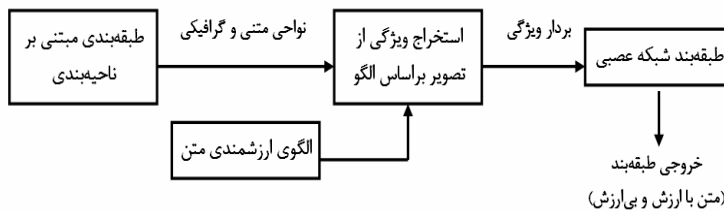
ب- یک ناحیه متنی در زیر لوگو که بیانگر عنوان نامه یا طرفین مکاتبه باشد.

ج- یک ناحیه متنی در زیر ناحیه عنوان که بیانگر بدنه نامه می‌باشد.

د- یک ناحیه گرافیکی در زیر ناحیه بدنه که بیانگر مهر یا امضاء نامه می‌باشد. در موارد ذکر شده فوق دو مسأله بصورت نسبی مطرح می‌گردد. یکی، موقعیت نواحی در صفحه و همچنین موقعیت نسبی نواحی نسبت به یکدیگر است. دیگری، مساحت نسبی هر ناحیه متنی یا گرافیکی می‌باشد. شکل (۳) الگوی نوعی یک متن با ارزش (نامه رسمی - اداری) را نشان می‌دهد.

### ۲-۳- طبقه‌بندی معنایی تصویر براساس الگوی ارزشمندی

مسأله‌ای که در طبقه‌بندی تصاویر متنی با ارزش و تصاویر متنی بی‌ارزش وجود دارد، بحث تعریف معیار یا به عبارتی الگوی با ارزش بودن متن موجود در تصویر است. همانگونه که مشخص است، تصمیم‌گیری نهایی در مورد محتوای یک تصویر ناشناخته، مسأله‌ای است که بصورت قابل توجهی به نظر و انتظار کاربر از سیستم طبقه‌بند دارد. بدین معنی که با توجه به تعریفی که کاربر از تصویر دلخواه یا غیردلخواه دارد، می‌توان تصمیم‌گیری در این مورد را انجام داد.



شکل (۴): روند نمای طبقه‌بندی معنایی تصاویر متنی مبتنی بر ویژگیهای مستخرج از الگوی ارزشمندی.

### ۳-۲- طبقه‌بندی نواحی تصویر

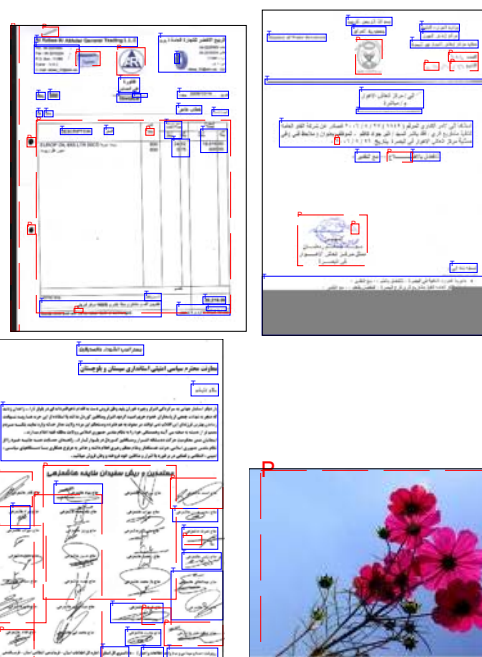
الگوریتم طبقه‌بندی مبتنی بر ناحیه‌بندی ارائه شده در [۱]، نواحی مختلف تصویر ورودی را شناسایی کرده و کلاس متنی یا گرافیکی بودن آن را مشخص می‌کند. در این قسمت، نتایج اعمال این الگوریتم بر روی چندین نمونه از تصاویر پایگاه داده نشان داده شده است. نمونه تصاویر بصورتی انتخاب شده است تا اینکه توانایی الگوریتم در حالتها و شرایط مختلف مورد ارزیابی قرار گیرد شکل (۶). در تصاویر ناحیه‌بندی شده شکل (۶) نواحی متنی با مستطیل‌های آبی رنگ با خطوط ممتد و نواحی غیرمتنی (نواحی عکسی) با مستطیل‌های قرمز رنگ با خطوط خط‌چین نشان داده شده است. در تصاویر ناحیه‌بندی شده مشاهده می‌گردد که در اکثر موارد نواحی متنی به درستی تشخیص داده شده‌اند و محل‌های امضاء، لوگوهای موجود در تصاویر و جداول به عنوان نواحی غیرمتنی (نواحی شکلی) در نظر گرفته شده‌اند. در یکی از تصاویر (گل قرمز) نیز کل تصویر به عنوان ناحیه غیرمتنی در نظر گرفته شده است و هیچ ناحیه متنی شناسایی نشده است. جدول (۱)، نتایج طبقه‌بندی نواحی مستخرج از تمام تصاویر پایگاه داده را نشان می‌دهد. در مجموع طبقه‌بندها ۱۹۸۴۲ ناحیه را به درستی شناسایی کرده‌اند. در حالیکه ۵۰۳ ناحیه را نتوانسته‌اند به درستی شناسایی کنند. یعنی در مجموع صحت طبقه‌بندی برابر ۹۷/۵٪ بوده است و تنها ۲/۵٪ نواحی شناسایی نشده‌اند جدول (۱).

### ۳-۳- طبقه‌بندی معنایی تصاویر

جهت بررسی عملکرد طبقه‌بند معنایی پیشنهادی، مجموعه‌ای از تصاویر پایگاه داده با نرم‌افزار طبقه‌بند تصویر به دو گروه تصاویر متنی با ارزش (مطلوب) و غیرمتنی (نامطلوب) طبقه‌بندی می‌شود شکل (۷). معیار مطلوب و نامطلوب بودن با توجه به تعریف الگوی ارزشمندی متن در نظر گرفته شده است. به این معنی که تصاویر حاوی نامه‌های رسمی و اداری به عنوان تصاویر با ارزش (تصاویر مطلوب) و بقیه تصاویر (شامل تصاویر windows و web، تصاویر خالص، روزنامه‌های اینترنتی و ...) به عنوان تصاویر بی‌ارزش در نظر گرفته شده است. همانگونه که در شکل (۷) مشاهده می‌گردد، تصاویر خالص که دارای طیف رنگی وسیعی هستند در گروه تصاویر نامطلوب طبقه‌بندی شده‌اند. در مقابل، تصاویر با محتوا رنگی ضعیف در گروه تصاویر مطلوب قرار می‌گیرند. از طرف دیگر، تمام موارد شامل نامه‌های اداری و رسمی، گذرنامه و متن‌های روزنامه‌ای به عنوان تصاویر با ارزش در نظر گرفته می‌شوند.



شکل (۵): نمونه‌ای از انواع مختلف تصاویر پایگاه داده.



شکل (۶): عملکرد الگوریتم پیشنهادی ناحیه‌بندی و طبقه‌بندی بر روی نمونه تصاویر. نواحی متنی با مستطیل‌های آبی رنگ با خطوط ممتد و نواحی غیرمتنی با مستطیل‌های قرمز رنگ با خطوط خط‌چین مشخص شده‌اند.

جدول (۱): صحت عملکرد طبقه‌بندی نواحی شناسایی شده.

نوع سند	تعداد موارد صحیح (TA)	تعداد موارد اشتباه (FA)	مجموع	صحت (%)
دست نوشته	۸۱۶	۲۳	۸۳۹	۹۷/۲۵
اداری	۱۶۸۰	۳۸	۱۷۱۸	۹۷/۷۸
کارت‌های تبلیغاتی	۱۴۲	۵	۱۴۷	۹۶/۶
روزنامه	۱۶۶۶	۳۰	۱۶۹۶	۹۸/۲۳
جدول‌بندی شده	۲۸۵۷	۷۵	۲۹۳۲	۹۷/۴۵
فرم‌های مبتنی بر web و windows	۴۰۴۶	۹۳	۴۱۳۹	۹۷/۷۵
شامل متن و تصویر باهم	۱۹۷۰	۷۵	۲۰۵۵	۹۵/۸۶
تصاویر محض	۲۲۶	۳۹	۲۶۵	۸۵/۲۸
اسناد دیگر	۶۴۳۹	۱۲۵	۶۵۶۴	۹۸/۰۹
جمع	۱۹۸۴۲	۵۰۳	۲۰۳۴۵	۹۷/۵۲

است. در این دسته روشها، تنها به نوع محتویات و موقعیت فیزیکی آنها پی می‌بریم در حالی که هیچ اطلاعاتی از مفاهیم و معانی نهفته در آنها در اختیار قرار نمی‌گیرد. در این مقاله با در نظر گرفتن این نکته، سعی شد به کمک ویژگیهای سطح پایینی همچون ویژگیهای بینایی تصویر مفاهیم سطح بالای موجود در آن تعریف شود. به عبارت دیگر یک تناظری بین مفاهیم سطح بالا و ویژگیهای سطح پایین برقرار گردد. موفقیت نسبی در برقراری این تناظر حداقل در دو مورد کمک شایانی به سیستم‌های جمع‌آوری اطلاعات می‌کند. اولاً، همانگونه که در بخش (۳-۳) آورده شد. اعمال این چنین الگوریتمی بر روی یک پایگاه داده بزرگ ۲۳۹۰ تصویری، باعث فیلتر شدن قسمت عمده‌ای از تصاویر بی‌ارزش موجود در مجاری اینترنت می‌شود. به عبارت دیگر، به جای اینکه کاربر ۲۳۹۰ تصویر را به صورت دستی مورد بررسی قرار دهد. با این الگوریتم تنها کافی کمتر از ۲٪ آن را یعنی ۴۵ تصویر را مورد بررسی دقیق‌تر قرار دهد. ثانیاً، ایجاد تناظر فوق وابستگی زیادی به نظرات کاربران دارد که این مسأله نکته مثبتی تلقی می‌گردد. چون که می‌توان با کمک گرفتن از نظرات کاربران در یک فرآیند بازخورد ربط جهت بهبود تناظر فوق استفاده کرد.

تنها موردی که به نظر می‌رسد بی‌ارزش باشد، تصویر مربوط به یک مدار چاپی است که در گروه تصاویر با ارزش دسته‌بندی شده‌اند. تمام موارد شامل تصاویر رنگی، تصاویر فرمهای نرم‌افزارها و صفحات وبگاه‌های اینترنتی به عنوان تصاویر بی‌ارزش در نظر گرفته می‌شوند. در مجموع الگوریتم پیشنهادی، تصاویر متنی با ارزش را از بقیه تصاویر با صحت ۹۸/۸٪ طبقه‌بندی می‌کند.

در ادامه برای ارزیابی کامل‌تر الگوریتم پیشنهادی، از یک پایگاه داده با ۲۳۹۰ تصویر اینترنتی استفاده شده است که از نظر فراوانی تصاویر متنی و گرافیکی با دنیای واقعی مطابقت کامل دارد. جدول (۲) نتایج طبقه‌بندی ۲۳۹۰ تصویر اینترنتی را نشان می‌دهد. در این طبقه‌بندی از مجموع ۹ تصویر با ارزش تنها یک تصویر نوشتاری با کیفیت پایین در دسته تصاویر بی‌ارزش طبقه‌بندی شده‌است. در مقابل نیز از مجموع ۲۳۸۱ تصویر بی‌ارزش تنها ۳۷ تصویر در دسته تصاویر با ارزش طبقه‌بندی شده‌است. به عبارت دیگر درصد پذیرش اشتباه (False Accept) به عنوان تصاویر با ارزش ۱/۵٪ می‌باشد. این نتایج نشان می‌دهد که کمتر از ۲٪ تصاویر اینترنتی برای بررسی توسط کارشناسان جدا می‌گردد.

جدول (۲): نتایج طبقه‌بندی تصاویر اینترنتی به دو گروه تصاویر با ارزش و بقیه تصاویر.

نوع تصویر	نتایج طبقه‌بندی		تعداد
	متنی با ارزش	بقیه تصاویر	
متنی با ارزش	۸	۱	۹
بقیه تصاویر	۳۷	۲۳۴۴	۲۳۸۱
مجموع	۴۵	۲۳۴۵	۲۳۹۰

#### ۴- بحث

دسترسی به اطلاعات موجود در تصاویر مسأله مهمی است که همواره مورد نظر محققان زمینه تحقیقاتی پردازش تصویر بوده است. در کاربرد تصاویر متنی، تنها ابزاری که برای نیل به این هدف مورد استفاده قرار می‌گیرد، روشهای تجزیه و تحلیل طرح‌بندی صفحات

**۵- نتیجه گیری**

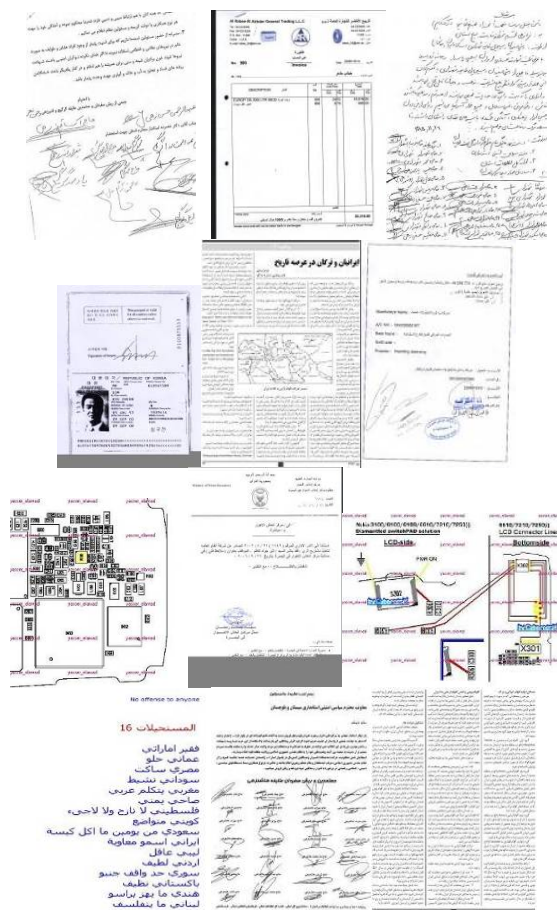
در این مقاله، جهت دستیابی به مفاهیم و معانی موجود در یک تصویر متنی، با استفاده از یک الگوی متنی با ارزش و نتایج حاصل از طبقه بندی نواحی تصویر، یک روش کارآمد جهت طبقه بندی تصاویر متنی با ارزش از بقیه تصاویر ارائه شد. در این ساختار، تصویر به وسیله الگوریتم ناحیه بندی دمرحله ای، ناحیه بندی می شود. سپس متنی بودن و غیرمتنی (عکسی) بودن نواحی حاصل به کمک چندین طبقه بندی در یک ساختار سلسله مراتبی طبقه بندی مشخص می گردد. در انتها، با استخراج ویژگی از الگوی متنی با ارزش و آموزش یک شبکه عصبی، تصاویر براساس محتوای متنی و عکسی، به دو گروه با ارزش و بی ارزش دسته بندی می شوند. الگوریتم پیشنهادی بر روی یک پایگاه داده تصاویر متنی و غیرمتنی که از تصاویر موجود در اینترنت فراهم شده است، مورد ارزیابی قرار گرفت. نتایج آزمایشها بیانگر کارایی روش پیشنهادی در طبقه بندی معنایی تصاویر براساس تعریف کاربر از ارزش بودن و بی ارزش بودن تصاویر متنی است.

**سپاسگزاری**

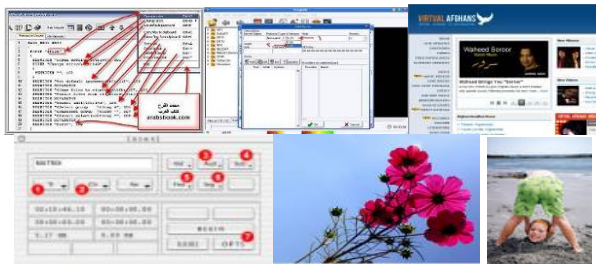
این مقاله مستخرج از طرح پژوهشی مصوب در دانشگاه آزاد اسلامی واحد نجف آباد می باشد.

**پی نوشت:**

- 1- Document Image Analysis
- 2- Coarseness
- 3- Contrast
- 4- Directionality
- 5- Linelikeness
- 6- Regularity
- 7- Roughness



(الف)



(ب)

شکل (۷): عملکرد الگوریتم طبقه بندی پیشنهادی بر روی یک مجموعه تصاویر آزمایشی. (الف) - تصاویر طبقه بندی شده با برچسب مطلوب و (ب) - تصاویر طبقه بندی شده با برچسب نامطلوب.

**مراجع**

[۱] حسین پورقاسم، "تجزیه و تحلیل طرح بندی تصاویر متنی مبتنی بر طبقه بندی نواحی در یک ساختار سلسله مراتبی تصمیم گیری"، مجله الکترونیک- قدرت دانشگاه آزاد اسلامی واحد نجف آباد، شماره دوم، ص: ۳۷-۴۴، بهار ۱۳۸۸.

[2] G.Nagy, "Twenty years of document image analysis in PAMI", IEEE Trans. Patt. Ana. and Mac. Intel., Vol.22, pp.38-62, 2000.

[3] N.Amamoto, S.Torigoe, Y.Hirogaki, "Block segmentation and text area extraction of vertically/horizontally written document", CDAR, pp.739-742, Tsukuba, Japan, 1993.

[4] M.Acharyya, M.K.Kundu, "Document image segmentation using wavelet scale-space features", IEEE Trans. Circ. and Sys. for Vid. Tech., Vol.12, No.12, pp.1117-1127, 2002.

[5] A.Busch, W.W.Boles, S.Sridharan, "Texture for script identification", IEEE Trans. Patt. Ana. and Mac. Intel., Vol.27, No.11, pp.1720-1732, Nov. 2005.

[6] S.Mao, T.Kanungo, "Empirical performance evaluation methodology and its application to page segmentation algorithms", IEEE Trans. Patt. Ana. and Mac. Intel., Vol.23, No.3, pp.242-256, March 2001.



- [7] Y.Qiao, Z.Lu, C.Song, d.S.Sun, "Document image segmentation using Gabor wavelet and kernel-based methods", Systems and Control in Aerospace and Astronautics, pp.451-455, Jan. 2006.
- [8] L.O'Gorman, "The document spectrum for page layout analysis", IEEE Trans. Patt. Ana. and Mac. Intel., Vol.15, pp.1162-1173, 1993.
- [9] G.Nagy, S.Seth, M.Viswanathan, "A prototype document image analysis system for technical journals", Computer, Vol.7, pp.10-22, 1992.
- [10] A.Jain, Y.Zhong, "Page segmentation using texture analysis", Patt. Rec., Vol.29, pp.743-770, 1996.
- [11] J.T.Laaksonen, J.M.Koskela, E.Oja, "Picsom - a framework for content-based image database retrieval using self-organizing maps", SCIA, June 1999, URL: <http://www.cis.hut.fi/picsom/publications.html>.
- [12] Q.Ye, Q.Huang, W.Gao, D.Zhao, "Fast and robust text detection in images and video frames", Ima. and Vis. Comp., Vol.23, No.6, pp.565-576, June 2005.
- [13] H.Tamura, S.Mori, T.Yamawaki, "Texture features corresponding to visual perception", IEEE Trans. on Sys., Man., and Cyb., Vol.8, No.6, June 1978.
- [14] W.Niblack, R.Barber, W.Equitiz, M.Flickner, E.Glasman, D.Petkovic, P.Yanker, C.Faloutsos, G.Taubin, "The QBIC project: Querying images by content using color, texture, and shape", SPIE, pp.173-187, 1993.
- [15] A.Pentland, R.W.Picard, S.Sclaroff, "Photobook: Content-based manipulation of image databases", Int. Jou. of Comp. Visi., Vol.18, No.3, pp.233-254, June 1996.
- [16] R.M.Haralick, K.Shanmugan, I.Dinstein, "Textural features for image classification", IEEE Trans. on Sys., Man. and Cyb., Vol.3, No.6, pp.610-621, 1973.
- [17] H.Li, D.Doermann, O.Kia, "Automatic text detection and tracking in digital video", IEEE Trans. Ima. Proc., pp.147-156, 2000.
- [18] R.Lienhart, A.Wernicke, "Localizing and segmenting text in images and videos", IEEE Trans. Circu. and Sys. for Video Tech., pp.256-268, 2002.

### رزومه



حسین پورقاسم استادیار گروه مخابرات دانشکده برق دانشگاه آزاد اسلامی واحد نجفآباد، دکتری تخصصی برق با گرایش مهندسی پزشکی را در سال ۱۳۸۷ از دانشگاه تربیت مدرس اخذ کرده است. وی مدرک کارشناسی و کارشناسی ارشد خود را در همین رشته در سالهای ۱۳۸۱ و ۱۳۸۳ دریافت کرده است. نامبرده در تمامی این دوره‌ها دانشجوی ممتاز بوده است. زمینه‌های تحقیقاتی مورد علاقه وی شامل زمینه‌های مختلف پردازش تصویر مانند نمایه‌گذاری و بازیابی تصاویر، بیومتریک، ردیابی اشیاء، مخفی کردن اطلاعات، شناسایی الگو و شبکه‌های عصبی مصنوعی می‌باشد و در این زمینه‌ها تاکنون بیش از ۲۰ مقاله در مجلات و کنفرانسهای خارجی و داخلی به چاپ رسانیده است.



محمد صادق هل فروش کارشناسی و کارشناسی ارشد خود را در رشته مهندسی الکترونیک به ترتیب از دانشگاه شیراز و دانشگاه صنعتی شریف در سالهای ۱۳۷۲ و ۱۳۷۴ دریافت کرد. او در سال ۱۳۸۵ موفق به اخذ مدرک دکتری در رشته مخابرات گرایش سیستم از دانشگاه تربیت مدرس شد. دکتر هل فروش اکنون استادیار گروه مخابرات در دانشگاه صنعتی شیراز است. زمینه‌های پژوهشی مورد علاقه او شناسایی الگو با تاکید بر سیستمهای بیومتریک و پردازش تصویر است.



سیلان دانشور فارغ‌التحصیل کارشناسی مهندسی پزشکی از دانشگاه علوم پزشکی شهید بهشتی در سال ۱۳۷۵، کارشناسی ارشد مهندسی پزشکی - بیوالکتریک سال ۱۳۸۲ و دکتری مهندسی پزشکی - بیوالکتریک سال ۱۳۸۶ بوده و در حال حاضر استادیار گروه مهندسی پزشکی دانشگاه صنعتی سهند تبریز هستند. زمینه‌های مورد علاقه تحقیقاتی ایشان پردازش تصاویر پزشکی، پردازش سیگنال و مدلسازی بیولوژیکی می‌باشد.