

## **Automatic Detection of Salient Video Frames from Wireless Capsule Endoscopy using Adaptive Singular Value Decomposition**

A. Biniiaz<sup>1</sup>, F. Abdolali<sup>2</sup>, R. Aghaeizadeh Zoroofi<sup>3\*</sup>, O. Haji Maghsoudi<sup>4</sup>, Y. Sato<sup>5</sup>

<sup>1</sup> Ph.D. Student, Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

<sup>2</sup> Graduated Ph.D. Student, Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

<sup>3</sup> Professor, Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

<sup>4</sup> Ph.D. Student, Department of Bioengineering, Temple University, Philadelphia, USA, PA 19144

<sup>5</sup> Professor, Graduate School of Information Science, Nara Institute of Science and Technology (NAIST), Nara, Japan

Receipt in the Online Submission System: 24 September 2018, Received in Revised Form: 13 December 2018, Accepted: 7 January 2019

---

### **Abstract**

Wireless capsule endoscopy is a non-invasive diagnosis method which allows recording a video as the capsule travels through the gastrointestinal tract. The practical drawback is producing a long clinical video up to 8 hours and it takes about 2 hours to review the exam by an experienced expert. Video summarization methods can reduce the time required by experts and errors in manual interpretation. This paper presents an automatic method based on unique properties of adaptive singular value decomposition through sliding window that can reduce the long annotation time. By utilizing these properties, we are able to summarize a WCE video by outputting a motion video summary. Moreover, we apply an effective approach based on adaptive contrast diffusion to correct uneven illumination that deal with the low contrast generally caused by poor visibility conditions of the GI tract, WCE power and its structure. Experimental results on WCE videos indicate that a significant reduction in the review time is feasible. Quantitative and qualitative results of summarization show the effectiveness of proposed method that can be adapted to various clinical applications, such as training of young physicians, computer assisted diagnosis, medical decision support or medical document management.

**Keywords:** *Adaptive Contrast Diffusion, Singular Value Decomposition, Sliding Window, Video Summarization, Wireless Capsule Endoscopy*

---

\*Corresponding Author

Address: Bioelectric Department, Electrical Engineering Faculty, University of Tehran, P. O. Box: 14399-51154, Tehran, Iran

Tel: +98-21-61114302

Fax: +98-21-61114302

E-mail: zoroofi@ut.ac.ir

## تشخیص خودکار فریم‌های برجسته در کپسول آندوسکوپی بی‌سیم با استفاده از تجزیه به مقادیر تکین انطباقی

عباس بی‌نیا<sup>۱</sup>، فاطمه عبدالعلی<sup>۲</sup>، رضا آقایی‌زاده ظروفی<sup>۳\*</sup>، امید حاجی‌مقصودی<sup>۴</sup>، یوشینوبو ساتو<sup>۵</sup>

<sup>۱</sup> دانشجوی دکتری مهندسی پزشکی، آزمایشگاه پردازش تصویر هوشمند، دانشکده مهندسی کامپیوتر و برق، دانشگاه تهران، تهران، ایران  
<sup>۲</sup> دانش آموخته دکتری مهندسی پزشکی، آزمایشگاه پردازش تصویر هوشمند، دانشکده مهندسی کامپیوتر و برق، دانشگاه تهران، تهران، ایران

<sup>۳</sup> استاد، آزمایشگاه پردازش تصویر هوشمند، دانشکده مهندسی کامپیوتر و برق، دانشگاه تهران، تهران، ایران  
<sup>۴</sup> دانشجوی دکتری مهندسی پزشکی، بخش مهندسی پزشکی، دانشگاه تمپل، فیلادلفیا، ایالات متحده آمریکا  
<sup>۵</sup> استاد، دانشکده تحصیلات تکمیلی علوم اطلاعات، موسسه علم و تکنولوژی NAIST، نارا، ژاپن

تاریخ ثبت در سامانه: ۱۳۹۷/۷/۲، بازنگری: ۱۳۹۷/۹/۲۲، پذیرش قطعی: ۱۳۹۷/۱۰/۱۷

### چکیده

آندوسکوپی به کمک کپسول بی‌سیم، یکی از روش‌های غیرتهاجمی آسیب‌شناسی مسیر گوارشی انسان به شمار آمده که با پیش‌روی کپسول در مسیر گوارشی و تصویربرداری از این مسیر، ویدئوی آندوسکوپی ضبط می‌شود. در عمل، حجم وسیعی از ویدئوهای آندوسکوپی با کیفیت‌های مختلف و زمان تصویربرداری طولانی، از مسیر گوارشی ثبت می‌شود که این زمان در حدود ۸ ساعت بوده و زمان بازنگری ویدئوی آندوسکوپی توسط پزشک متخصص و مجرب نیز در حدود ۲ ساعت و یا بیش‌تر می‌باشد. روش‌های خلاصه‌سازی ویدئو، می‌توانند زمان بازنگری و هم‌چنین خطاهای تفسیر و بازنگری دستی توسط پزشک را کاهش دهند. در این پژوهش، به منظور غلبه بر محدودیت‌های موجود در استخراج فریم‌های کلیدی، روشی مبتنی بر بینایی ماشین به کمک خصوصیات منحصر به فرد تجزیه به مقادیر تکین انطباقی مبتنی بر پنجره لغزشی، پیشنهاد شده است که با به کارگیری این ویژگی‌ها، امکان خلاصه‌سازی ویدئوی کپسول آندوسکوپی فراهم می‌گردد. علاوه بر این، به دلیل محدودیت‌های حجم و توان کپسول آندوسکوپی و فیزیک مسیر گوارشی، تصاویر کپسول دارای روشنایی پایینی بوده و غیریک‌نواخت می‌باشند که برای حل این مشکل، از فیلتر انتشار تطبیقی بهره گرفته شده است. نتایج آزمایش روی کپسول بی‌سیم، نشان می‌دهد که زمان بازنگری به طور قابل ملاحظه‌ای کاهش یافته است. ارزیابی‌های کمی و کیفی، نشان می‌دهند که استفاده از این روش پیشنهادی، تاثیر مثبتی روی فرایند خلاصه‌سازی داشته است. از این‌رو، می‌توان در سایر کاربردهای بالینی، از قبیل آموزش پزشکان جوان، آسیب‌شناسی به کمک کامپیوتر، تصمیم‌گیری سریع و مدیریت اطلاعات پزشکی نیز از این روش بهره گرفت.

کلیدواژه‌ها: انتشار کنتراست تطبیقی، تجزیه مقادیر تکین، پنجره لغزشی، خلاصه‌سازی ویدئو، کپسول بی‌سیم آندوسکوپی

\* نویسنده مسئول

نشانی: گروه بیوالکترونیک، دانشکده مهندسی برق و علوم کامپیوتر، پردیس دانشکده‌های فنی، دانشگاه تهران، تهران، ایران، کد پستی: ۱۴۳۹۹-۵۱۱۵۴

تلفن: ۶۱۱۱۴۳۰۲ (۲۱) ۹۸+

دورنگار: ۶۱۱۱۴۳۰۲ (۲۱) ۹۸+

پست الکترونیکی: zoroofi@ut.ac.ir

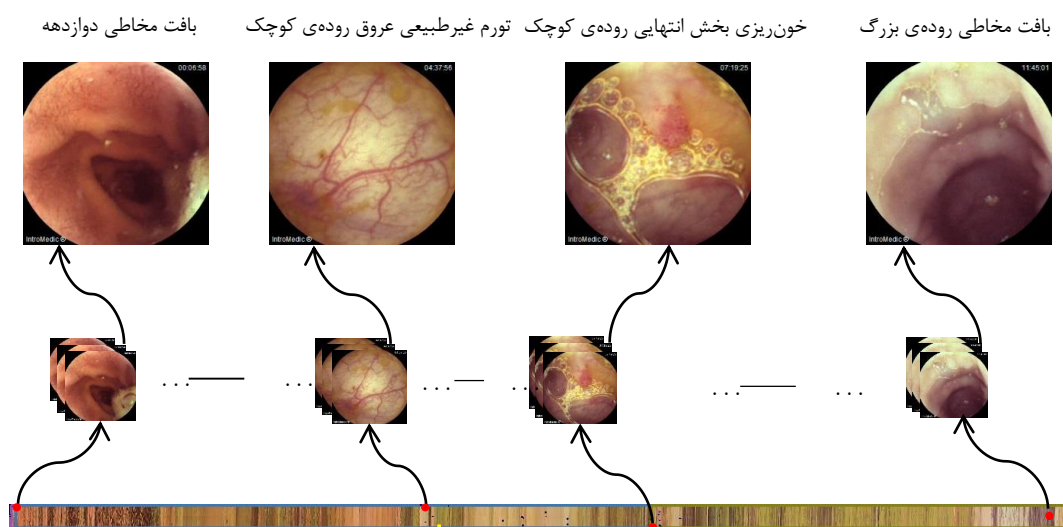
## ۱- مقدمه

امروزه، استفاده از روش‌هایی با حداقل تهاجم، در علم پزشکی از اهمیت ویژه‌ای برخوردار بوده که آندوسکوپی یکی از مهم‌ترین این روش‌ها به شمار می‌رود. در این میان، تصویربرداری آندوسکوپی به وسیله‌ی کپسول آندوسکوپی بی‌سیم<sup>۱</sup> (WCE)، به عنوان یک روش تشخیصی بسیار مناسب، امکان مشاهده‌ی بخش‌های داخلی بدن را فراهم ساخته و هم‌چنین در مواردی با به کارگیری ابزار مناسب در کنار آندوسکوپ، امکان نمونه‌برداری از بافت نیز میسر گشته که بسیار حایز اهمیت می‌باشد.

خلاصه‌سازی ویدئو، فرایند بسیار دشواری است، چرا که دستیابی به یک فرایند بی‌عیب و نقص برای انتخاب فریم‌های برجسته، به دانش معنایی نیاز دارد. در این جا باید اشاره شود که منظور از فریم‌های برجسته، فریم‌هایی با بیش‌ترین بازنمایی<sup>۲</sup> از صحنه<sup>۳</sup> می‌باشد. با انتخاب فریم‌های برجسته، فریم‌هایی که دارای محتوای تصویری کم‌تر و یا مشابه هم هستند، نادیده گرفته می‌شوند. در حالت کلی، هر ویدئو می‌تواند شامل تعدادی صحنه باشد که هر کدام از صحنه‌ها از ترکیب شات‌های<sup>۴</sup> ویدئویی ایجاد می‌شوند [۱]. از طرفی، هر شات با هر بازه‌ی زمانی، دارای فریم‌های متوالی بوده که معمولاً محتوای تصویری

یکسانی دارند و به مرور زمان و با جابه‌جایی کپسول در مسیر گوارشی، با محتوای جدیدی جای‌گزین می‌شوند. هدف از انجام پژوهش حاضر، انتخاب تصویری با بیش‌ترین بازنمایی از هر محتوا می‌باشد. با توجه به توضیحات فوق و به منظور شفافیت بیش‌تر موضوع، تعدادی از فریم‌های برجسته‌ی یک ویدئوی کپسول آندوسکوپی در شکل (۱) ارائه شده است.

لازم به ذکر است که خلاصه‌سازی تصاویر آندوسکوپی همواره با چالش‌های خاصی روبه‌رو است. نخست این که ویدئوهای کپسول آندوسکوپی، دارای محتوای بافت اندک و بسیار تغییرپذیر می‌باشد. هم‌چنین، به دلیل محدودیت‌های حجم و توان کپسول آندوسکوپی، تصاویر به دست آمده، غیریک‌نواخت، تاریک (دارای روشنایی پایین)، به شدت فشرده و به دلیل وجود تراشه‌ی سنسور دوربین<sup>۵</sup> (CMOS)، دارای نویز می‌باشند. از طرفی، علاوه بر حرکات تصادفی دوربین بی‌سیم، ساختار مسیر گوارشی نیز مستقل از وضعیت و فیزیک بدن، حرکات دودی و موضعی<sup>۶</sup> خاص خود را دارد. فراتر از آن، بعضی از ساختارها مانند حباب‌ها و مایع گوارشی نیز به شدت گمراه‌کننده هستند که مجموعه‌ی عوامل فوق‌الزمان بازنگری توسط پزشک را طولانی‌تر می‌کند [۲].



شکل (۱) - ارائه‌ی نمونه‌هایی از فریم‌های برجسته در یک ویدئوی کپسول آندوسکوپی بی‌سیم مربوط به پایگاه داده‌ی بیمارستان فیروزگر تهران

تحلیل دیداری یک ویدئوی آندوسکوپی با حدود ۶۰۰۰۰ فریم، بسته به مهارت پزشک، در حدود ۴۵ دقیقه تا چند ساعت می‌باشد. بنابراین، به کارگیری روش‌های محاسباتی در ویدئوهای آندوسکوپی، به کاهش زمان بازنگری و نیز کاهش تفسیرهای نادرست پزشک، کمک شایانی خواهد کرد [۴].

به دلیل طولانی بودن ویدئوی کپسول آندوسکوپی، تشخیص پزشک، همواره کاملاً صحیح و بی‌نقص نیست. به طوری که درصد تشخیص درست خون‌ریزی‌های خاص گوارشی، زخم‌ها، پلیپ‌ها و خون‌ریزی‌های شدید گوارشی، به ترتیب برابر با ۶۹، ۳۸، ۴۶ و ۱۷ درصد گزارش شده است [۳]. زمان لازم برای

<sup>۱</sup> Shot

<sup>۵</sup> Complementary Metal Oxide Semiconductor

<sup>۶</sup> Peristalsis

<sup>۱</sup> Wireless Capsule Endoscopy

<sup>۲</sup> Most Representative

<sup>۳</sup> Scene

ایسماعیل و هم‌کارانش، یک الگوریتم خلاصه‌سازی ویدئوی آندوسکوپی مبتنی بر یادگیری بدون سرپرست، به همراه تمایز ویژگی را ارائه کردند [۹]. آن‌ها برای پیدا کردن کلاس‌های مشابه و حذف نویز از ویدئوی کپسول آندوسکوپی، عضویت احتمالی فریم‌های ویدئو در کلاس‌ها را بر اساس توصیف کننده‌های دیداری و زمانی، محاسبه کردند. هر چند تلاش نویسندگان بر این است که تعداد کلاس‌های بهینه را یافته و اثر نویز را در فرایند کلاس‌بندی کاهش دهند، اما یافتن تعداد کلاس‌های بهینه، از نقاط ضعف این روش به حساب می‌آید.

محمد و هم‌کارانش، برای تقسیم‌بندی تصاویر کپسول آندوسکوپی روده‌ی کوچک به دو کلاس حاوی اطلاعات مفید و فاقد اطلاعات مفید، ابتدا ویدئوی آندوسکوپی را بر اساس تراکم طیف توان، به سوپرفریم‌ها کلاس‌بندی کرده و سپس ویژگی‌های سطح پایین و عمیق را برای این فریم‌های نماینده، استخراج کردند. در گام بعدی این روش، ویژگی‌های استخراج شده به کمک دیکشنری مبتنی بر یادگیری، کد گذاری تنک شده و به عنوان ورودی طبقه‌بند SVM<sup>۶</sup>، برای یادگیری دو کلاس مفید و غیرمفید به کار برده می‌شوند. صرف نظر از هزینه‌ی محاسباتی بالا، در این روش تمام فریم‌های مفید و غیرمفید، بدون در نظر گرفتن تعداد و تکرار فریم‌ها و تاثیر آن‌ها در زمان بازنگری، طبقه‌بندی می‌شوند [۲].

یاکوویدیس و هم‌کارانش، به منظور کاهش زمان بازنگری ویدئوی کپسول آندوسکوپی، از روش تصویرکاوی بدون سرپرست استفاده کردند. آن‌ها روشی مبتنی بر الگوریتم کاهش داده بدون سرپرست به کمک NFM<sup>۷</sup> را پیشنهاد کردند. در پژوهش آن‌ها، هر فریم به عنوان عضوی از فضای برداری در نظر گرفته می‌شود. در پایان، با محاسبه‌ی فاصله‌ی اقلیدسی و ژنودزیک<sup>۸</sup>، یک مجموعه از بردارهای عمود بر هم استخراج می‌شوند [۱۰]. محمود و هم‌کارانش، برای ارسال برخط فریم‌های موثر در آسیب‌شناسی آندوسکوپی و هیستروسکوپی، به جای ارسال کل فریم‌ها، به منظور صرفه‌جویی در هزینه‌ی ارسال و پهنای باند، روشی دیگر را پیشنهاد کردند [۱۱، ۱۲]. در روش پیشنهادی آن‌ها، فضای رنگ به COC<sup>۹</sup> تغییر کرده، سپس یک تصویر کامل<sup>۱۰</sup> برای پردازش‌های برخط استخراج شده، برای هر فریم، یک نقشه‌ی برجستگی<sup>۱۱</sup> از ترکیب

کاهش زمان ویدئو، از طریق تشخیص بخش‌های مرتبط و حذف بخش‌های غیرمرتبط صورت می‌گیرد. آیمال و هم‌کارانش، در پژوهشی به بررسی برخی از روش‌های خلاصه‌سازی در حوزه‌ی ویدئوهای غیرپزشکی پرداختند [۵]. هم‌چنین، آویلا و هم‌کارانش، الگوریتم<sup>۱</sup> VSUMM را برای استخراج فریم‌های کلیدی ویدئوهای غیرپزشکی ارائه کردند. پس از جداسازی فریم‌های ویدئویی با نرخ نمونه‌برداری مشخص، از هیستوگرام رنگ برای استخراج ویژگی‌ها بهره گرفته می‌شود. در مرحله‌ی بعد، ویژگی‌ها با استفاده از الگوریتم k-means، به شات‌های مورد نظر تقسیم‌بندی شده و سپس فریم‌های مشابه، از طریق تعیین آستانه، حذف می‌شوند. تعیین مقدار k، یکی از چالش‌های موجود در این روش به شمار می‌رود [۶].

ژانگ و هم‌کارانش، استفاده از الگوریتم<sup>۲</sup> CVAS را برای خلاصه‌سازی مبتنی بر آگاهی از زمینه، پیشنهاد کردند. آن‌ها فریم‌های کلیدی معنادار را از طریق کدگذاری تنک، با به کارگیری یک دیکشنری از گراف‌های هم‌بستگی ویژگی‌های زمانی-مکانی، به تصویر در آوردند. در این روش، بخش‌های مهم ویدئوی آزمایشی<sup>۳</sup>، استخراج می‌شود اما وجود تعدادی از فریم‌های اضافه در ویدئوی خلاصه‌سازی شده، به عنوان یکی از نقایص این روش به شمار می‌رود [۷].

ونگ و هم‌کارانش، برای استخراج فریم‌های کلیدی یک ویدئو، روش انتخاب دیکشنری را مورد استفاده قرار دادند. آن‌ها ویدئوی آندوسکوپی مربوط به بخش فوقانی دستگاه گوارش را مورد بررسی قرار دادند [۸]. در این روش، در مرحله‌ی پردازش کیفیت، فریم‌های مات، تاریک و اشباع، از طریق آموزش جنگل تصادفی، حذف می‌گردند، سپس فریم‌های ویدئوی آندوسکوپی، برش زده می‌شوند تا بخش‌های اضافی و تیره‌ی پیرامون آن حذف گردد. برای کاهش پیچیدگی محاسبات روش پیشنهادی، شات‌ها و مرز آن‌ها، به کمک ناحیه‌بندی زمانی، انتخاب می‌شود. هم‌چنین، ویژگی‌های مقاوم به چرخش، از قبیل هیستوگرام رنگ و هیستوگرام بافت<sup>۴</sup> LBP، برای پردازش‌های بعدی توسط مساله‌ی انتخاب دیکشنری، استخراج می‌شوند. این روش در آندوسکوپی بخش فوقانی دستگاه گوارش کاربرد داشته و ویدئوهایی با نور کافی، وضوح بالا<sup>۵</sup> و نرخ فریم زیاد در تصویربرداری را به همراه دارد.

<sup>۶</sup> Non-Negative Matrix Factorization

<sup>۸</sup> Geodesic Distance

<sup>۹</sup> Color-Opponent Coding

<sup>۱۰</sup> Integral-Image

<sup>۱۱</sup> Saliency Map

<sup>۱</sup> Video Summarization

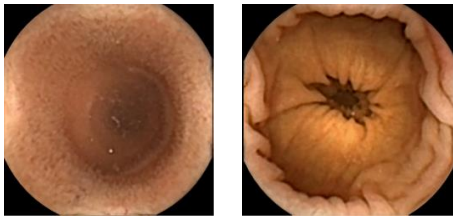
<sup>۲</sup> Context-Aware Surveillance Video Summarization

<sup>۳</sup> Test

<sup>۴</sup> Local Binary Pattern

<sup>۵</sup> High Definition

<sup>۶</sup> Support Vector Machine



شکل (۲) - نمونه‌ای از تصاویر پایگاه داده (KID) شکل راست: روده‌ی کوچک، شکل چپ: کولون

بر اساس نظر متخصصان گوارش، طراحی یک سیستم متمرکز برای تشخیص و نمایش خودکار فریم‌های ویدئویی، به لحاظ بالینی از اهمیت بسیار زیادی برخوردار است. با توجه به قابلیت بالای روش تجزیه به مقادیر تکین انطباقی، در این پژوهش، یک چارچوب موثر برای خلاصه‌سازی و نمایش ویدئوهای کپسول آندوسکوپی ارائه شده است. برای دستیابی به این هدف، هیستوگرام رنگ برای نمایش هر فریم به کار گرفته شده و ماتریس ویژگی، توسط پنجره‌ی لغزشی تجزیه به مقادیر تکین، نمایش داده می‌شود. با تخمین رتبه‌ی ماتریس ویژگی (تعداد سطرها و یا ستون‌های مستقل خطی در این ماتریس) و حذف مقادیر ویژه‌ی کم‌تر از آستانه، تنها فریم‌های کلیدی انتخاب می‌شوند. در فاز بعد، به کمک فیلتر انتشار کنتراست تطبیقی، غیریک‌نواختی روشنایی، بهبود داده می‌شود. این غیریک‌نواختی، به دلیل روشنایی اندک و محیط بسیار پیچیده‌ی مسیر گوارشی ایجاد می‌شود.

در ادامه‌ی این مقاله، در ابتدای بخش ۲، الگوریتم خلاصه‌سازی ویدئوهای آندوسکوپی شرح داده شده و سپس یک روش کاربردی برای اصلاح غیریک‌نواختی روشنایی در ویدئوی کپسول آندوسکوپی، به اختصار مورد بررسی قرار می‌گیرد. در بخش ۳، نتایج شبیه‌سازی‌ها و ارزیابی‌های آماری آزمایش‌های انجام شده، ارائه خواهد شد. در نهایت، در بخش ۴، به بحث و نتیجه‌گیری پرداخته خواهد شد.

## ۲- روش پیشنهادی

با توجه مباحث مطرح شده و اهمیت انتخاب فریم‌های برجسته در ویدئوی کپسول آندوسکوپی، در این مقاله با تلفیق رویکردهای مهندسی و ریاضی، از پنجره‌ی لغزشی مبتنی بر تجزیه به مقادیر تکین (SVD<sup>۳</sup>) برای انتخاب فریم‌های کلیدی استفاده می‌شود. بدین منظور، برای خلاصه‌سازی و بهبود کیفیت ویدئوی کپسول آندوسکوپی، بلوک دیگرام کلی روش

ویژگی‌های گشتاور، انحنا و کنتراست چندسطحی، تهیه شده و در مرحله‌ی آخر، از نقشه‌ی برجستگی برای استخراج فریم‌های کلیدی بهره گرفته می‌شود. در این مقاله، به منظور اعتبارسنجی روش پیشنهادی، از دو پژوهش اخیر بهره گرفته شده است.

با توجه به مباحث ذکر شده، روشن است که طراحی یک الگوریتم برای خلاصه‌سازی ویدئوی کپسول آندوسکوپی، با چالش‌های مختص خود روبه‌رو می‌باشد. عواملی مانند روشنایی غیریکنواخت، وضوح پایین، نرخ فشرده‌سازی بالا و نرخ متغیر تیزی<sup>۱</sup> کانونی تصویر، از جمله‌ی این چالش‌ها به شمار می‌آیند. علاوه بر این، ممکن است که فریم‌های حاوی ضایعه، دارای ساختارهای گمراه کننده‌ی مانند حباب‌ها، مایع گوارشی و ... نیز باشند. زمانی در حدود ۲ ساعت (یا بیش‌تر)، به عنوان زمان بازنگری در هر آزمایش کپسول آندوسکوپی، زمانی طولانی به حساب آمده و برای پزشک امری طاقت‌فرسا می‌باشد. با توجه به توضیحات فوق، یک فرایند کامپیوتری موثر می‌تواند از طریق استخراج فریم‌های برجسته، میزان کار پزشک در آسیب‌شناسی ویدئوی کپسول آندوسکوپی را کاهش دهد. هدف اصلی از انجام تحقیق حاضر، تلاش برای دستیابی به این مهم می‌باشد.

ذکر این نکته ضروری است که جهت انجام این پژوهش، از پایگاه داده‌ی محلی و پایگاه داده‌ی غیربومی استفاده شده است. پایگاه داده‌ی محلی، توسط بیمارستان فیروزگر تهران جمع‌آوری شده که فریم‌هایی از آن در شکل (۱) ارائه شده است. ویدئوهای این مجموعه، توسط کپسول آندوسکوپی بی‌سیم Miro-Cam، با وضوح  $320 \times 320$  پیکسل تهیه شده است. میزان روشنایی این مجموعه، ۲۵۶ سطح رنگ است که با فرمت ذخیره‌ی mec و نرخ نمونه‌برداری ۳ فریم بر ثانیه تصویربرداری می‌شود. ویدئوهای مربوط به ۳۰ بیمار از این پایگاه داده‌ی محلی، شامل ۱۲ زن و ۱۸ مرد، با متوسط سن و انحراف معیار  $54 \pm 15$ ، توسط یک متخصص گوارش به همراه یک دست‌یار، مورد تجزیه و تحلیل قرار گرفته است. علاوه بر این، به منظور ارزیابی بهتر روش پیشنهادی، از تصاویر کپسول آندوسکوپی پایگاه داده‌ی KID<sup>۲</sup> نیز استفاده شده است [۱۳، ۱۴]. این پایگاه داده، شامل تعداد زیادی ویدئوی کپسول آندوسکوپی بی‌سیم از نوع Pill-Cam، به همراه استاندارد طلایی با وضوح  $320 \times 320$  پیکسل می‌باشد. تمام ویدئوهای این پایگاه داده، بدون نام بوده و توسط یک تیم تحقیقاتی، مورد بررسی و تفسیر قرار گرفته است. نمونه‌هایی از تصاویر این پایگاه داده، در شکل (۲) ارائه شده است.

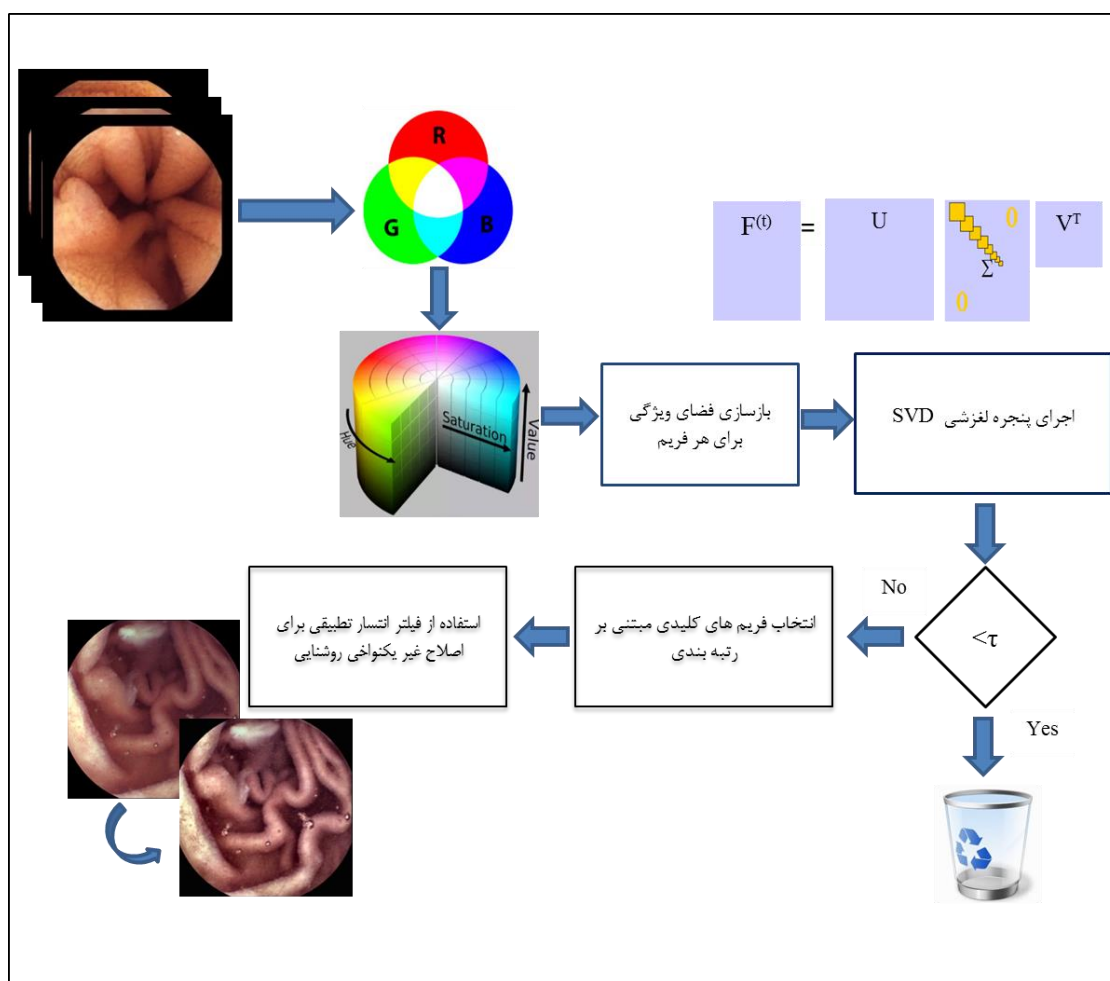
<sup>۳</sup> Singular Value Decomposition

<sup>۱</sup> Sharpness

<sup>۲</sup> Koulaouzidis-Iakovidis Database

تجزیه می‌شوند. در این مرحله، با محاسبه‌ی نوسانات رتبه و مقایسه‌ی آن‌ها با هم، فریم‌هایی که بازنمایی بیش‌تری از صحنه دارند، به عنوان فریم‌های برجسته انتخاب می‌شوند. پس از انتخاب فریم‌های کلیدی، غیریک‌نواختی روشنایی و کنتراست فریم‌های مورد نظر، به کمک فیلتر انتشار بهبود داده شده تا پزشک بتواند مشاهده و درک بهتری از صحنه‌ی مورد نظر داشته باشد.

پیشنهادی، در شکل (۳) ارائه شده است. همان‌طور که در این شکل مشاهده می‌شود، فریم‌های ورودی آندوسکوپی، پس از تغییر فضای رنگ از RGB به HSV، در قالب بردارهای ویژگی بازسازی می‌شوند. سپس با ترکیب این بردارهای مربوط به فریم‌های ویدئوی کپسول بی‌سیم آندوسکوپی، ماتریس‌های ویژگی ایجاد می‌شوند. در مرحله‌ی بعد، مقادیر تکین ماتریس‌های ویژگی، که در قالب یک پنجره‌ی زمانی هستند،



شکل (۳) - بلوک دیاگرام روش پیشنهادی

از RGB (که با سیستم ماشین سازگارتر است) به HSV تغییر داده می‌شود. فضای رنگ HSV با سیستم بینایی و سلول‌های مخروطی و میله‌ای چشم انسان، سازگاری بیش‌تری دارد [۶]. محاسبه‌ی هیستوگرام در فضای رنگ جدید، تاثیر زیادی روی هزینه‌ی محاسبات داشته و قابلیت انعطاف‌پذیری الگوریتم برای ویدئوهای آندوسکوپی طولانی‌تر را فراهم می‌کند. در این حالت، فضای HSV را برای مثال می‌توان به ۱۶ زیر مجموعه، کوانتیزه نمود. بنابراین، با وجود کاهش قابل توجهی از حجم داده‌ها، اطلاعات کم‌تری از دست می‌رود. به کارگیری هیستوگرام در یافتن تفاوت‌های کلی بین فریم‌های مجزا، مفید می‌باشد [۱۵].

## ۲-۱- معادلات ریاضی تجزیه به مقادیر تکین SVD

همان‌طور که گفته شد، فریم‌های کلیدی، فریم‌هایی با بیش‌ترین بازنمایی از صحنه بوده و به عنوان معادلی از تصویر در فضای ابعاد بالا به زیر مجموعه‌ای با ابعاد کم‌تر می‌باشند. زیر مجموعه‌ی فریم‌های کلیدی  $S$  از ویدئوی  $V$  با  $k$  فریم، به صورت زیر نمایش داده می‌شود:

$$S = P_{Keyframe}(V) = \{fr_1, fr_2, \dots, fr_k\} \quad (1)$$

که در آن،  $P_{keyframe}$  به فرایند استخراج فریم‌ها اشاره دارد. پس از جداسازی فریم‌ها، برای انجام پردازش‌های بعدی، فضای رنگ

## ۲-۲- توصیف آماری داده‌ها با پنجره‌ی لغزشی SVD

در ادامه، فرض کنید  $A^{(t)}$  فریم ویدئویی در لحظه‌ی  $t$  در فضای رنگ HSV باشد که متناظر با آن، سه هیستوگرام  $h_H$ ،  $h_S$  و  $h_V$  توصیف می‌شوند. در این صورت، بردار ویژگی  $f^{(t)}$  با طول  $L=L_H+L_S+L_V$  به صورت زیر ایجاد می‌شود:

$$f^{(t)} = [h_H \ h_S \ h_V] \quad (۴)$$

برای هر فریم ویدئوی اندوسکوپی، در زمان  $t > N$  ماتریس ویژگی  $F^{(t)}$  ( $t=N, N+1, \dots, T$ ) به صورت زیر خواهد بود:

$$F^{(t)} = \begin{bmatrix} f^{(t)} \\ f^{(t-1)} \\ \vdots \\ f^{(t-N+1)} \end{bmatrix} \quad (۵)$$

$T$  و  $N$  به ترتیب عرض پنجره و تعداد کل فریم‌های ویدئویی می‌باشند. بردار  $F^{(t)}$  معرف ویژگی فریم فعلی در زمان  $t$  و تعداد  $N$  فریم قبلی است [۱۷].

تجزیه به مقادیر تکین، روش رایجی برای استخراج ویژگی از فضایی با ابعاد بالا به فضایی با ابعاد پایین‌تر می‌باشد [۱۸]. بردار ویژگی  $F^{(t)}$  توسط SVD به مقادیر ویژه تجزیه می‌شود.

$$F^{(t)} = U \Sigma V^T \quad (۶)$$

$\Sigma$  ماتریس مقادیر تکین  $F^{(t)}$ ، یک ماتریس قطری است که مقادیر آن به صورت غیرصعودی مرتب شده‌اند.  $U$  و  $V$ ، ماتریس‌های مثلثی در محدوده‌ی ابعاد سطر و ستون  $F^{(t)}$  هستند. در این پژوهش، برای محاسبه‌ی سریع‌تر SVD، از کتابخانه‌ی استاندارد LAPACK در حل مسایل جبر عددی بهره گرفته شده است [۱۹].

به کارگیری رویکرد مطرح شده، به جای خلاصه‌سازی مبتنی بر مرز میان‌شات‌ها، بیش‌تر در بر گیرنده‌ی خلاصه‌سازی مبتنی بر محتوا می‌باشد. فریم‌های با تغییرات دیداری بیش‌تر، به نقاطی دورتر از مبدا در فضای ویژگی تصویر می‌شوند. در صورتی که مقادیر ویژه‌ی تکین کم‌تر از آستانه، حذف گردند، محورهای نیمه وابسته و کم اهمیت در فضای ویژگی، به صورت خطی حذف می‌شوند.

فرض کنید  $s_1, s_2, \dots, s_i, \dots, s_N$  مقادیر تکین یک ماتریس باشند. رتبه‌ی ماتریس  $F^{(t)}$ ، برابر با تعداد مقادیر ویژه‌ی تعریف

به طور کلی، برای خلاصه‌سازی ویدئوها، با نرخ کم‌تری از آن‌ها نمونه‌برداری می‌شود. برای هر ماتریس ویژگی  $A=[A_1 \ A_2 \ \dots \ A_n]$  و فریم  $i$  در مجموعه‌ی نمونه‌برداری شده‌ی  $n$ -تایی، بردار  $m$  بعدی  $A_i$  ایجاد می‌شود ( $n \leq m$ ). در این پژوهش، برای تشکیل  $A_i$  از هیستوگرام رنگ HSV استفاده می‌شود. برای ادغام اطلاعات مکانی، می‌توان هر فریم را به بلوک‌های  $3 \times 3$  تقسیم کرد تا از هر بلوک بردارهای ویژگی مجزا استخراج گردد. معمولاً از آن‌جا که تعداد کمی از میله‌های هیستوگرام انتخاب شده و هر بلوک شامل تمام رنگ‌ها نیست، ماتریس ویژگی، تنگ خواهد بود. الگوریتم SVD روی این ماتریس، در مقایسه با SVD معمولی روی یک فریم کامل، سریع‌تر اجرا شده و حافظه‌ی کم‌تری نیز مصرف می‌کند. در واقع، SVD، ماتریس ویژگی  $A_{m \times n}$  را از فضای  $m$  بعدی به فضای بردارهای ویژه‌ی  $T$  بعدی ( $T < m$ ) نگاشت می‌دهد [۱۶]:

$$A = U \Sigma V^T \quad (۲)$$

در این رابطه،  $U=[u_{ij}]$ ، یک ماتریس متعامد<sup>۱</sup> ستونی است که هر ستون آن معادل بردار ویژه‌ی سمت چپی می‌باشد.  $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$  یک ماتریس قطری  $n \times n$  است که عناصر آن، مقادیر ویژه‌ی غیرصفر بوده که به صورت نزولی مرتب شده‌اند.  $V=[v_{ij}]$  یک ماتریس متعامد  $n \times n$  است که بردار مقادیر ویژه‌ی چپ می‌باشد. با فرض برقراری رابطه‌ی  $\text{rank}(A)=r$ ،  $\Sigma$  شرایط زیر را ارضا خواهد کرد:

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq \sigma_{r+1} = \dots = \sigma_n = 0 \quad (۳)$$

SVD، ماتریس ویژگی  $A_{m \times n}$  را از فضای  $m$ -بعدی داده‌ی خام به فضای بردارهای ویژه‌ی  $r$ -بعدی با محورهای مستقل خطی نگاشت می‌دهد. این نگاشت، بردار  $A_i=[a_{i1} \ a_{i2} \ \dots \ a_{im}]^T$  از ماتریس  $A$  (که شامل هیستوگرام‌های سرهم‌بندی فریم  $i$  است) را روی بردار ستونی  $\psi_i=[v_{i1} \ v_{i2} \ \dots \ v_{ir}]^T$  از ماتریس  $V^T$  تصویر کرده و هر سطر  $j$  از ماتریس  $A$  (که  $j$  ورودی هیستوگرام سرهم‌بندی شده‌ی تعداد رویدادها در هر فریم ویدئویی است) را به بردار سطری  $\phi_j=[u_{j1} \ u_{j2} \ \dots \ u_{jr}]^T$  از ماتریس  $U$  نگاشت می‌دهد. اجرای SVD، منوط به بزرگ‌تر بودن  $m$  از  $n$  بوده و اگر تعداد فریم‌ها ( $n$ ) از تعداد عناصر سرهم‌بندی شده در هر هیستوگرام ( $m$ ) بیش‌تر باشد، SVD روی  $A^T$  انجام شده و نقش  $U$  و  $V$  نیز تغییر می‌کند.

<sup>۱</sup> Orthonormal

ورودی‌های ماتریس  $L$  به صورت  $[L_{pq}]_{xx}$ ،  $[L_{pq}]_{yy}$ ،  $[L_{pq}]_{xy}$  و  $[L_{pq}]_{yx}$  مشخص می‌شود. بنابراین، گرادیان‌ها به صورت زیر می‌باشند:

$$\frac{L^{n+1} - L^n}{\Delta t} + \frac{L^{n+1}}{\tau} = \frac{F^n}{\tau}$$

$$\frac{u_{pq}^{n+1} - u_{pq}^n}{\Delta t} - \frac{\Delta_x^x ([L_{pq}]_{xx}^n \Delta_x^x u_{pq}^n + [L_{pq}]_{xy}^n \Delta_y^y u_{pq}^n)}{h^2}$$

$$- \frac{\Delta_y^y ([L_{pq}]_{yx}^n \Delta_x^x u_{pq}^n + [L_{pq}]_{yy}^n \Delta_y^y u_{pq}^n)}{h^2} = 0 \quad (9)$$

$\tau$ ، فاکتور مقیاس زمانی است.  $F$ ، از انتشار در طول مرزهای مشخص تصویر، جلوگیری کرده و به صورت زیر تعریف می‌شود [۲۳]:

$$F_s(\nabla u) = \begin{cases} P_{\nabla u^\perp} |\nabla u| \geq s \\ \frac{3}{2} \left( 1 - \frac{|\nabla u|^2}{s^2} \right) + \frac{|\nabla u|^2}{s^2} P_{\nabla u^\perp} \quad o.w. \end{cases} \quad (10)$$

$P_{\nabla u^\perp}$ ، افکنش<sup>۴</sup> در جهت عمود بر گرادیان تصویر است و  $s$  پارامتر آستانه‌ی ثابت می‌باشد. حالت گذرای این مدل، شامل تصاویری با الگوهای همگن و پارامتر آزادی  $\tau$ ، برای تعیین مقیاس الگوها می‌باشد. در مرحله‌ی بعد، ماتریس انتشار به صورت زیر بازسازی می‌شود:

$$L_{pq}^{n+1} = \left( \frac{1}{1 + \alpha} \right) (\alpha L_{pq}^n + F_{pq}^n) \quad (11)$$

که  $\alpha = \frac{\tau}{\Delta t}$  و  $F_{pq}$  توسط رابطه‌ی (۱۰) تعیین شده و این امر الگوی واضح شرایط مثبت ماتریس  $L$  را حفظ می‌کند. این الگوی تکراری به هر سه کانال  $R$ ،  $G$  و  $B$  اعمال شده و تصویر تصحیح شده از طریق بازسازی این کانال‌ها ایجاد می‌شود.

### ۳- یافته‌ها و بحث

#### ۳-۱- اعتبارسنجی روش

نتایج یک خلاصه‌سازی موثر توسط الگوریتم خودکار، مطابقت زیادی با نتایج خلاصه‌سازی توسط انسان دارد. در این‌جا به منظور ارزیابی عمل‌کرد الگوریتم پیشنهادی برای استخراج فریم‌های کلیدی ویدئوی کپسول آندوسکوپی، از معیارهای

می‌شود که مقدار آن‌ها از یک آستانه‌ی مشخص  $\lambda$  بیش‌تر باشند ( $s_i/s_{\max} \geq \lambda$ ). در حقیقت، به کمک SVD بریده شده<sup>۱</sup>، ساختارهای مهم و اساسی باقی مانده و نوسانات جزئی و نویزها حذف می‌گردند [۲۰].

در بررسی‌های انجام شده روی ویدئو، مشخص شده است که اگر رتبه‌ی ماتریس ویژگی فریم فعلی ( $F^{(t)}$ ) از رتبه‌ی ماتریس ویژگی فریم قبلی ( $F^{(t-1)}$ ) بزرگ‌تر باشد، محتوای دیداری فریم فعلی با فریم قبلی، به طور قابل ملاحظه‌ای تفاوت خواهد داشت، از این‌رو، نمی‌توان آن را با مقادیر تکین بردارهای لحظه‌ی  $t-1$ ، بازنمایی کرد. علاوه بر این، اگر رتبه‌ی ماتریس ویژگی فعلی از ماتریس قبلی کم‌تر باشد، محتوای دیداری رشته‌ی ویدئو به اندازه‌ی است که می‌توان شات قبلی را نادیده گرفت [۲۱]. برای مقارنه‌ی اولیه‌ی الگوریتم، از  $N$  فریم اول برای محاسبه‌ی ماتریس ویژگی در لحظه‌ی  $t=N$  استفاده می‌شود. بنابراین حلقه‌ی اصلی الگوریتم، از لحظه‌ی  $t=N+1$  شروع خواهد شد.

#### ۲-۳- اصلاح غیریک‌نواختی روشنایی و بهبود تصویر

همان‌طور که در بخش مقدمه مطرح شد، غیریک‌نواختی روشنایی در تصاویر آندوسکوپی، از جمله مواردی است که تشخیص پزشک را تحت تاثیر قرار می‌دهد. در این پژوهش، برای اصلاح این غیریک‌نواختی، از فیلتر انتشار کنتراست تطبیقی<sup>۲</sup> استفاده شده است [۲۲]. در معادله‌ی اصلی انتشار، مرزها حفظ شده و محتوا، نرم می‌گردد، اما در مورد ویدئوهای کپسول آندوسکوپی، به دلیل زمینه‌ی متفاوت و پیچیده، باید از پارامتر نرم‌سازی دیگری استفاده شود. به همین دلیل، معادله‌ی انتشار تطبیقی روی فضای رنگ توسعه داده شده و به صورت زیر ارائه می‌شود:

$$\frac{\partial u(x, y)}{\partial t} - \text{div}(L \nabla u) = 0 \quad (7)$$

که در آن،  $u$  تصویر ورودی،  $x$  و  $y$ ، مختصات تصویر،  $t$  زمان انتشار و  $L$  ماتریس انتشار ناهم‌سان‌گرد<sup>۳</sup> می‌باشد. فرض کنید  $u_{pq}$  مقدار  $u$  در پیکسل ( $x=ph$  و  $y=qh$ ) باشد، آن‌گاه، اختلاف جزئی در راستای  $x$  و  $y$ ، به صورت زیر تعریف می‌شود:

$$\begin{aligned} \Delta_x^+ u_{p,q} &= u_{p+1,q} - u_{p,q} \\ \Delta_x^- u_{p,q} &= u_{p,q} - u_{p-1,q} \\ \Delta_y^+ u_{p,q} &= u_{p,q+1} - u_{p,q} \\ \Delta_y^- u_{p,q} &= u_{p,q} - u_{p,q-1} \end{aligned} \quad (8)$$

<sup>۱</sup> Anisotropic Diffusion Matrix

<sup>۲</sup> Projection

<sup>۳</sup> Truncated SVD

<sup>۴</sup> Adaptive Contrast Diffusion Filter



پیشنهادی در تحقیق حاضر، مبتنی بر رویکرد پنجره‌ی لغزشی مقادیر تکین ویژه می‌باشد. با بهره‌گیری از پنجره‌ی لغزشی SVD، رتبه‌ی ماتریس ویژگی فعلی محاسبه شده و فریم‌های نماینده با بیش‌ترین بازنمایی، از طریق بررسی نوسانات رتبه در طول زمان، تشخیص داده می‌شوند.

نتایج شبیه‌سازی‌ها حاکی از آن است که الگوریتم پیشنهادی، عمل کرد بسیار موثری در خلاصه‌سازی داشته است. از آن‌جا که این الگوریتم، برای پردازش، به تمام ویدئو نیاز ندارد، می‌توان برای پردازش آنلاین ویدئو نیز از آن بهره گرفت. نتایج آزمایشی نشان می‌دهد که این الگوریتم در مقابل غیریک‌نواختی روشنایی، مقاوم است. الگوریتم پیشنهادی توسط یک کامپیوتر شخصی، با پردازنده‌ی P4 (۳ GHz) و حافظه‌ی دسترسی ۸ GB و در محیط نرم‌افزار متلب ۸.۱ پیاده‌سازی شده است. میانگین زمان انجام پردازش برای هر بیمار، در حدود ۵ دقیقه بوده که نشان دهنده‌ی مقرون به صرفه بودن الگوریتم خلاصه‌سازی به لحاظ محاسباتی می‌باشد. با توجه به مطالب مطرح شده در بخش اعتبارسنجی، نتایج خلاصه‌سازی توسط پارامترهای صحت، دقت و معیار  $F$ ، ارزیابی می‌شوند که نتایج مربوط به پایگاه داده‌ی KID، در جدول (۱) ارائه شده است.

همان‌طور که نتایج کمی ارائه شده در جدول (۱) نشان می‌دهد، متخصصان گوارش می‌توانند با جستجو در ویدئوی کپسول آندوسکوپی، بدون نیاز به بررسی مجزای هر فریم از ویدئو، مدت زمان لازم برای بررسی و تشخیص را به مقدار زیادی کاهش دهند. پس از مشاهده‌ی فریم‌های استخراج شده به کمک روش پیشنهادی، در صورتی که بخشی از ویدئو نیاز به بازنگری مجدد داشته باشد، پزشک متخصص می‌تواند تنها همان بخش مورد نظر را مجدداً از روی ویدئوی کپسول بی‌سیم آندوسکوپی، بررسی نماید. بررسی مسایل آسیب‌شناسی توسط روش پیشنهادی، برای پزشک بسیار مقرون به صرفه می‌باشد. نتایج آزمایش روی پایگاه داده‌ی KID، با مد نظر قرار دادن پارامترهای صحت، دقت و معیار  $F$ ، عمل کرد مناسب روش پیشنهادی را نشان می‌دهد. برای بررسی بهتر عمل کرد روش پیشنهادی، از داده‌های محلی نیز استفاده شده است. نتایج کمی پنجره‌ی لغزشی مبتنی بر تجزیه به نقاط تکین، در ویدئوهای محلی کپسول بی‌سیم آندوسکوپی، در جدول (۲) ارائه شده است. همان‌طور که مشاهده می‌شود، به کارگیری روش پیشنهادی پنجره‌ی لغزشی تجزیه به مقادیر تکین، روی داده‌های محلی نیز عمل کرد مناسبی داشته است.

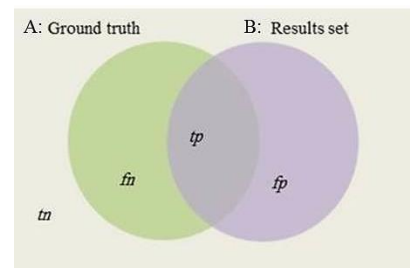
ارزیابی صحت<sup>۱</sup>، دقت<sup>۲</sup> و مشابهت<sup>۳</sup> ( $F$ ) بهره گرفته شده است. معیار  $F$ ، در قالب میانگین هندسی صحت و دقت تعریف می‌شود:

$$Recall = \frac{tp}{tp + fn} = \frac{m}{|k_1|} \quad (12)$$

$$Precision = \frac{tp}{tp + fp} = \frac{m}{|k_2|} \quad (13)$$

$$F - measure = \frac{2(Recall \times Precision)}{Recall + Precision} \quad (14)$$

همان‌طور که در شکل (۴) مشاهده می‌شود،  $tp$  برابر با تعداد فریم‌هایی است که متعلق به کلاس A بوده و به درستی طبقه‌بندی شده‌اند.  $tn$  برابر با تعداد فریم‌هایی است که متعلق به کلاس A نبوده و به درستی طبقه‌بندی شده‌اند.  $fp$  برابر با تعداد فریم‌هایی است که متعلق به کلاس A نبوده اما به اشتباه متعلق به کلاس A در نظر گرفته شده‌اند.  $fn$  برابر با تعداد فریم‌هایی است که متعلق به کلاس A بوده اما به اشتباه به عنوان جزئی از این کلاس طبقه‌بندی نشده‌اند. به طور کلی، می‌توان نتیجه گرفت که  $N = tp + tn + fp + fn$  برابر با تعداد کل فریم‌ها می‌باشد [۲۴].



شکل (۴) - هم‌پوشانی بین استاندارد طلایی و نتایج الگوریتم [۲۴]

مقادیر بیش‌تر معیار  $F$  و نرخ فشرده‌سازی، عمل کرد بهتر الگوریتم پیشنهادی را نشان می‌دهند که نسبت به دو معیار قبلی (نرخ درستی و نرخ خطا)، دارای تقارن بیش‌تری بوده، محدود به بازه‌ی ۰ تا ۱ هستند و به صورت بهتری قابل درک و تفسیر شدن می‌باشند.

## ۲-۲- نتایج شبیه‌سازی

همان‌طور که بیان شد، بازنگری تصاویر در ویدئوی کپسول آندوسکوپی، امری پیچیده و چالش برانگیز می‌باشد. پزشکان متخصص، که به صورت مداوم ویدئوهای آندوسکوپی را بررسی می‌کنند، اظهار کرده‌اند که یافتن فریم‌های نماینده با بیش‌ترین بازنمایی، فرایندی پیچیده، وقت‌گیر و طاقت‌فرسا است. روش

<sup>۱</sup> F-Measure

<sup>۱</sup> Precision

<sup>۲</sup> Recall

جدول (۱) - ارزیابی کمی عمل‌کرد روش پیشنهادی، شامل صحت ( $P$ )، دقت ( $R$ ) و معیار  $F$ ، روی پایگاه داده‌ی KID.

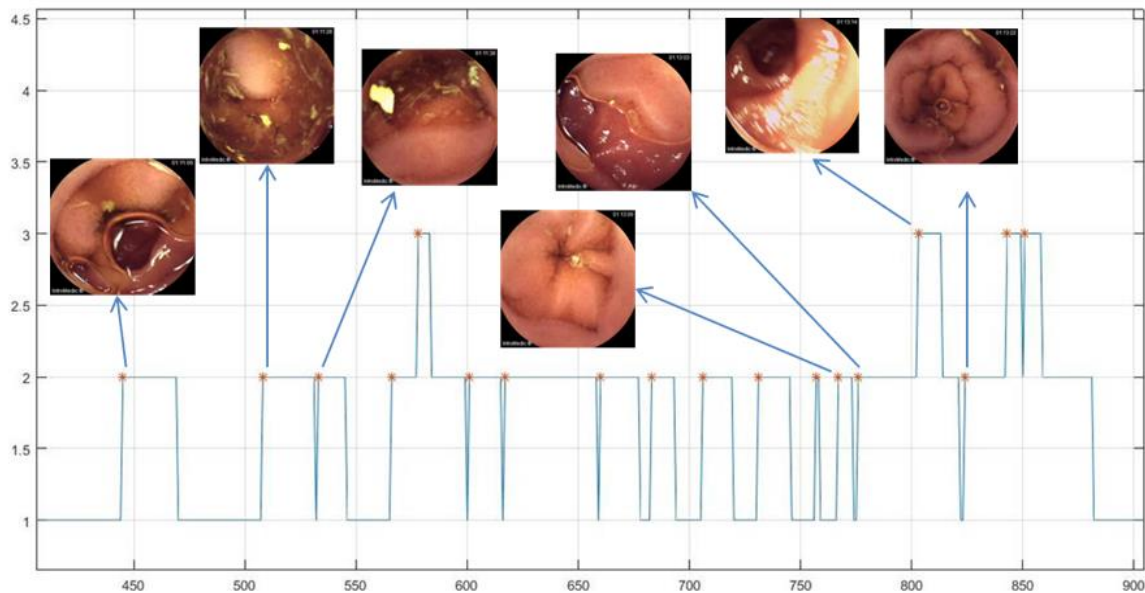
دیف	Iakovidis و هم‌کارانش			Mahmud و هم‌کارانش			روش پیشنهادی		
	$F$	$P$	$R$	$F$	$P$	$R$	$F$	$P$	$R$
۱	۰/۷۸	۰/۶۷	۰/۷۲	۰/۸۴	۰/۸۰	۰/۸۲	۰/۸۴	۰/۸۵	۰/۸۴
۲	۰/۷۰	۰/۷۱	۰/۷۰	۰/۸۳	۰/۸۲	۰/۸۲	۰/۸۵	۰/۸۶	۰/۸۵
۳	۰/۷۲	۰/۵۸	۰/۶۴	۰/۸۲	۰/۷۹	۰/۸۰	۰/۸۴	۰/۸۳	۰/۸۴
۴	۰/۷۱	۰/۷۲	۰/۷۱	۰/۷۵	۰/۷۹	۰/۷۷	۰/۸۴	۰/۸۷	۰/۸۴
۵	۰/۶۹	۰/۵۰	۰/۵۸	۰/۸۱	۰/۸۰	۰/۸۰	۰/۸۲	۰/۸۳	۰/۸۲
۶	۰/۶۵	۰/۷۲	۰/۶۹	۰/۷۵	۰/۸۲	۰/۷۸	۰/۸۰	۰/۸۳	۰/۸۰
۷	۰/۷۹	۰/۸۱	۰/۸۰	۰/۸۲	۰/۸۳	۰/۸۲	۰/۸۵	۰/۸۶	۰/۸۵
۸	۰/۷۴	۰/۷۹	۰/۷۶	۰/۷۲	۰/۷۸	۰/۷۵	۰/۷۸	۰/۷۷	۰/۷۸
۹	۰/۸۰	۰/۷۳	۰/۷۶	۰/۸۴	۰/۸۱	۰/۸۲	۰/۸۵	۰/۸۴	۰/۸۵
۱۰	۰/۷۸	۰/۶۷	۰/۷۲	۰/۸۸	۰/۸۵	۰/۷۲	۰/۸۷	۰/۸۷	۰/۸۷
۱۱	۰/۶۶	۰/۷۰	۰/۶۸	۰/۸۰	۰/۷۵	۰/۷۷	۰/۸۲	۰/۸۱	۰/۸۲
۱۲	۰/۷۲	۰/۵۹	۰/۶۵	۰/۷۵	۰/۷۷	۰/۷۶	۰/۷۹	۰/۸۰	۰/۷۹
۱۳	۰/۷۲	۰/۸۱	۰/۷۶	۰/۷۰	۰/۷۴	۰/۷۲	۰/۷۸	۰/۸۱	۰/۷۸
۱۴	۰/۷۱	۰/۷۴	۰/۷۲	۰/۷۶	۰/۷۷	۰/۷۶	۰/۷۸	۰/۷۹	۰/۷۸
Avg	۰/۷۳	۰/۷۰	۰/۷۱	۰/۷۹	۰/۷۷	۰/۷۸	۰/۸۲	۰/۸۳	۰/۸۲

جدول (۲) - ارزیابی کمی عمل‌کرد روش پیشنهادی، شامل صحت ( $P$ )، دقت ( $R$ ) و معیار  $F$ ، روی پایگاه داده‌ی محلی

ردیف	Iakovidis و هم‌کارانش			Mahmud و هم‌کارانش			روش پیشنهادی		
	$F$	$P$	$R$	$F$	$P$	$R$	$F$	$P$	$R$
۱	۰/۷۲	۰/۷۵	۰/۷۳	۰/۷۷	۰/۷۶	۰/۷۶	۰/۸۱	۰/۸۳	۰/۸۱
۲	۰/۶۸	۰/۷۳	۰/۷۰	۰/۷۲	۰/۷۶	۰/۷۴	۰/۸۳	۰/۸۴	۰/۸۳
۳	۰/۷۱	۰/۷۴	۰/۷۲	۰/۷۹	۰/۷۷	۰/۷۸	۰/۸۶	۰/۸۷	۰/۸۶
۴	۰/۷۴	۰/۷۵	۰/۷۴	۰/۷۶	۰/۷۹	۰/۷۷	۰/۸۵	۰/۸۴	۰/۸۵
۵	۰/۷۷	۰/۷۶	۰/۷۶	۰/۷۹	۰/۷۴	۰/۷۶	۰/۸۰	۰/۸۲	۰/۸۰
۶	۰/۶۵	۰/۷۲	۰/۶۸	۰/۸۱	۰/۸۴	۰/۸۲	۰/۸۴	۰/۸۵	۰/۸۴
۷	۰/۷۳	۰/۷۷	۰/۷۵	۰/۷۷	۰/۷۹	۰/۷۸	۰/۷۹	۰/۸۲	۰/۷۹
۸	۰/۸۰	۰/۷۴	۰/۷۷	۰/۷۳	۰/۷۵	۰/۷۴	۰/۸۵	۰/۸۷	۰/۸۵
۹	۰/۷۵	۰/۷۹	۰/۷۷	۰/۷۶	۰/۷۷	۰/۷۶	۰/۸۶	۰/۸۷	۰/۸۶
۱۰	۰/۷۱	۰/۷۵	۰/۷۹	۰/۷۳	۰/۷۹	۰/۷۶	۰/۷۸	۰/۷۸	۰/۷۸
۱۱	۰/۷۲	۰/۷۱	۰/۷۱	۰/۷۴	۰/۷۶	۰/۷۵	۰/۸۲	۰/۸۴	۰/۸۲
۱۲	۰/۶۹	۰/۷۲	۰/۷۰	۰/۷۷	۰/۷۶	۰/۷۶	۰/۷۹	۰/۸۰	۰/۷۹
۱۳	۰/۷۶	۰/۸۱	۰/۷۸	۰/۷۹	۰/۷۴	۰/۷۷	۰/۸۲	۰/۸۳	۰/۸۲
۱۴	۰/۷۱	۰/۷۲	۰/۷۱	۰/۷۴	۰/۷۸	۰/۷۵	۰/۸۵	۰/۸۷	۰/۸۵
۱۵	۰/۷۸	۰/۷۵	۰/۷۶	۰/۸۱	۰/۷۹	۰/۸۰	۰/۸۴	۰/۸۲	۰/۸۴
۱۶	۰/۷۲	۰/۷۴	۰/۷۳	۰/۷۵	۰/۷۷	۰/۷۶	۰/۸۱	۰/۸۳	۰/۸۱
۱۷	۰/۷۵	۰/۶۹	۰/۷۲	۰/۷۶	۰/۷۲	۰/۷۴	۰/۸۳	۰/۸۰	۰/۸۳
۱۸	۰/۷۳	۰/۷۵	۰/۷۴	۰/۷۲	۰/۷۴	۰/۷۳	۰/۸۵	۰/۸۵	۰/۸۵
۱۹	۰/۷۹	۰/۷۱	۰/۷۵	۰/۸۱	۰/۷۶	۰/۷۸	۰/۸۷	۰/۸۶	۰/۸۷
۲۰	۰/۷۲	۰/۶۹	۰/۷۰	۰/۷۹	۰/۷۸	۰/۷۸	۰/۸۳	۰/۸۰	۰/۸۳
Avg	۰/۷۳	۰/۷۴	۰/۷۳	۰/۷۶	۰/۷۷	۰/۷۶	۰/۸۳	۰/۸۳	۰/۸۳

صحنه می‌باشد. نکته‌ی قابل توجه این است که به منظور بررسی دقیق‌تر و تشخیص بهتر توسط پزشک، استخراج حداقل یک فریم برای هر صحنه ضروری می‌باشد. زیرا علی‌رغم تشابه نسبی فریم‌ها در کپسول بی‌سیم، موقعیت کپسول در گذر زمان به طور تدریجی در طول روده‌ی کوچک تغییر می‌کند. همان‌طور که اشاره شد، با تعیین رتبه‌ی مناسب برای پنجره‌ی لغزشی در هر تکرار، با توجه به نوسانات رتبه، می‌توان فریم‌هایی با بیش‌ترین بازنمایی در ویدئوی آندوسکوپی را استخراج نمود. بدین منظور، برای بررسی بهتر عمل‌کرد روش پیشنهادی، نمونه‌هایی از فریم‌های استخراج شده با توجه به نوسانات رتبه‌ی پنجره‌ی لغزشی، در شکل (۵) ارائه شده است.

با توجه به نتایج به دست آمده، می‌توان بیان کرد که در اغلب ویدئوها، خلاصه‌سازی خودکار، همبستگی قابل قبولی با خلاصه‌سازی‌های دستی داشته است. یکی از مزایای خلاصه‌سازی با پنجره‌ی لغزشی SVD، این است که نیازی به مشخص کردن تعداد فریم‌ها وجود نداشته و تعداد فریم‌های مورد نیاز، به صورت خودکار انتخاب می‌شوند. روش‌های پیچیده‌تر، ممکن است دقت بالاتری داشته باشند، اما با توجه به هزینه‌های بالای محاسباتی، استفاده از آن‌ها به صورت برخط در زمان انجام آزمایش کپسول آندوسکوپی، توصیه نمی‌شود. یکی از ویژگی‌های منحصر به فرد کپسول آندوسکوپی بی‌سیم، این است که کل سری زمانی تصویر ضبط شده، تنها به عنوان یک شات ویدئویی محسوب می‌شود، اما دارای تعداد زیادی



شکل (۵) - استخراج فریم‌هایی با بیش‌ترین بازنمایی، با توجه به عمل‌کرد روش پیشنهادی و مقایسه‌ی رتبه‌ی پنجره‌ی لغزشی SVD

هزینه‌های محاسباتی، از روش‌های سریع‌تری برای تخمین رتبه‌ی ماتریس‌های بزرگ بهره گرفته شود. همان‌طور که گفته شد، غیریک‌نواختی روشنایی در تصاویر آندوسکوپی، از جمله مواردی است که تشخیص پزشک را تحت تاثیر قرار می‌دهد. در این مقاله، غیریک‌نواختی روشنایی به کمک فیلتر انتشار تطبیقی، بهبود داده شده است. ارزیابی کیفی فیلتر انتشار کنتراست تطبیقی، در شکل (۶) ارائه شده است. در این شکل، ستون اول شامل فریم‌های اصلی مری، روده‌ی کوچک و معده بوده و فریم‌های تصحیح شده از نظر روشنایی، در ستون دوم ارائه شده است. واضح است که با استفاده از این روش، کیفیت تصاویر از لحاظ روشنایی، بسیار بهبود یافته است.

معمولاً روش‌های خلاصه‌سازی، در راستای ارسال فریم‌های خلاصه‌سازی شده برای سایر پزشکان و نیز آندوسکوپی از راه دور<sup>۱</sup>، بسیار ساده‌تر بوده و به هزینه و زمان کم‌تری نیاز دارند، زیرا در این روش‌ها، حجم بالایی از فریم‌های نامربوط و زاید حذف می‌شوند. از طرفی، در حالت عادی، به دلیل طولانی بودن ویدئوی کپسول آندوسکوپی، ممکن است درصد قابل توجهی از نارسایی‌های بالینی، از دید پزشک دور بماند. از این‌رو، با کاهش موثر تعداد فریم‌های زاید و نمایش فریم‌های برجسته و حاوی اطلاعات مفید به کمک روش‌های خلاصه‌سازی، امکان بررسی دقیق‌تر و تشخیص بهتر برای پزشکان فراهم می‌گردد. پیشنهاد می‌شود که در کارهای بعدی، به منظور کاهش هرچه بیش‌تر

<sup>۱</sup> Tele-Endoscopy

در این مقاله، علاوه بر کاهش محسوس فریم‌های ویدئویی، با به کارگیری یک مدل مبتنی بر فیلتر انتشار کنتراست تطبیقی، غیریک‌نواختی روشنایی و کنتراست تصاویر کپسول آندوسکوپی نیز تصحیح شده و بهبود یافته است. در کارهای بعدی، می‌توان امکان پیاده‌سازی این روش روی سایر داده‌های سیستم مراقبت از سلامت، مانند هیستروسکوپی<sup>۱</sup> و اکوکاردیوگرام<sup>۲</sup>، را مورد بررسی قرار داد. هم‌چنین، یکی دیگر از رویکردهایی که می‌تواند مورد بررسی گیرد، تلفیق اطلاعات پزشکی و بیمار، با استفاده از تکنیک‌هایی مانند نهان‌نگاری<sup>۳</sup> و مخفی‌سازی<sup>۴</sup> در فریم‌های کلیدی و انجام یک خلاصه‌سازی ایمن می‌باشد.

## ۵- سپاس‌گزاری

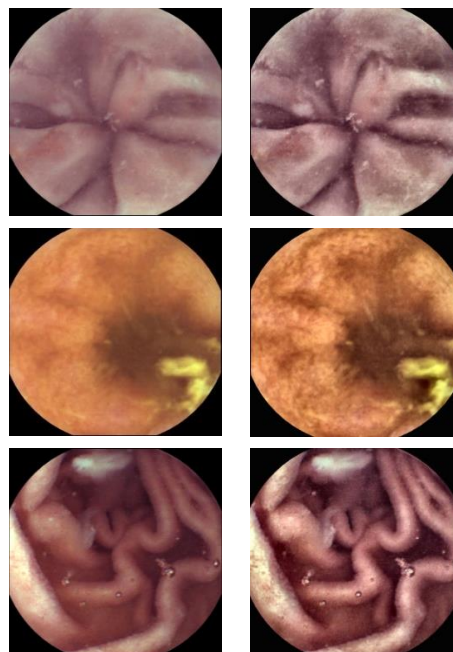
نویسندگان این مقاله، از هم‌کاری بخش تحقیقات گوارش و کبد بیمارستان فیروزگر تهران و پزشکان متخصص این بیمارستان، دکتر هاشم فخر یاسری و دکتر مسعود رضا سهرابی، به پاس در اختیار قرار دادن داده‌های محلی، ارائه‌ی نکات کلیدی و حمایت‌های آن‌ها، صمیمانه تشکر و قدردانی می‌نمایند.

## ۶- مراجع

- [1] S. P. Rajendra and N. Keshaveni, "A Survey of Automatic Video Summarization Techniques," *Int. J. Electron. Comput. Syst.*, vol. 3, no. 1, pp. 1-6, 2014.
- [2] A. Mohammed, S. Yildirim, M. Pedersen, O. Hovde, and F. Cheikh, "Sparse Coded Handcrafted and Deep Features for Colon Capsule Video Summarization," *Proc. - IEEE Symp. Comput. Med. Syst.*, vol. 2017-June, pp. 728-733, 2017.
- [3] S. Seguí *et al.*, "Generic feature learning for wireless capsule endoscopy analysis," *Comput. Biol. Med.*, vol. 79, pp. 163-172, 2016.
- [4] D. K. Iakovidis and A. Koulaouzidis, "Software for enhanced video capsule endoscopy: Challenges for essential progress," *Nat. Rev. Gastroenterol. Hepatol.*, vol. 12, no. 3, pp. 172-186, 2015.
- [5] M.H. A. Muhammad Ajmal Muhammad Shakir and F. A. S. Yasir Abbas, "Video Summarization Techniques and Classification," in *Computer Vision and Graphics. ICCVG*, 2012, pp. 1-13.
- [6] S. Eliza, F. De Avila, A. Paula, B. Lopes, L. Jr, and A. De Albuquerque, "VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method," *Pattern Recognit. Lett.*, vol. 32, no. 1, pp. 56-68, 2011.
- [7] S. Zhang, Y. Zhu, and A. K. Roy-Chowdhury, "Context-Aware Surveillance Video Summarization," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5469-5478, 2016.

<sup>۲</sup> Watermarking

<sup>۴</sup> Steganography



شکل (۶) - بهبود غیریک‌نواختی و کنتراست روشنایی، پس از انتخاب فریم‌های کلیدی، ستون سمت چپ: تصاویر واقعی از مری، روده‌ی کوچک و معده، ستون سمت راست: تصاویر بهبود یافته توسط فیلتر انتشار تطبیقی

## ۴- نتیجه‌گیری

در این مقاله، یک چارچوب کلی برای نمایش و خلاصه‌سازی ویدئوی کپسول آندوسکوپی، ارائه شده است. با به کارگیری این روش پیشنهادی، متخصصان گوارش می‌توانند محتوای ویدئوی کپسول آندوسکوپی را جستجو کرده و از این طریق، زمان لازم برای آنالیز هر فریم ویدئویی را کاهش دهند. با توجه به چارچوب ارائه شده، در پروسه‌ی SVD انطباقی، ابتدا فریم‌های کلیدی و با اهمیت انتخاب شده و سپس یک مرحله‌ی پس‌پردازش، برای بهبود و نمایش بهتر کنتراست فریم‌ها، انجام می‌شود. با توجه به قابلیت و انعطاف‌پذیری رویکرد تجزیه به مقادیر تکین SVD و ترکیب آن با پنجره‌ی لغزشی، درصد قابل توجهی از فریم‌های تکراری به کمک روش پیشنهادی حذف خواهد شد. روش پیشنهادی با بهره بردن از خواص منحصر به فرد SVD، هم ویژگی‌های زمانی و هم ویژگی‌های مکانی داده‌ی آندوسکوپی را تحت فضای برداری مقادیر تکین، ادغام می‌کند. از این‌رو، می‌توان ویدئوهای آندوسکوپی را به بخش‌هایی طبقه‌بندی کرد که محتوای دیداری مشابهی دارند. با توجه به نتایج کمی و کیفی، حجم بالای کاهش فریم‌ها منجر به کاهش زمان بازنگری و افزایش بهره‌وری متخصصان می‌شود.

<sup>۱</sup> Hysteroscopy

<sup>۳</sup> Echocardiogram

- [16] Y. Gong and X. Liu, "Video summarization and retrieval using singular value decomposition," *Multimed. Syst.*, vol. 168, pp. 157–168, 2003.
- [17] R. Badeau, G. Richard, and B. David, "Sliding window adaptive SVD algorithms," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 1–10, 2004.
- [18] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *J. Am. Soc. Inf. Sci.*, vol. 41, no. 6, p. 391, 1990.
- [19] E. Anderson *et al.*, "LAPACK: A portable linear algebra library for high-performance computers," in *Proceedings of the 1990 ACM/IEEE conference on Supercomputing*, 1990, pp. 2–11.
- [20] G. H. Golub and C. F. Van Loan, *Matrix computations*, vol. 3. JHU Press, 2012.
- [21] Y. Gong and X. Liu, "Video summarization using singular value decomposition," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, 2000, vol. 2, pp. 174–180.
- [22] B. Li and M. Q.-H. Meng, "Wireless capsule endoscopy images enhancement via adaptive contrast diffusion," *J. Vis. Commun. Image Represent.*, vol. 23, no. 1, pp. 222–228, 2012.
- [23] G.-H. Cottet and M. El Ayyadi, "A Volterra type model for image processing," *IEEE Trans. image Process.*, vol. 7, no. 3, pp. 292–303, 1998.
- [24] A. Biniiaz and A. Abbasi, "Unsupervised ACO : Applying FCM as a supervisor for ACO in medical image segmentation," *J. Intell. Fuzzy Syst.*, vol. 27, pp. 407–417, 2014.
- [8] S. Wang *et al.*, "Scalable gastroscopic video summarization via similar-inhibition dictionary selection," *Artif. Intell. Med.*, vol. 66, pp. 1–13, 2016.
- [9] M. M. Ben Ismail, O. Bchir, and A. Z. Emam, "Endoscopy Video Summarization based on Unsupervised Learning and Feature Discrimination," in *Visual Communications and Image Processing (VCIP)*, 2013, pp. 1–6.
- [10] D. K. Iakovidis, S. Tsevas, and A. Polydorou, "Reduction of capsule endoscopy reading times by unsupervised image mining," *Comput. Med. Imaging Graph.*, vol. 34, no. 6, pp. 471–478, 2010.
- [11] K. Muhammad, M. Sajjad, M. Young, and S. Wook, "Efficient visual attention driven framework for key frames extraction from hysteroscopy videos," *Biomed. Signal Process. Control*, vol. 33, pp. 161–168, 2017.
- [12] I. Mehmood, M. Sajjad, and S. W. Baik, "Video summarization based tele-endoscopy : a service to efficiently manage visual data generated during wireless capsule endoscopy procedure," *J. Med. Syst.*, 2014.
- [13] A. Koulaouzidis and D. K. Iakovidis, "KID: A capsule endoscopy database for medical decision support," *United Eur. Gastroenterol. Week (UEGW). Barcelona, Spain*, 2015.
- [14] "kid." [Online]. Available: <https://mdss.uth.gr/datasets/endoscopy/kid/>.
- [15] A. Vadivel, S. Sural, and A. K. Majumdar, "Human color perception in the HSV space and its application in histogram generation for image retrieval," in *Color Imaging: Processing, Hardcopy, and Applications*, 2005, pp. 598–609.