یک طرح ترکیبی کارآمد برای استخراج فریم کلیدی و موقعیت یابی متن در ویدئو

خلاصه

الگوریتم های بهینه برای ثبت متن و موقعیت یابی آن در سکانس های ویدئویی در بازار چند رسانه ای و استخراج داده امروز بسیار پرطرفدار هستند. به خاطر چالش هایی از قبیل وضوح تصویر پایین، کنتراست پایین، پیش زمینه پیچیده و متنون با استایل، جهت، رنگ و چینش مختلف، استخراج متن از تصویر ویدئویی یک کار چالش بر انگیز است. در این مقاله روشی برای استخراج بهینه و کارآمد فریم های کلیدی از ویدئو بر اساس لحظات رنگی و پس از آن موقعیت یابی متن تنها بر روی همین فریم های کلیدی انجام میشود. به این خاطر که اطلاعات متن با هر فریم تغییر نمیکند، استخراج متن تنها از این فریم های کلیدی میتواند به کاهش هزینه محاسباتی و زمانی الگوریتم کمک شایانی بکند. علاوه بر این، این مقاله یک روش کارآمد هایبرید برای محلی کردن صحنه و متن گرافیکی در فریم های ویدئویی آن هم با استفاده از DWT (تبدیل موج دو بعدی هار)[1] ، لاپلاس فیلتر گاوسی و روش تفاوت حداکثری گرادیان ارائه میکند. DWT یک روش تجزیه سریع تصویر ارائه میکند که تصویر را به سه بخش جزئیات تخمینی میشکند. این سه جزء اطلاعاتی درباره لبه های عمودی، افقی، محوری از تصویر در خود دارند که باعث میشود متن سریعتر تشخیص داده شود. روش تفاوت گرادیان حداکثری نیز برای موقعیت یابی و محلی کردن بیشتر متن در تصویر به کار میرود، دامنه تفاوت گرادیان هم در فرآیند حد نصاب سنجی[2] به کار میرود. یک تکنیک حد نصاب سنجی پویا برای تبدیل نوع تصویر به فرم باینری مورد استفاده قرار گرفته است. از آنجایی که این تکنیک مقادیر متنوعی برای تصاویر مختلف حاصل میکند، میتواند برای موقعیت یابی اتوماتیک متن در تصاویر ویدئویی به کار رود. دو عملگر ماسک هم برای به دست آوردن یک معادله به کار رفته اند و زمانی استفاده میشوند که پیکسل ها مساوی با مقدار حد نصاب تعیین شده

[1] 2-D haar discrete wavelet transform
[2] thresholding process

باشند. مثبت و منفی ها با استفاده از عملگرهای مورفولوژیکی حذف میشوند و آنالیز اجزای به هم پیوسته صورت میگیرد تا در نهایت جایگاه متن مشخص گردد. معیارهای مقایسه در نتایج نشان میدهند که روش ارائه شده عملکرد مناسبی در نرخ شناسایی، نرخ هشدار نادرست و نرخ شناسایی نادرست ارائه میکند.

کلمات کلیدی: شناسایی عکس، حالات رنگی، استخراج فریم های کلیدی، تبدیل موج گسسته، لاپلاس فیلتر گاوسی، تفاوت گرادیان.

1. معرفی

با پیشرفت های اخیر در تکنولوژی چند رسانه ای، افزایش قابل توجهی در پایگاه داده تصاویر و ویدئوهای دیجیتالی بوجود آمده است. در نتیجه آن نیاز به شاخص گذاری چند رسانه ای کارآمد و تکنیک های استخراج حس میشود. برچسب گذاری ویدئو بر اساس محتوا یکی از حوزه های در حال رشد از تحقیقات گذشته است. محتوای ویدئو را میتوان به صورت زیر دسته بندی کرد: الف. محتوای ادراکی، مبتنی بر ویژگی هایی از جمله شکل، شدت، رنگ، بافت و تغییرات موقت و ب. محتوای معنایی – بر اساس اشیا موجود در ویدئو، دسته بندی کرد. متون قرار گرفته در ویدئوها اطلاعات ارزشمندی دارند و به سادگی میتوانند برای منظور برچسب گذاری معنایی ویدئو به کار روند. متاسفانه روشی مستحکم و قوی برای اینکار وجود ندارد تا بتواند متون را از تمامی انواع ویدئوها استخراج کند. متنون ویدئویی را میتوان به دسته های زیر تقسیم کرد: الف. متون صحنه که به صورت طبیعی از ویدئو ثبت میشوند، ب. متن زیرنویس که به صورت مجزا در ویدئو جاسازی شده اند. خصوصیات نامطلوب دیگری هم در ویدئو وجود دارد از جمله: پس زمینه پیچیده، وضوح پایین، شدت پایین، و اندازه ها، استایلها، رنگ ها و جهات متنوع متن در ویدئو چالش هایی هستند که پیش راه محققان این حوزه قرار گرفته اند. در بین متون صحنه و متن زیرنویس، کاملا واضح است که استخراج متون صحنه بسیار دشوارتر است.

در این مقاله، یک شمای کارآمد برای استخراج اولیه فریم های کلیدی از ویدئو با استفاده از لحظات رنگی و پس از آن تبدیل موج گسسته، تفاوت حداکثر گرادیان و عملگر های مورفولوژیکی برای موقعیت یابی متن در فریم های کلیدی ویدئو استفاده میشوند. باقی مطالب مقاله به صورت زیر تنظیم شده اند، بخش دوم، یک چشم انداز کلی از روش های بکار رفته و کارهای مرتبط با این حوزه ارائه میکند. روش پیشنهادی در بخش سوم نمایش داده شده است. نتایج آزمایشی و معیارهای مقایسه در بخش چهارم ارائه شده اند. در نهایت، نتایج در بخش پنجم مشخص شده اند.

2. کارهای قبلی

الگوریتم های بیشماری برای موقعیت یابی، استخراج و تشخیص متن در سکانس های تصویری ویدئو در سال های اخیر ارائه شده اند. شناسایی متن و تکنیک های موقعیت یابی را میتوان به دو دسته متفاوت دسته بندی کرد: الف. بر اساس منطقه، ب. تکنیک های مبتنی بر بافت.

روش های مبتنی بر منطقه بر ویژگی های مناطق تصویر برای استخراج متن استفاده میکند با این فرض که تفاوت عمده ای بین خصوصیات ویدئو/تصویر و متن و پس زمینه کنارش وجود دارد. ویژگی های لبه ها، رنگها و روش های اجزای متصل از جمله تکنیک های به کار رفته عمده در پیاده سازی این دسته موقعیت یابی هستند. روش های مبتنی بر مناطق تصویر به شیوه ای از پایین به بالا کار میکنند. در ابتدا تصویر به مناطق محتمل حاوی کاراکتر متون تقسیم میشود؛ این مناطق در ادامه بیشتر تقسیم شده و خطوط حاوی متن را حاصل میکنند. قدم نهایی مشخص کردن مناطق حاوی متن و فاقد متن است.

روش های مبتنی بر بافت از معیارهای کمی برای تنظیم شدت/رنگ زیر عناصر در منطقه ای از ویدئو استفاده میکنند تا آن محدوده را از پس زمینه جدا کنند. این تکنیک ها معمولا از فیلترهای گابور استفاده میکنند، همچنین از تجزیه موج، تبدیل کسینوسی گسسته، FFT، واریانس فضایی و غیره. در ابتدا ویژگی های بافت ار تصویر/ویدئو جدا میشوند و پس از آن محدوده های تصویر با استفاده از این روش مشخص میگردند. با اینکه روش های مبتنی بر بافت تصویر

بسیار قوی و کارآمد هستند برای پس زمینه های پیچیده تصاویر در مقایسه با روش مبتنی بر منطقه تصویر پیچیدگی محاسباتی بیشتری دارند.

چونگ-وی لیانگ از DWT و عملگرهای مورفولوژیکی در کار خود استفاده کرده بود، برای استخراج متن به صورت منطقه ای در تصاویر استاتیک یا سکانس های ویدئو با استفاده از DWT و عملگرهای ذکر شده. وی یک شمای هرمی برای شناسایی متن در تصاویر استفاده کرده است. دی. چن هم یک روش دو مرحله مشابه هم برای تشخیص و شناسایی متن در تصاویر پیچیده و فریم های ویدئویی ارائه کرده است. این روش شامل مراحل زیر میشود: الف. فرآیند موقعیت یابی متن سریع باعث میشود نرمالیزه شدن اندازه متن انجام شود و ب. یک روش قوی یادگیری ماشین برای فرآیند اعتبار سنجی اعمال میشود که بر ویژگی های مستقل در پس زمینه دلالت دارد. شیواکومارا هم روشی ارائه کرده است که با استفاده از روش تمایز گرادیان برای جدا کردن بخش های محتمل متن به کار میرود.

روش پیشنهادی زمان محاسبه را کاهش میدهد. با استفاده از استخراج متن تنها از فریم های کلیدی به جای کل فریم های ویدئو که میتواند بسیار زمان بر باشند. علاوه بر این، متون ویدئویی معمولا چندین جهت گیری و چینش متفاوت دارند و با استفاده از DWT میتواند جزئیات لبه را به صورت همزمان در جنبه های افقی، عمودی و روی محورهای اصلی مشخص کرد.

3. روش پیشنهادی

فلوچارت چارچوب پیشنهادی برای موقعیت یابی متن در ویدئو در تصویر 1 آورده شده است. جزئیات هر بلوک فرآیند در زیر تشریح شده اند.

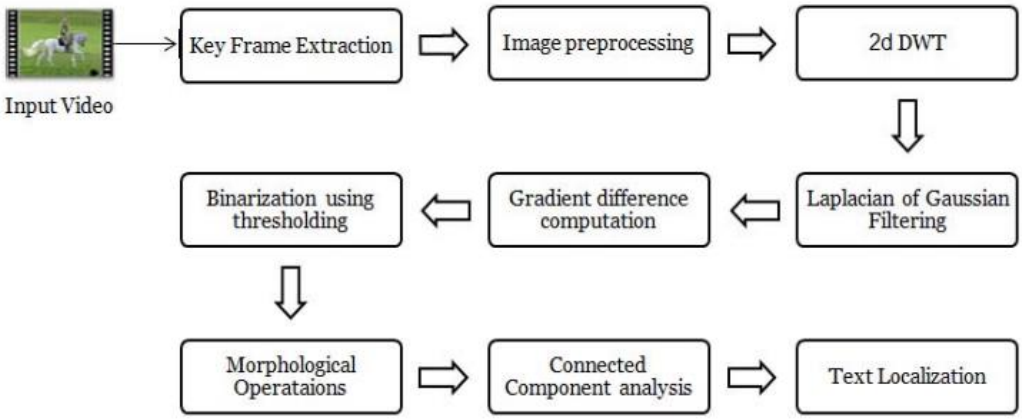


Fig. 1 Flowchart of the proposed method

A. استخراج فریم کلیدی

یک ویدئو از چندین فریم تشکیل یافته است که به صورت دنباله ای از هم در یک ویدئو نمایش داده میشوند. وقتی یک ویدئو ضبط میشود تمامی این تصاویر به صورت مداوم با یک دوربین ثبت شده و به عنوان یک عمل دنباله دار از نظر زمانی و فضایی ارائه میشود. این واحد های موقتی پایه شات، یا برداشت نامیده میشوند. برداشت ها با گذارهای تدریجی به یکدیگر متصل میشوند. این گذارها را میتوان با استخراج خصوصیات ویژه هر فریم شناسایی کرد و اگر شاخص تمایز برای دو فریم پشت سر هم بسیار بالا باشد، نشان دهنده گذار از برداشتی به برداشت دیگر هستیم. در مدل پیشنهادی، یک شات را میتوان بر اساس لحظه رنگی استخراج کرد. این مفهوم یک فاصله اقلیدسی است که تجانس بین فریم ها را اندازه میگیرد.

برای محاسبه لحظه رنگی، ابتدا مقادیر RGB هر فریم به مدل رنگی YIQ تبدیل میشوند. نگاشت رنگی YIQ روشنایی Y و رنگ I و Q را تفکیک میکند. شدت Y در تصویر با فرمول زیر حاصل میشود:

$$Y = 0.299R + 0.587G + 0.114B \qquad (1)$$

فرمول برای محاسبه رنگ $h^{th};\ h = 1,2,3,\ldots,$ از i^{th} جزء رنگ چنین حاصل میشود:

$$M_i^h = \left(\frac{1}{N}\sum_{k=1}^{M}(p_{i,k} - M_i^1)^h\right)^{\frac{1}{h}} \qquad (2)$$

در اینجا، $p_{i,k}$ شدت i^{th} بخش رنگ از k^{th} پیکسل فریم است و N تعداد کل پیکسل های موجود در فریم است. هر چند برای راحتی محاسبه، در اینجا تنها از کانال Y استفاده کرده ایم و تنها دو لحظه رنگی ابتدایی را در محاسبات وارد نموده ایم (انحراف خالص و استاندارد). بر اساس این لحظات رنگی، یک بردار ویژگی از هر فریم ساخته میشوند، بطوریکه:

$$F_j = [\alpha_1 M_1^1, \alpha_1 M_1^2, \dots, \alpha_1 M_1^H, \alpha_1 M_1^1, \alpha_2 M_2^2, \dots, \alpha_2 M_2^H] \quad (3)$$

پس از آن فاصله اقلیدسی بین فریم های پشت سرهم k_j و k_{j-1} به صورت زیر محاسبه میشوند:

$$D = d(F_j - F_{j-1}) - d(F_{j-1} - F_{j-2}) \qquad (4)$$

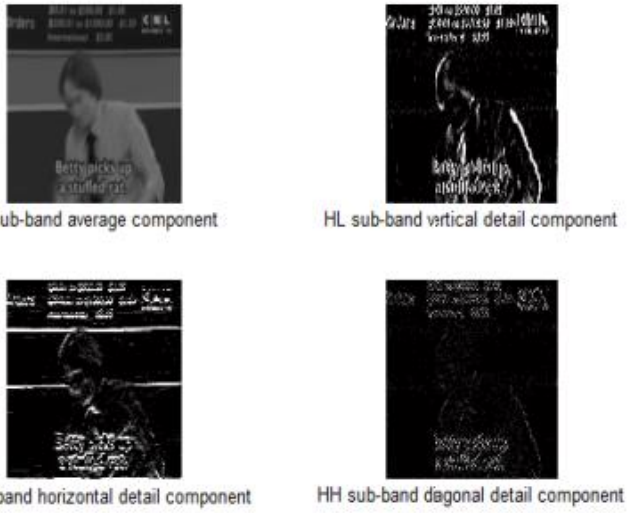
در اینجا $d(F_j - F_{j-1}) = \sum_{k=1}^{z}|F_j(k) - F_{j-1}(k)|$ و q مقدار 2 دارند. اگر $D > T$ باشد، پس یک فریم کلیدی شناسایی شده است. حد نصاب T هم مقداری خالص مساوی با تمام فواصل اقلیدسی خواهد داشت.

B. استخراج لبه

1) DWT: وقتی که فریم های کلیدی استخراج شدند، هر فریم به یک تصویر سیاه و سفید تبدیل میشود، اگر ویدئو رنگی باشد. DWT یکی از تبدیلات مفید برای آنالیز چند وضوحه تصاویر است. در DWT دو بعدی، سیگنال تصویر ورودی به چهار زیر باند تقسیم میشود؛ LL – LH – HL – HH. اینکار با یکبار فیلتر کردن تصویر به صورت سطری و حذف کردن دوتا در میان انجام میشود. زیر تصویر بوجود آمده بار دیگر به صورت ستونی فیلتر میشود و با حذفیات مشابه. زیر باند LL اجزای میانی تصویر را نمایش میدهد، در حالیکه LH – HL و HH به ترتیب جزئیات اجزای افقی، عمودی و قطری را نشان میدهند. معمولا متن هر سه نوع از لبه ها را ارائه میکند که از این جزئیات اجزا حاصل شده اند. دلیل اصلی برای استفاده از DWT دو بعدی جهت استخراج لبه ها این است که میتواند تمامی سه نوع لبه را به صورت همزمان شناسایی کند. در مقایسه با روش های سنتی در این زمینه زمان محاسبه کاهش پیدا میکند. یکی

دیگر از مزایای DWT در این است که میتواند نویز را حذف کند، در حالیکه سایر شناسایی کننده ها نویز را نیز به عنوان لبه حساب میکنند.

(a)



(b)

Fig. 2. (a) Gray image of the original image (b) DWT coefficients

2) فیلترینگ لاپلاس ماسک گاوسی: سه جزئیات لبه ای که با روش DWT تفکیک شده بودند با استفاده از یک عملگر فیلترینگ لاپلاس ماسک گاوسی 5x5 فیلتر میشوند تا بلوک های حاوی متن در هر یک از اجزای جزئیات استخراج شود. از فیلترهای سطح پایین تر برای صیقلی کردن تصویر به واسطه حذف کردن نویز ها استفاده میشود. این فیلترها معمولا عملگرهای ماسک را اعمال میکنند. برای شناسایی لبه از روش لاپلاس از بلوک هایی از تصویر استفاده میشود که تغییر شدت رنگ آنی داشته باشند. تصویر حاصل پس از اعمال فیلتر لاپلاسی و تغییر بین این مقادیر برای شناسایی

متن و گذارهای پس زمینه استفاده میشود. صیقلی کردن لاپلاسی مشکلات حساسیت به نویز را کاهش میدهد و در عین حال با محدود کردن تصویر به چند باند فرکانس معین گذارهای صفر را نیز حذف میکند. تابع LoG دو بعدی با مرکزیت صفر به صورت زیر است:

$$LoG(x,y) = -\frac{1}{\pi\sigma^4}\left[1 - \frac{x^2+y^2}{2\sigma^2}\right]e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (5)$$

C. تمایز گرادیان ماکزیمم

اطلاعات گرادیان مناطق حاوی متن در یک تصویر به طور بخصوصی از پس زمینه تصویر متفاوت تر است. مقادیر مثبت و منفی در اجزای فیلتر شده توسط LoG در یک تصویر با استفاده از روش تمایز گرادیان ماکزیمم (MGD) استخراج میشود. MGD به عنوان یک تمایز بین مقادیر ماکزیمم و مینیمم داخل پنجره $1 \times N$ است. برای تصویر فیلتر شده با کمک LoG یعنی f ، MGD در پیکسل (x,y) به صورت زیر محاسبه میشود:

$$MGD(x,y) = \max\big(F(x,y)\big) - \min\left(F(x,y)\right) \qquad (6)$$

به طور معمول مناطق حاوی متن مقدار MGD بزرگتری نسبت به پس زمینه دارند بنابراین مناطق متنی روشنتر نمایش داده میشوند.

D. فرآیند باینری کردن

پس از حصول نگاشت گرادیان، هر یک از اجزای جزئیات یک تصویر به شکل باینری تبدیل میشوند. در اینجا ما از روش حد نصاب سنجی دینامیک استفاده کرده ایم. برای تعیین مقدار حد نصاب، دو عملگر ماسک برای حصول معادله به کار رفته اند. با اعمال این معادله بر روی هر پیکسل در کنار پیکسل های همسایه اش مقدار حد نصاب بازگردانده میشود. مقدار حد نصاب برای نگاشت گرادیان تمامی سه اجزای جزئیات محاسبه میشود. این یک روش حد نصاب

سنجی دینامیک است. پس G نگاشت گرادیان یک بخش جزئیات است و حد نصاب متناظر با آن یعنی T_G به صورت زیر محاسبه میگردد:

$$T_G = \frac{\sum(G(i,k)\times h(i,k))}{\sum h(i,k)} \qquad (7)$$

در اینجا، $h(i,k) = Max(|m_1 ** G(i,k)|)$ و $m1 = [-1 \quad 0 \quad 1]$ و $m2 = [-1 \quad 0 \quad 1]^t$ عملگرهای ماسک هستند.

E. موقعیت یابی متن

گام های پایه بکار رفته برای موقعیت یابی متن در این روش از قرار زیر هستند:

1. عملگرهای مورفولوژیکی: اتساع مورفولوژیکی بر روی تصویر باینری با 3 بخش جزئیات و با استفاده از عناصر ساختاری متفاوت برای هر بخش انجام میشود. در این مثال، یک مستطیل 3×8 به عنوان یک عنصر ساختاردهنده برای اجزای افقی و قطری و یک مستطیل 5×8 برای اجزای عمودی به کار رفته است.

2. AND منطقی: از آنجایی که متن حاوی اجزای افقی، عمودی و قطری است، این سه بخش با استفاده از یک عملگر AND منطقی با هم ترکیب میشوند تا مناطق حاوی متن از هم جدا شود.

3. حذف مثبت و منفی ها: در تصویر نهایی اجزای به هم متصل شده با استفاده از اتصال-8 برچسب گذاری میشوند اما همچنان ممکن است حاوی مقادیر مثبت نادرست باشد که از قواعد هندسی استفاده شده برای کاندید کردن مناطق حاوی متن به جا مانده باشد. این مقادیر به صورت تجربی چنین تعیین شده اند:

a) $Area \leq 70$
b) $Width > 60, Height < 20$

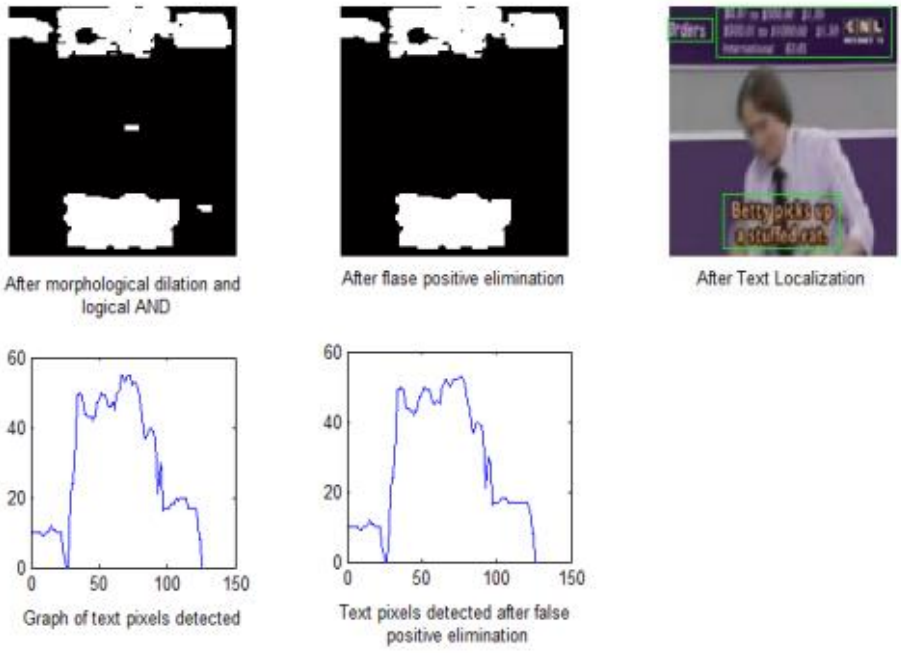


Fig.4 Intermediate results for text localization

4. نتایج آزمایشی و معیارهای مقایسه و ارزیابی

الگوریتم پیشنهاد شده با استفاده از نرم افزار متلب پیاده سازی شده است. برای اهداف آزمایشی ما مجموعه داده ویدئو خودمان را ایجاد کردیم که شامل ویدئوهای کارتونی، ورزشی و آموزشی میشد. الگوریتم بر روی چند تصویر هم آزمایش شده است. نتایج موقعیت یابی متن در ویدئوهای ساده و تصاویر در تصویرهای 5 و 6 نمایش داده شده اند.

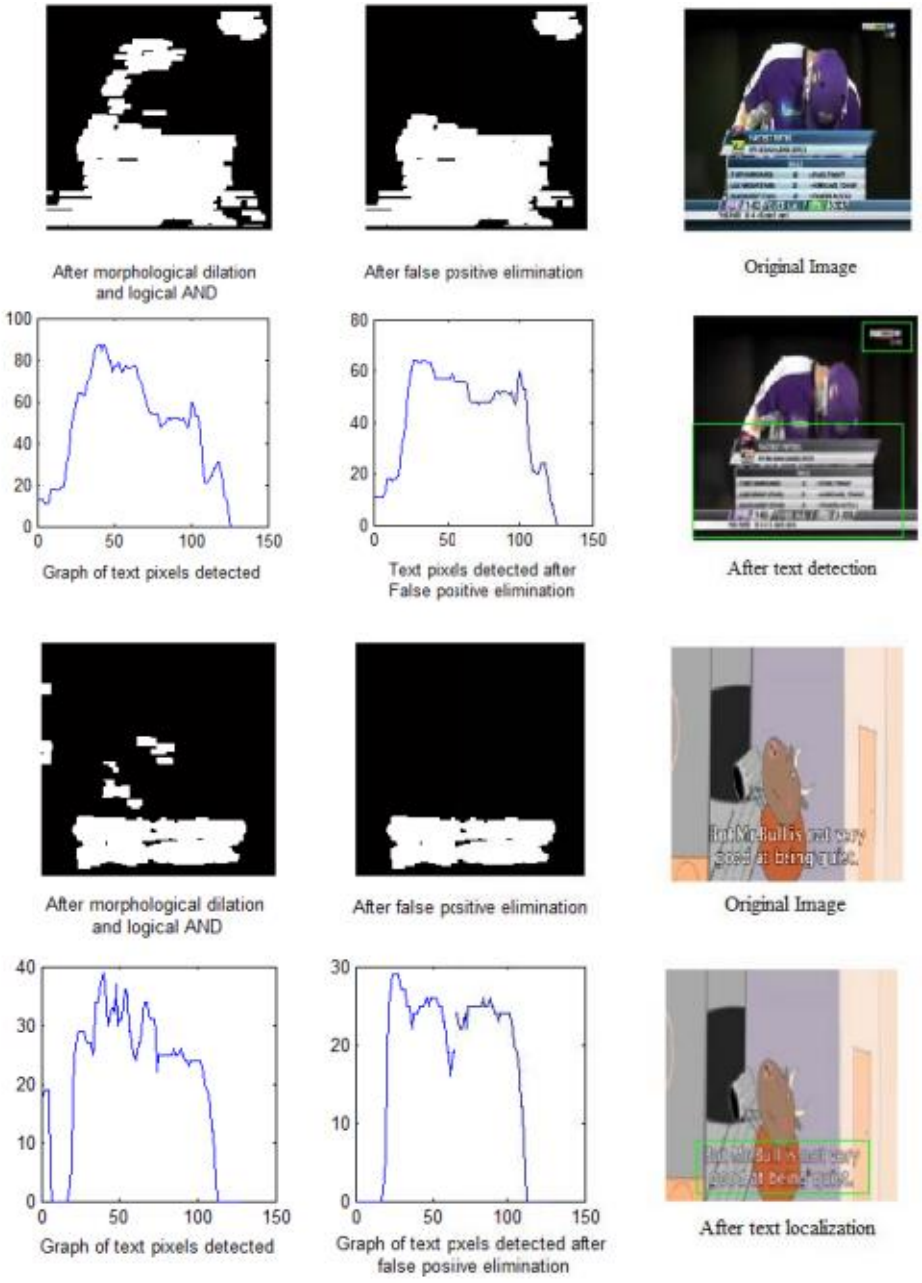


Fig. 5. Text localization results for sample videos

ارزیابی عملکرد این روش با استفاده از روش موجود و مشهور «دقت و حافظه» و F-measure انجام شده است.

Fig. 6. Text localization result for sample images

$$P = \frac{CorrectlyDetectedBlocks}{CorrectlyDetectedBlocks+FalsePositives} \qquad (8)$$

$$R = \frac{CorrectlyDetectedBlocks}{CorrectlyDetectedBlocks+FalseNegatives} \qquad (9)$$

در اینجا، مثبت های نادرست همان مناطقی در تصویر هستند که مناطق حاوی متن شناسایی شده اند، اما در واقع حاوی متن نیستند. منفی های نادرست هم مناطق کاندید شده هستند که در واقع حاوی متن هستند، اما شناسایی نشده اند. F-measure یک معیار ترکیبی است که دقت و حافظه و تبادل بین این دو را مشخص میکند. به صورت زیر:

$$F = \frac{2PR}{P+R} \qquad (10)$$

بلوکهای اعداد نهایی شناسایی شده برای موقعیت یابی متن به صورت دستی شمارش شده و به دسته های زیر تقسیم میشوند:

1) بلوک های کاندید متن (CTB): تعداد کل بلوک های شناسایی شده است.

2) بلوک های متن حقیقی (TTB): تعداد بلوک های کاندید شده ای که حاوی متن هستند.

3) بلوک های متن نادرست (FTB): تعداد بلوک متن های نادرست شناسایی شده هستند، برای مثال بدون حتی یک کاراکتر از متن

4) بلوک متن های جا افتاده (MTB): آنهایی هستند که ناقص شناسایی شده اند، برای مثال بلوک هایی با چند کاراکتر جا افتاده از متن.

با استفاده از مقادیر پارامتر فوق، دقت، حافظه و F-measure برای انواع ویدئوها محاسبه شده و نتایج به صورت زیر در جدول آمده است:

جدول 1: معیارهای مقایسه برای ویدئوهای نمونه

نمونه ویدئو	فریم های کلیدی	CTB	درصد دقت	درصد حافظه	F-measure
کارتونی	9	9	100	100	100
ورزشی	21	41	87.8	85.7	86
آموزشی	7	21	72	90	80

نتایج حاصل شده نشان میدهند که روش پیشنهادی نرخ شناسایی و حافظه خوبی برای هر دو نوع فریم ویدئوها و تصاویر ثابت دارد.

5. نتیجه گیری

در این مقاله، یک الگوریتم هایبرید (ترکیبی) ساده و کارآمد برای موقعیت یابی متن در فریم های ویدئو ارائه شد. روش نتایج خوبی برای تصاویر ثابت هم ارائه کرد. استخراج فریم کلیدی و استفاده از DWT دو بعدی کارآیی الگوریتم را بهبود بخشید و در عین حال زمان محاسباتی را کاهش داد. تکنیک حد نصاب سنجی مورد استفاده برای باینری کردن کاملا دینامیک بوده و بر پایه طیف تمایز گرادیان ماکزیمم عمل میکند. این روش حد نصاب های مختلف را برای تصاویر مختلف تولید میکند. محدودیت این روش در اینجاست که در ویدئوهایی با زمینه های پیچیده و روشنایی بالا تنها بخشی از متن را تشخیص میدهد. این روش در مواردی که پس زمینه و رنگ غیر قابل تمیز هستند نا کارآمد است. کارهای آتی انجام شده در این زمینه میتواند بر روی دسته بندی تصاویر به تصاویر با وضوح پایین و بالا تمرکز کرده و الگوریتم های مختلف بر پایه هر دسته را به کار برد.

REFERENCES

[1] Chung-Wei Liang and Po-Yueh Chen, "DWT Based Text Localization", International Journal of Applied Science and Engineering, 2004, pp. 105-116.

[2] Wei, Y. C., & Lin, C. H., "A robust video text detection approach using SVM", Expert Systems with Applications, 39(12), 2012, pp. 10832- 10840.

[3] Chen, D., Odobez, J. M., and Bourlard, H., "Text detection and recognition in images and video frames", Pattern Recognition, 37(3),2004, pp. 595-608.

[4] Shivakumara, P., Basavaraju, H. T., Guru, D. S., & Tan, C. L., "Detection of Curved Text in Video: Quad Tree Based Method", Document Analysis and Recognition (ICDAR), IEEE, August, 2013, pp. 594-598.

[5] Shekar, B., Kumari, M. S., and Holla, R., "An efficient and accurate shot boundary detection technique based on colour moments", International Journal of Artificial Intelligence and Knowledge Discovery, vol. 1, no. 1, 2011, pp. 77–80.

[6] Mallat, S. G., "A theory for multiresolution signal decomposition: the wavelet representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 7, 1989, pp. 674-693.

[7] Wong, E. K., and Chen, M., "A new robust algorithm for video text extraction", Pattern Recognition, 36, 2003, pp. 1397-1406.

[8] Phan, T. Q., Shivakumara, P., and Tan, C., "A laplacian method for video text detection," in Document Analysis and Recognition (ICDAR), 2009, pp. 66–70.

[9] Shivakumara, P., Phan, T. Q., and Tan, C., "A gradient difference based technique for video text detection," in Document Analysis and Recognition, ICDAR, 2009, pp. 156–160.

[10] Shivakumara, P., Phan, T. Q., & Tan, C. L., "A laplacian approach to multi-oriented text detection in video", Pattern Analysis and Machine Intelligence, IEEE, 33(2), 2011, pp. 412-419.

[11] B.H.Shekhar, Smitha M.L, Shivkumara, P., "Discrete wavelet transform and gradient difference based approach for text localization in videos", IEEE , 2014, pp.280-284.