مقاله پژوهشی

تنظیم تطبیقی پارامتر کوانتیزاسیون در سطح فریم بر اساس ویژگی‌های سیستم بینایی انسان در کدگذار ویدئویی

مریم کلانتری، عباس ابراهیمی مقدم و مرتضی خادمی درح

چکیده: کدگذاری ویدئو نقش مهمی در کاهش حجم داده و بهینه‌سازی انتقال و ذخیره‌سازی محتوای چندرسانه‌ای دارد. در سال‌های اخیر، کدگذاری ادراکی با بهره‌گیری از ویژگی‌های سیستم بینایی انسان به‌عنوان رویکردی مؤثر برای کاهش نرخ بیت بدون افت قابل‌تشخیص کیفیت مطرح شده است. در این مقاله، یک روش ادراکی برای تنظیم تطبیقی پارامتر کوانتیزاسیون (ΔQP) در کدگذاری ویدئو با ساختار سلسله‌مراتبی زمانی ارائه می‌شود. روش پیشنهادی با استفاده از مدل آستانه تفاوت قابل‌تشخیص (JND)، حساسیت ادراکی هر فریم را تخمین زده و مقدار ΔQP را متناسب با ویژگی‌های محتوایی و اهمیت زمانی آن تنظیم می‌کند، به‌گونه‌ای که فریم‌های با حساسیت ادراکی کمتر فشرده‌تر شده و کیفیت بصری فریم‌های حساس‌تر حفظ می‌شود.

نتایج آزمایش‌ها بر روی توالی‌های استاندارد و با استفاده از کدگذار HEVC نشان می‌دهد که روش پیشنهادی به‌طور متوسط ۲٫۵٪ کاهش نرخ بیت همراه با ۱٫۶۷٪ کاهش زمان اجرا ایجاد می‌کند. افزون بر این، ارزیابی ذهنی مبتنی بر شاخص I-MOS بهبود میانگین ۰٫۱۷ واحدی را نسبت به نسخه مرجع نشان می‌دهد و در اکثر توالی‌ها کیفیت بصری نسخه پیشنهادی بهتر یا برابر نسخه مرجع است. این نتایج نشان می‌دهد که تنظیم تطبیقی ΔQP مبتنی بر JND رویکردی کارآمد، کم‌پیچیدگی و سازگار با HEVC برای کاربردهای ویدئویی است.

***کلیدواژه:* کدگذاری ویدئو، کدگذاری ادراکی، پارامتر کوانتیزاسیون (QP)، آستانه تفاوت قابل‌تشخیص (JND)، ساختار سلسله‌مراتبی زمانی، فشرده‌سازی ویدئو با کارایی بالا (HEVC).**

۱- مقدمه

کدگذاری ویدئو نقش اساسی در کاهش حجم داده و انتقال کارآمد محتوای چندرسانه‌ای دارد [۱]. استاندارد فشرده‌سازی ویدئو با کارایی بالا[1] (HEVC) به‌عنوان یکی از مهم‌ترین دستاوردهای اخیر در حوزه فشرده‌سازی، نسبت به نسل‌های پیشین بهبود چشمگیری در نسبت نرخ‌بیت به کیفیت ارائه کرده است [۲]. با وجود این مزایا، روش‌های کلاسیک کدگذاری همچنان عمدتاً بر معیارهای سیگنال‌محور مانند PSNR[2] تکیه دارند؛ معیارهایی که همبستگی محدودی با کیفیت ادراکی سیستم بینایی انسان[3] (HVS) دارند و همین موضوع ضرورت بهره‌گیری از مدل‌های ادراکی را برجسته می‌سازد.

کدگذاری ادراکی با استفاده از ویژگی‌های HVS، امکان کاهش نرخ‌بیت بدون افت قابل‌تشخیص کیفیت بصری را فراهم می‌کند. یکی از ابزارهای کلیدی در این حوزه، آستانه تفاوت قابل‌تشخیص[4] (JND) است که به‌طور گسترده در کاربردهای کدگذاری تصویر و ویدئو مورد استفاده قرار می‌گیرد [۳]. با این حال، بسیاری از روش‌های مبتنی بر JND ارائه‌شده برای HEVC یا هزینه محاسباتی بالایی دارند، یا عمدتاً در سطح بلوک عمل می‌کنند و کمتر به نقش ساختار سلسله‌مراتبی زمانی در تعیین پارامتر کوانتیزاسیون[5] (QP) توجه کرده‌اند.

در این پژوهش، روشی ارائه می‌شود که در آن میزان جبران پارامتر کوانتیزاسیون (ΔQP) هر فریم بر اساس آستانه ادراکی JND و اهمیت لایه زمانی آن در ساختار سلسله‌مراتبی و نحوه تخصیص اولیه ΔQP مطابق مستند HM16.20 [۴] تعیین می‌شود. بدین ترتیب، فریم‌های کلیدی‌تر که نقش مهم‌تری در بازسازی توالی دارند، ΔQP کوچک‌تری دریافت می‌کنند و فریم‌های لایه‌های بالاتر امکان فشرده‌سازی بیشتری پیدا می‌کنند. ترکیب مدل ادراکی JND با ساختار زمانی HEVC موجب می‌شود بخش‌هایی از محتوا که حساسیت ادراکی کمتری دارند، فشرده‌تر شده و کیفیت بصری بخش‌های مهم‌تر حفظ شود.

ساختار مقاله به این ترتیب سازمان‌دهی شده است: بخش دوم به مرور پژوهش‌های مرتبط اختصاص دارد؛ در بخش سوم، روش پیشنهادی به‌طور تفصیلی تشریح می‌شود؛ بخش چهارم به ارائه و تحلیل نتایج عینی و ذهنی می‌پردازد؛ و در نهایت، بخش پنجم جمع‌بندی و نتیجه‌گیری را ارائه می‌کند.

۲- پیشینه پژوهش

مرور پیشینه در این پژوهش در دو محور متمرکز می‌شود: محور نخست، روش‌های تنظیم QP در سطح فریم که با هدف کنترل اعوجاج و مدیریت نرخ‌بیت ارائه شده‌اند؛ و محور دوم، مطالعات مبتنی بر مدل‌های آستانه JND که تلاش می‌کنند با تقلید از رفتار HVS، کیفیت ادراکی را بهبود داده و نرخ داده را کاهش دهند. این دسته‌بندی، چارچوب مفهومی پژوهش را شکل می‌دهد و مسیر تحلیل و توسعه روش پیشنهادی را

این مقاله در تاریخ ۸ دی ماه ۱۴۰۴ دریافت و در تاریخ ۲۲ فروردین ماه ۱۴۰۵ بازنگری شد.

مریم کلانتری، گروه برق، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران، (email: maryam.kalantari@um.ac.ir).

عباس ابراهیمی مقدم (نویسنده مسئول)، گروه برق، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران، (email: a.ebrahimi@um.ac.ir).

مرتضی خادمی، گروه برق، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران، (email: khademi@um.ac.ir).

1. High Efficiency Video Coding
2. Peak Signal-to-Noise Ratio
3. Human Visual System
4. Just Noticeable Difference
5. Quantization Parameter

روشن می‌سازد.

در حوزه تنظیم QP در سطح فریم، پژوهش‌های متعددی با هدف مدیریت نرخ‌بیت و بهبود رفتار نرخ–اعوجاج[1] (R–D) ارائه شده‌اند. این مطالعات عمدتاً بر مدل‌سازی وابستگی‌های بین‌فریمی متمرکز بوده و نشان می‌دهند که اعوجاج ایجادشده در فریم‌های مرجع می‌تواند به فریم‌های پیش‌بینی‌شده منتقل شده و بر کیفیت آن‌ها تأثیرگذار باشد. بر این اساس، مدل‌سازی انتشار اعوجاج با بهره‌گیری از اطلاعات حرکتی و انرژی پسماند، به‌عنوان مبنایی برای تنظیم تطبیقی QP مورد استفاده قرار گرفته است. در ادامه، رویکردهای پیشرفته‌تری با استفاده از شبکه‌های عصبی برای تحلیل هم‌زمان وابستگی اعوجاج و حرکت در سطح واحدهای کدینگ پیشنهاد شده‌اند که امکان تنظیم QP مستقل از ساختار ثابت گروه تصاویر[2] (GOP) را فراهم می‌کنند [۵]؛ هرچند این روش‌ها معمولاً با افزایش قابل‌توجه پیچیدگی محاسباتی همراه هستند.

جریان مهم دیگری از مطالعات، تنظیم QP را بر مبنای بهینه‌سازی مستقیم R–D دنبال کرده است. در این رویکردها، رابطه نرخ و اعوجاج با استفاده از مدل‌های تحلیلی یا آماری تقریب می‌شود و مقدار QP با حل یک مسئله بهینه‌سازی تعیین می‌گردد؛ به‌گونه‌ای که تخمین دقیق‌تر مدل R–D به کاهش قابل‌توجه نرخ‌بیت منجر می‌گردد [۶] و [۷]. افزون بر این، برخی پژوهش‌ها با درنظرگرفتن پدیده انتشار اعوجاج در طول فریم‌ها نشان داده‌اند که لحاظ‌کردن اثر اعوجاج فریم‌های مرجع در تابع هزینه R–D می‌تواند به بهبود محسوسی در BD-Rate[3] بینجامد [۸]. همچنین، بهره‌گیری از ساختارهای سلسله‌مراتبی پیش‌بینی یا گروه‌بندی مجازی فریم‌ها برای تصمیم‌گیری چندمرحله‌ای در تخصیص QP، امکان بهینه‌سازی هم‌زمان در سطح فریم و لایه‌های زمانی را فراهم کرده و راهکارهایی کارآمد برای مدیریت QP ارائه می‌دهد [۹].

در کنار این رویکردهای تحلیلی، دسته‌ای از مطالعات به جنبه‌های محتوایی پرداخته‌اند و مقدار QP را بر اساس ویژگی‌هایی مانند پیچیدگی بافت، شدت حرکت، روشنایی یا الگوهای فرکانسی تعیین کرده‌اند. در این میان، استفاده از شبکه‌های عصبی عمیق برای استخراج ویژگی‌های بصری نیز مورد توجه قرار گرفته است؛ به‌طوری‌که روش مبتنی بر شبکه VGG-16 در [۱۰] کاهش چشمگیری در نرخ‌بیت ایجاد کرده است، هرچند هزینه محاسباتی بالای آن می‌تواند کاربرد عملی چنین روش‌هایی را محدود سازد. همچنین، نتایج پژوهش‌های اخیر نشان می‌دهد که تحلیل محتوای فریم‌ها می‌تواند مبنای مؤثری برای تنظیم تطبیقی QP و بهبود بازده فشرده‌سازی باشد؛ به‌گونه‌ای که این ایده در استانداردهای جدیدتر کدگذاری ویدئو نیز با رویکردهای مبتنی بر استخراج ویژگی‌های بصری دنبال شده است [۱۱].

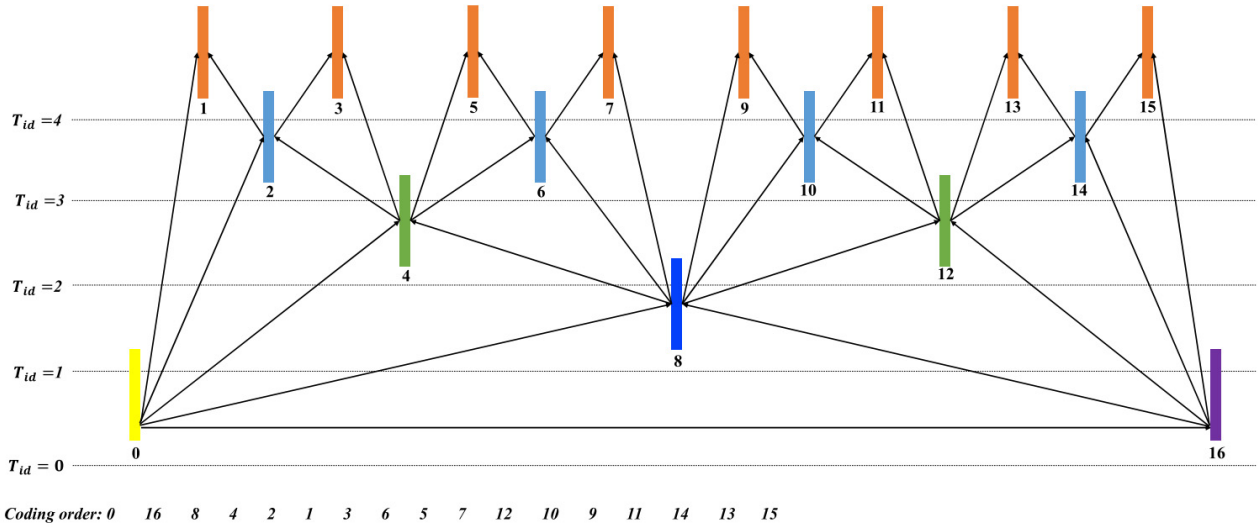
با وجود پیشرفت‌های قابل‌توجه در تنظیم QP (به‌ویژه در سطح فریم و لایه‌های زمانی[4])، اغلب روش‌های موجود همچنان بر ویژگی‌های ساختاری یا آماری محتوا تکیه دارند و از درنظرگرفتن ویژگی‌های ادراکی بیننده غفلت می‌کنند. این محدودیت موجب شده است که مدل‌های آستانه JND به‌عنوان رویکردی مؤثر در فشرده‌سازی ادراکی مطرح شوند. مدل‌های JND با شناسایی اعوجاج‌های غیرقابل‌تشخیص برای سیستم بینایی انسان، امکان کاهش نرخ‌بیت را بدون افت محسوس کیفیت فراهم می‌کنند و به همین دلیل در سال‌های اخیر به یکی از ابزارهای کلیدی در کدگذاری ویدئویی مبتنی بر HEVC تبدیل شده‌اند. در همین راستا، بخش قابل‌توجهی از پژوهش‌ها به کاربرد JND در بهبود کوانتیزاسیون، کنترل نرخ‌بیت و افزایش بازده فشرده‌سازی ادراکی پرداخته‌اند [۴].

مدل‌های JND در بهینه‌سازی کوانتیزاسیون با تخمین حساسیت بصری در نواحی مختلف تصویر، مقدار QP را به‌صورت تطبیقی تنظیم کرده و محدودیت‌های روش‌های مبتنی بر QP یکنواخت را کاهش می‌دهند. این رویکردها با درنظرگرفتن پیچیدگی بافت، حرکت و حساسیت زمانی، اعوجاج را در سطح آستانه ادراکی کنترل کرده و کاهش نرخ‌بیت را بدون افت محسوس کیفیت امکان‌پذیر می‌سازند. پژوهش‌های اخیر نشان داده‌اند که تنظیم تطبیقی QP بر اساس مدل‌های ادراکی می‌تواند به بهبود معنادار کارایی فشرده‌سازی منجر شود [۱۲] و [۱۳]؛ به‌گونه‌ای که رویکردهای مشابهی در استانداردهای جدیدتر کدگذاری ویدئو نیز با بهره‌گیری از مدل‌های ادراکی و یادگیری‌محور برای بهینه‌سازی QP دنبال شده‌اند [۱۴].

در حوزه کنترل نرخ‌بیت، پژوهش‌ها نشان داده‌اند که مدل‌های JND می‌توانند به‌عنوان مؤلفه‌ای ادراکی در کنار سازوکارهای کلاسیک بهینه‌سازی نرخ–اعوجاج، نقش مؤثری ایفا کنند. در این رویکردها، آستانه JND برای هر فریم یا واحد کدینگ تخمین زده می‌شود و این اطلاعات ادراکی در فرآیندهای مختلفی از جمله تخصیص بیت، تنظیم QP و یا اصلاح تابع لاگرانژی مورد استفاده قرار می‌گیرد؛ به‌گونه‌ای که نواحی کم‌اهمیت بصری فشرده‌تر شده و منابع بیت برای بخش‌های حساس‌تر ذخیره گردند. برای نمونه، مطالعه [۱۵] با ترکیب روشنایی ادراکی و مقادیر JND در سطح واحد کدینگ درختی[5] (CTU)، از طریق بهبود تخصیص بیت و کاهش نوسانات کیفیت، واریانس نرخ‌بیت را کاهش داده و کیفیت ادراکی یکنواخت‌تری فراهم کرده است. همچنین، روش [۱۶] با به‌کارگیری JND به‌عنوان ضریب وزنی در تخصیص بیت و اصلاح تابع لاگرانژی، کنترل پایدارتر کیفیت بین‌فریمی را بدون اتکای صرف به تنظیم QP امکان‌پذیر ساخته است.

فشرده‌سازی ادراکی ویدئو[6] یکی از مهم‌ترین کاربردهای مدل‌های JND در HEVC به شمار می‌رود. در این رویکرد، داده‌هایی که زیر آستانه تشخیص سیستم بینایی انسان قرار می‌گیرند شناسایی و حذف می‌شوند تا افزونگی بصری کاهش یابد. برای مثال، مطالعه [۱۷] با تحلیل ویژگی‌های فضایی–زمانی، حذف دقیق‌تری از داده‌های زیرآستانه ارائه کرده است. همچنین روش‌های [۱۸] و [۱۹] با سرکوب ضرایب تبدیل کسینوسی گسسته[7] (DCT) در کدگذاری درون‌فریمی، کاهش چشمگیری چشمگیری در نرخ‌بیت ایجاد کرده‌اند بدون آن‌که کیفیت بصری افت قابل‌تشخیصی داشته باشد. افزون بر این، چارچوب‌های ادراکی مبتنی بر JNDو توجه بصری، مانند [۲۰]، با بهره‌گیری هم‌زمان از مدل‌های برجستگی بصری و آستانه JND ، نشان داده‌اند که می‌توان افزونگی‌های ادراکی را به‌صورت هدفمند شناسایی و حذف کرد و بازده فشرده‌سازی را بدون افت محسوس کیفیت افزایش داد. همچنین، رویکردهای یادگیری‌محور مبتنی بر JND، از جمله روش ارائه‌شده در [۲۱] برای پیش‌پردازش HDR[8] در HEVC و مدل‌های مبتنی بر شبکه‌های عمیق

1. Rate–Distortion
2. Group of Pictures
3. Bjøntegaard Delta Rate
4. Temporal Layer
5. Coding Tree Unit
6. Perceptual Video Coding
7. Discrete Cosine Transform
8. High Dynamic Range



شکل ۱: ساختار سلسله‌مراتبی دسترسی تصادفی در HEVC با GOP = ۱۶ [۹].

مانند [۲۲]، نتایج مشابهی را در جهت فشرده‌سازی ادراکی گزارش کرده‌اند.

بخشی از پژوهش‌ها به کاربرد هم‌زمان مدل‌های JND در چندین ماژول کدگذار پرداخته‌اند؛ جایی که اطلاعات ادراکی به‌صورت یکپارچه در کوانتیزاسیون، فیلترگذاری پسماند و بهینه‌سازی نرخ–اعوجاج به کار گرفته می‌شود تا بازده فشرده‌سازی افزایش یابد. در این راستا، مطالعه [۲۳] با ترکیب مدل JND و فیلترگذاری ادراکی پسماند، اعوجاج‌های غیرقابل‌مشاهده را حذف کرده و کیفیت کدگذاری درون‌فریمی را بهبود داده است. همچنین، طرح JMCP در [۲۴] با به‌کارگیری هم‌زمان کوانتیزاسیون تطبیقی، فیلترگذاری پسماند و بهینه‌سازی R–D، و با در نظر گرفتن معیارهای ادراکی از جمله مدل JND، عملکردی کارآمدتر و ادراکی‌تر نسبت به روش‌های تک‌ماژوله فراهم کرده است. رویکردهای مشابهی با تمرکز بر یکپارچه‌سازی ماژول‌های ادراکی، در استانداردهای جدیدتر کدگذاری ویدئو نیز گزارش شده‌اند که نشان‌دهنده تداوم این خط پژوهشی در نسل‌های جدید کدگذارهاست [۲۵].

جمع‌بندی این مرور نشان می‌دهد که مدل‌های آستانه JND با تکیه بر ویژگی‌های سیستم بینایی انسان و تحلیل فضایی–زمانی محتوا، در دسته رویکردهای محتوایی–ادراکی قرار گرفته و توان بالایی در کاهش نرخ‌بیت همراه با حفظ کیفیت بصری در کدگذاری HEVC دارند. در مقابل، بسیاری از روش‌های متداول تنظیم QP همچنان رویکردی سیگنال‌محور داشته و بر شاخص‌های تحلیلی یا قواعد ازپیش‌تعریف‌شده مبتنی بر ساختار زمانی فریم‌ها متکی‌اند. هرچند رویکردهای ادراکی در سطوح مکانی و بلوکی مورد توجه قرار گرفته‌اند، تنظیم QP در سطح فریم در اغلب روش‌ها هنوز از این اطلاعات بهره‌مند نشده است؛ ازاین‌رو، روش پیشنهادی این پژوهش با هدف پر کردن این خلأ، تنظیم QP فریمی را در چارچوب یک رویکرد محتوایی مبتنی بر مدل JND و در تعامل با ساختار سلسله‌مراتبی فریم‌ها انجام می‌دهد.

۳- روش پیشنهادی

در استاندارد HEVC، اختلاف QP میان لایه‌های زمانی در پیکربندی دسترسی تصادفی[1] (RA) معمولاً مقداری ثابت در نظر گرفته می‌شود [۲]، [۳]، در حالی‌که فریم‌ها از نظر نقش در پیش‌بینی و حساسیت ادراکی یکسان نیستند. در ساختار سلسله‌مراتبی RA، همان‌گونه که در شکل ۱ نشان داده شده است، فریم‌ها بر اساس لایه‌های زمانی سازمان‌دهی می‌شوند که این لایه‌ها با نماد T_{id}[2] مشخص شده‌اند و نشان‌دهنده‌ی سطح زمانی هر فریم در ساختار سلسله‌مراتبی پیش‌بینی است. فریم‌های لایه‌های پایین‌تر، به‌ویژه فریم‌های درون‌کدگذاری‌شده[3] (I-frames) و پیش‌بینی‌شده[4] (P-frames)، به دلیل تأثیرگذاری بیشتر بر زنجیره پیش‌بینی، باید با مقادیر ملایم‌تری از QP کدگذاری شوند. در مقابل، فریم‌های لایه‌های بالاتر به دلیل نقش محدودتر در بازسازی نهایی، ظرفیت بیشتری برای فشرده‌سازی داشته و می‌توانند با QP بزرگ‌تر کدگذاری شوند [۲۶].

ثابت بودن ΔQP در ساختار سلسله‌مراتبی RA با ویژگی‌های ادراکی انسان و تنوع محتوایی فریم‌ها همخوانی ندارد و ممکن است به تخصیص غیربهینه نرخ‌بیت منجر شود. برای رفع این محدودیت، در این پژوهش مقدار ΔQP هر فریم بر اساس حساسیت ادراکی انسان و جایگاه آن در لایه زمانی به‌صورت تطبیقی تنظیم می‌شود [۳] و [۲۶].

روش پیشنهادی دو مرحله دارد: در گام نخست، آستانه JND هر فریم با بهره‌گیری از مدل مرجع [۲۷] و با درنظرگرفتن مؤلفه‌های ادراکی شامل تابع حساسیت کنتراست[5]، روشنایی، کنتراست و حرکت محاسبه می‌شود. این مدل در حوزه DCT عمل کرده و مقدار JND هر ضریب بر اساس (۱) تعیین می‌گردد:

$$JND(x,y) = T_B(x,y) \times F_{LA}(x,y) \times F_{CM}(x,y) \times F_{motion}(x,y) \quad (۱)$$

که در آن $T_B(x,y)$ آستانه پایه مدل JND است که بر اساس تابع حساسیت کنتراست محاسبه می‌شود و تغییرات قابل‌تشخیص برای چشم انسان را در فرکانس‌های فضایی مختلف توصیف می‌کند. ضریب $F_{LA}(x,y)$ بیانگر تطبیق روشنایی است و اثر سطح روشنایی محلی را بر حساسیت ادراکی مدل می‌کند؛ به‌طوری‌که در نواحی بسیار تاریک یا بسیار روشن، حساسیت بینایی کاهش می‌یابد [۲۷]. ضریب $F_{CM}(x,y)$ اثر

1. Random Access
2. Temporal ID
3. Intra-coded Frames
4. Predictive Frames
5. Contrast Sensitivity Function



(الف)



(ب)

شکل ۲: (الف) فریم شماره ۲۰ از توالی BQMall و (ب) نقشه‌ی آستانه‌ی *JND* برای فریم ۲۰ از توالی BQMall.

پوشانندگی کنتراست[1] را توصیف کرده و افزایش آستانه ادراکی در حضور بافت‌ها و لبه‌های قوی را بازتاب می‌دهد؛ در این نواحی، وجود جزئیات پیچیده موجب پنهان‌شدن اعوجاج‌های ضعیف‌تر می‌شود. در نهایت، ضریب $F_{motion}(x,y)$ نقش حرکت را در آستانه ادراکی لحاظ می‌کند و بیان می‌دارد که چشم انسان در نواحی دارای حرکت نسبت به اعوجاج‌های کوانتیزاسیون حساسیت کمتری دارد [۲۷].

با ترکیب این چهار مؤلفه، مقدار نهایی *JND* برای هر ضریب DCT محاسبه شده و آستانه ادراکی قابل‌تشخیص برای هر زیرباند به‌دست می‌آید. شکل ۲- الف و ۲- ب به‌ترتیب نمونه‌ای از فریم‌های توالی BQMall [۲۶] و نقشه JND استخراج‌شده متناظر با آن را نشان می‌دهند؛ به‌طوری‌که نواحی روشن‌تر بیانگر آستانه بالاتر و در نتیجه حساسیت ادراکی کمتر هستند

پس از محاسبه مقادیر *JND* در سطح ضرایبDCT ، شاخص ادراکی کل فریم با میانگین‌گیری از مقادیر JND تمامی ضرایب DCT موجود در فریم مطابق (۲) محاسبه می‌شود:

$$JND_{frame} = \frac{1}{BN^2}\sum_{b=1}^{B}\sum_{i=1}^{N}\sum_{j=1}^{N} JND_b(i,j) \qquad (۲)$$

در این رابطه، میانگین *JND* هر بلوک DCT برای بلوکی با ابعاد $N \times N$ در نظر گرفته شده و B تعداد بلوک‌های DCT فریم را نشان می‌دهد. شاخص حاصل بیانگر میزان حساسیت ادراکی فریم نسبت به اعوجاج‌های ناشی از کوانتیزاسیون بوده و به‌عنوان مبنای تنظیم تطبیقی ΔQP مورد استفاده قرار می‌گیرد.

در روش پیشنهادی، هدف از به‌کارگیری مدل *JND* استخراج یک شاخص ادراکی در سطح فریم برای تعیین مقدار ΔQP هر فریم است؛ بنابراین، استفاده از اطلاعات *JND* در سطح بلوک یا ناحیه در چارچوب این روش ضرورتی ندارد، زیرا خروجی مورد نیاز الگوریتم یک مقدار ΔQP برای کل فریم است. در مطالعات پیشین نیز اطلاعات ادراکی *JND* متناسب با هدف الگوریتم، در سطوح مختلفی از جمله CTU و کنترل نرخ به‌کار گرفته شده است [۱۵], [۱۶]. هرچند استفاده از ضرایب ناحیه‌ای یا بلوکی می‌تواند کیفیت ادراکی را بهبود دهد، اما با افزایش پیچیدگی محاسباتی همراه است. ازاین‌رو، در این پژوهش رویکرد فریمی با هدف ارائه روشی ساده، با پیچیدگی کم و سازگار با ساختار کدگذار HEVC انتخاب شد. بررسی تخصیص ΔQP در سطح ناحیه یا بلوک نیز به‌عنوان یکی از مسیرهای پژوهش‌های آینده در نظر گرفته شده است.

برای ایجاد یک مقیاس کنترل‌پذیر در فرآیند تنظیم، مقدار JND_{frame} با استفاده از نرمال‌سازی خطی به بازه عددی (۰٫۵–۱٫۵) نگاشت می‌شود. انتخاب نرمال‌سازی خطی به‌منظور حفظ رابطه یکنواخت بین تغییرات شاخص ادراکی و میزان تنظیم ΔQP و جلوگیری از تقویت بیش‌ازحد نوسانات کوچک شاخص صورت گرفته است. نرمال‌سازی مطابق (۳) انجام می‌گیرد:

$$JND_{norm} = ۰٫۵ + \frac{JND_{frame} - JND_{min}}{JND_{max} - JND_{min}} \qquad (۳)$$

در مرحله دوم، مقدار نهایی ΔQP هر فریم با ترکیب ضریب ادراکی $JND_{norm}(f)$ و وزن متناظر با لایه زمانی آن تعیین می‌شود. وزن‌های لایه‌های زمانی مطابق جدول ۱، با در نظر گرفتن اهمیت هر لایه در ساختار سلسله‌مراتبی پیکربندی RA و نیز مقادیر ثابت مورد استفاده در نسخه مرجع انتخاب شده‌اند تا فریم‌های حساس‌تر مقدار ΔQP کوچک‌تری دریافت کنند و فریم‌های با حساسیت ادراکی پایین‌تر (به‌ویژه در لایه‌های بالاتر) مقادیر بزرگ‌تری از ΔQP به آن‌ها اختصاص یابد. رابطه کلی تنظیم به‌صورت (۴) تعریف می‌شود:

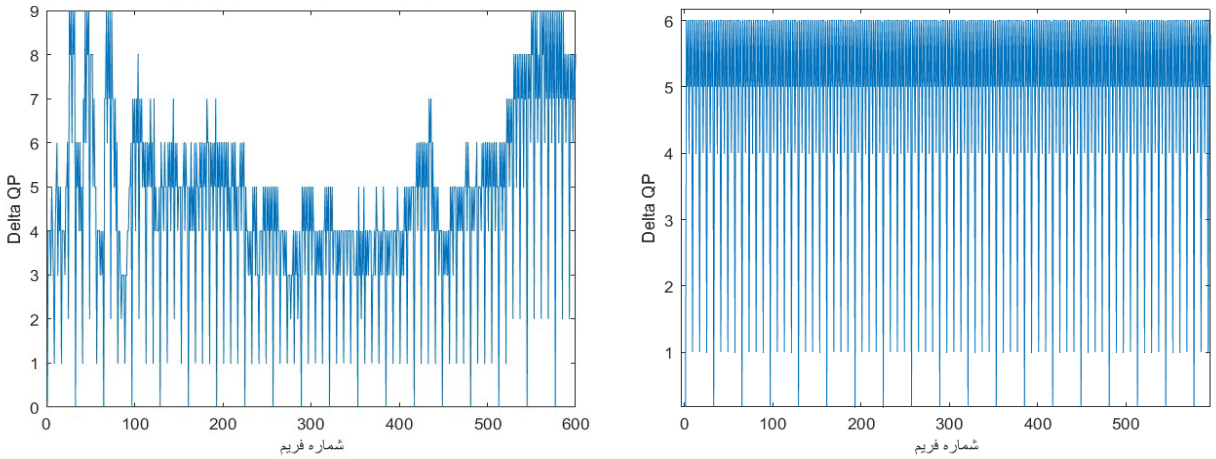
$$\Delta QP_f = JND_{norm}(f) \times W_{Layer}(f) \qquad (۴)$$

که در آن، f شماره فریم در حال کدگذاری است. به‌این‌ترتیب، اختلاف QP میان لایه‌های زمانی از یک مقدار ثابت و ازپیش‌تعریف‌شده به یک سازوکار تطبیقی و وابسته به محتوا تبدیل می‌شود.

شکل‌های ۳- الف و ۳- ب به‌ترتیب الگوی ΔQP در روش مرجع HEVC و روش پیشنهادی را نمایش می‌دهند. همان‌گونه که در شکل ۳- الف مشاهده می‌شود، در روش مرجع، مقدار ΔQP ثابت بوده و تنها بر اساس لایه زمانی فریم تعیین می‌شود [۳], [۲۶]. در مقابل، شکل ۳- ب نشان می‌دهد که روش پیشنهادی با بهره‌گیری از آستانه ادراکی *JND* و ساختار سلسله‌مراتبی RA، مقدار ΔQP را به‌صورت تطبیقی و متناسب با محتوای هر فریم تنظیم می‌کند. حاصل این رویکرد، الگوی تغییرات پیوسته‌تر و سازگار با ویژگی‌های بصری فریم است.

در این روش تنها سازوکار تعیین مقدار ΔQP در سمت کدگذار تغییر یافته است و هیچ داده جانبی (side information) یا کانال فرعی به بیت‌استریم اضافه نمی‌شود. مقدار نهایی QP مطابق سازوکار استاندارد HEVC در بیت‌استریم کد می‌شود؛ از این رو، دیکودر استاندارد بدون هیچ‌گونه تغییر قادر به بازسازی ویدئو است.

1. Contrast Masking



(ب) (الف)

شکل ۳: مقدار ΔQP فریم های متوالی ویدئو آزمون BQMall (الف) روش مرجع HM۱۶٫۲۰ و (ب) روش پیشنهادی.

٤- روش شبیه‌سازی

در این بخش، عملکرد روش پیشنهادی در مقایسه با نسخه مرجع HEVC مورد ارزیابی قرار می‌گیرد. به‌منظور تحلیل جامع رفتار الگوریتم، مجموعه‌ای از آزمایش‌های عینی و ذهنی بر روی توالی‌های استاندارد کلاس‌های مختلف HEVC انجام شده است. نتایج عینی شامل محاسبه BD-Rate برای معیارهای PSNR [۲۸] و MS-SSIM [۲۹]، اندازه‌گیری تغییرات نرخ‌بیت (ΔBR) و بررسی تغییر زمان اجرا (ΔT) است؛ همچنین، ارزیابی ذهنی کیفیت بصری نسخه پیشنهادی با استفاده از آزمون I-MOS[1] و با الهام از توصیه ITU-T P.۹۱۰ انجام شد [۳۰]. برای ارزیابی کیفیت بصری این ارزیابی‌ها امکان بررسی دقیق تأثیر مدل JND و سازوکار تنظیم تطبیقی ΔQP را بر کیفیت بصری و بازده فشرده‌سازی فراهم می‌کنند.

تمامی آزمایش‌ها تحت پیکربندی استاندارد RA اجرا شده‌اند [۲۶] که مطابق با تنظیمات رسمی JCT-VC[2] است. پارامترهای کلیدی شبیه‌سازی در جدول ۲ ارائه شده‌اند. شبیه‌سازی‌ها با استفاده از نرم‌افزار مرجع HEVC نسخه HM۱۶٫۲۰ و بر روی یک رایانه شخصی با پردازنده Core i۵–۱۳۶۰۰K@۵٫۱ GHz، ۳۲ گیگابایت حافظه RAM و سیستم‌عامل (۶۴ بیتی) Windows ۱۱ انجام شده‌اند. هر دو نسخه مرجع و پیشنهادی تحت تنظیمات یکسان پردازش موازی و با بهره‌گیری از تمامی هسته‌های در دسترس در محیط شبیه‌سازی Microsoft Visual Studio ۲۰۱۳ اجرا شده‌اند. در روش پیشنهادی، تنها ماژول محاسبه شاخص ادراکی JND و تعیین تطبیقی ΔQP به کد مرجع افزوده شده و سایر اجزای کدگذار، شامل پیش‌بینی بین‌فریمی، تبدیل، کوانتیزاسیون پایه و بهینه‌سازی نرخ–اعوجاج، بدون تغییر باقی مانده‌اند؛ بنابراین، تغییرات گزارش‌شده در زمان اجرای کدگذاری صرفاً بازتاب‌دهنده تأثیر الگوریتم پیشنهادی هستند.

به‌منظور حفظ سازگاری با ساختار کدگذار مرجع و جلوگیری از افزایش پیچیدگی محاسباتی، محاسبه شاخص ادراکی JND در حوزه تبدیل و بر پایه بلوک‌های DCT با ابعاد ۸×۸ انجام شده است. این اندازه، یکی از مقیاس‌های متداول در مدل‌های ادراکی مبتنی بر DCT بوده و توازن مناسبی میان دقت ادراکی و سربار محاسباتی فراهم می‌کند.

جدول ۱: تخصیص وزن به لایه‌های زمانی در ساختار RA.

لایه‌ی زمانی T_{id}	وزن $W_{Layer}(f)$
۰	۰
۱	۱
۲	۴
۳	۵
٤	۶

جدول ۲: پارامترهای کلیدی شبیه‌سازی.

پارامتر	مقدار
ساختار GOP	۱۶
تعداد فریم‌های مرجع[3]	۴
مد کدگذاری[4]	دسترسی تصادفی(RA)
سطوح QP آزمایش شده	۲۲، ۲۷، ۳۲ و ۳۷
فرمت نمونه‌برداری رنگ	۴:۲:۰ با عمق بیت ۸ بیت
پروفایل انتخابی	Main Profile

٤-۱ نتایج عینی

در این بخش، نتایج عینی[5] حاصل از مقایسه نسخه پیشنهادی با نسخه مرجع ارائه شده‌اند. معیارهای BD-Rate مبتنی بر PSNR و MS-SSIM محاسبه شده و تغییرات نرخ‌بیت و زمان اجرای کدگذاری برای تمامی مقادیر QP آزمایشی مورد بررسی قرار گرفته‌اند. در تحلیل نتایج، تمرکز اصلی بر معیار MS-SSIM قرار داده شده است، زیرا این شاخص همبستگی بالاتری با کیفیت ادراکی دارد.

برای تحلیل اثر روش پیشنهادی بر پیچیدگی محاسباتی، تغییر زمان اجرای کدگذاری نسبت به نسخه مرجع HEVC به‌صورت درصدی مطابق (۵) محاسبه شده است:

$$\Delta T(\%) = \frac{T_{prop} - T_{HM}}{T_{HM}} \times ۱۰۰ \qquad (۵)$$

1. Informal Mean Opinion Score
2. Joint Collaborative Team on Video Coding
3. Reference Frames
4. Coding Structure
5. Objective Results

که در آن T_{prop} و T_{HM} به‌ترتیب زمان اجرای نسخه پیشنهادی و نسخه مرجع HEVC هستند. مقدار منفی ΔT بیانگر کاهش زمان اجرا و بهبود کارایی محاسباتی است، در حالی‌که مقادیر مثبت نشان‌دهنده افزایش سربار پردازشی می‌باشند.

جدول ۳ نتایج تجمیع‌شده ۱۷ توالی آزمایشی را گزارش می‌کند و امکان ارزیابی عملکرد روش پیشنهادی در محتوای متنوع را فراهم می‌سازد.

بررسی جدول نشان می‌دهد که روش پیشنهادی در بخش قابل‌توجهی از توالی‌ها موجب کاهش نرخ‌بیت شده و در عین حال کیفیت ادراکی بسیار نزدیک یا مشابه نسخه مرجع را حفظ کرده است. معیار MS-SSIM (که شاخص قابل‌اعتمادتری نسبت به PSNR از منظر ادراکی محسوب می‌شود) در بسیاری از موارد بهبود یا تغییر ناچیز داشته است؛ در حالی‌که کاهش نرخ‌بیت غالباً منفی بوده و در برخی توالی‌ها به بیش از ۷٪ نیز رسیده است (از جمله PeopleOnStreet، BasketballDrive و vidyo3). این نتایج نشان می‌دهد که افزایش تطبیقی ΔQP در فریم‌هایی که حساسیت ادراکی پایین‌تری دارند، توانسته است بدون ایجاد افت قابل‌مشاهده در کیفیت، نرخ‌بیت را کاهش دهد.

از سوی دیگر، در برخی توالی‌ها افزایش BD-Rate مشاهده شده است. این موارد عمدتاً مربوط به توالی‌هایی هستند که دارای بافت‌های یکنواخت همراه با نویز مصنوعی یا حرکت‌های نامنظم با فرکانس بالا هستند (مانند PartyScene و RaceHorses). چنین محتوایی به‌طور ذاتی نسبت به تغییرات کوچک در کوانتیزاسیون حساس‌تر است و مدل *JND* فضای محدودی برای افزایش ΔQP فراهم می‌کند؛ بنابراین حتی تغییر اندک در QP می‌تواند منجر به BD-Rate مثبت شود. با این حال، اختلاف معیار MS-SSIM در این توالی‌ها نسبت به نسخه مرجع معمولاً بسیار ناچیز است و نشان می‌دهد که افزایش BD-Rate بر اساس PSNR لزوماً به معنای افت کیفیت ادراکی نیست.

برای تحلیل دقیق‌تر رفتار روش پیشنهادی، سه توالی از کلاس‌های مختلف انتخاب شده و منحنی‌های نرخ–اعوجاج آنها در شکل‌های ۴- الف تا ۴- ج ارائه شده‌اند. این سه نمونه طیف متفاوتی از محتوای ویدئویی شامل رزولوشن بالا، حرکت سریع و بافت‌های حساس را پوشش می‌دهند.

در توالی PeopleOnStreet (کلاس A)، روش پیشنهادی در نرخ‌بیت‌های متوسط تا بالا عملکردی پایدار داشته و صرفه‌جویی نرخ‌بیت همراه با حفظ MS-SSIM به‌دست آمده است. در توالی BasketballDrive (کلاس B)، منحنی MS-SSIM نسخه پیشنهادی در بخش‌هایی از دامنه نرخ‌بیت اندکی بالاتر از نسخه مرجع قرار گرفته و کاهش نرخ‌بیت بیش از ۷ درصد بیانگر مدیریت مؤثرتر تخصیص بیت توسط مدل JND است. توالی BlowingBubbles (کلاس D) که دارای بافت‌های تکراری و حساس به کوانتیزاسیون است، همچنان رفتار پایداری از خود نشان داده و در برخی مقادیر QP کیفیتی مشابه نسخه مرجع همراه با کاهش معنی‌دار نرخ‌بیت ارائه کرده است. همچنین بهبود معیار MS-SSIM در بخش‌هایی از منحنی نشان‌دهنده موفقیت مدل در کنترل نویز ادراکی است.

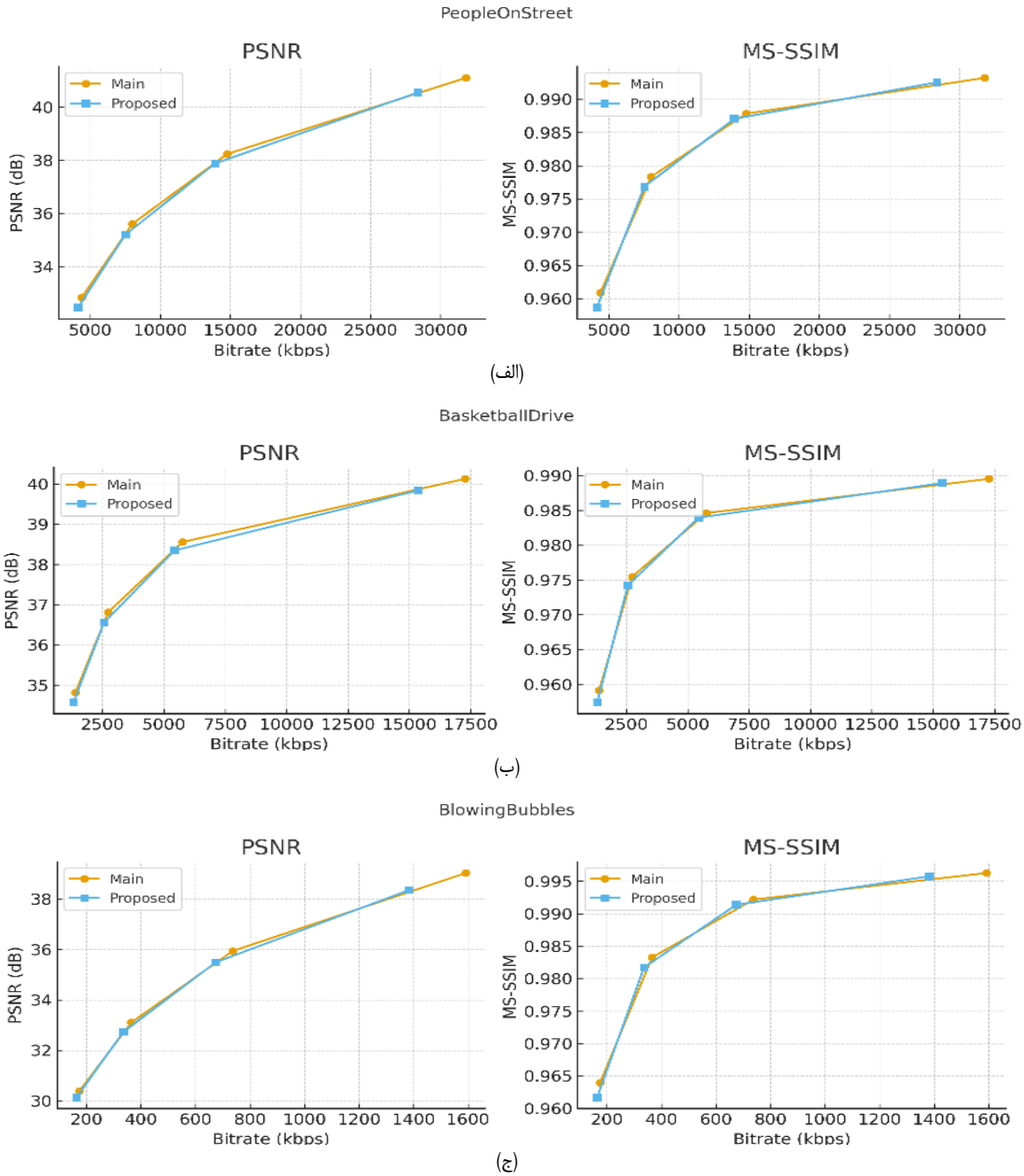
از نظر زمان اجرا نیز روش پیشنهادی در بسیاری از توالی‌ها باعث کاهش زمان کدگذاری شده است. علت اصلی این رفتار، افزایش تطبیقی ΔQP در فریم‌هایی است که آستانه ادراکی آنها بالاتر بوده و نیاز کمتری به کدگذاری دقیق دارند. این افزایش در QP موجب کاهش تعداد بیت‌های تولیدی و ساده‌تر شدن برخی مراحل پردازش شده و در نتیجه

جدول ۳: نتایج عینی برای توالی‌های آزمایشی استاندارد HEVC.

Class	Sequence	BD_BR % (PSNR)	BD_BR % (MS-SSIM)	ΔBR %	ΔT %
A	*PeopleOnStreet 2560x1600_30_f150*	۲٫۸۶	۰٫۰۷	-۷٫۱۲	-۴٫۷۴
	Traffic-150 2560x1600_30_f300	۱٫۳۲	-۰٫۲۰	-۲٫۶۷	-۱٫۰۳
B	*BasketballDrive 1920x1080_50_f500*	۴٫۷۲	۰٫۵۳	-۷٫۱۴	-۲٫۷۲
	ParkScene 1920x1080_24_f240	۱٫۷۹	۰٫۱۳	-۵٫۸۱	-۳٫۸۱
	Cactus 1920x1080_50_f500	۲٫۰۴	۱٫۵۳	-۲٫۰۳	-۰٫۹۱
	Kimono1 1920x1080_24_f240	۲٫۹۰	۱٫۴۹	-۲٫۶۳	-۲٫۴۳
C	*BasketballDrill 832x480_f500*	۱٫۱۲	-۰٫۵۴	-۵٫۰۲	-۲٫۷۴
	PartyScene 832x480_50_f500	۱۷٫۶۹	۱٫۵۲	۲٫۹۷	۱٫۹۹
	BQMall 832x480_60_f600	۲٫۰۷	۰٫۳۰	-۰٫۵۹	۰٫۷۱
	RaceHorses 832x480_30_f300	۳٫۵۷	۲٫۷۶	۷٫۱۰	۳٫۰۷
D	*BasketballPass 416x240_50_f500*	۲٫۲۹	-۰٫۶۲	-۴٫۵۶	-۱٫۲۶
	BlowingBubbles 416x240_50_f500	۱٫۹۱	-۲٫۶۴	-۸٫۹۶	-۳٫۸۹
	BQSquare 416x240_60_f600	۲٫۵۲	-۱٫۱۲	-۱٫۱۱	-۱٫۷۵
	RaceHorses 416x240_30_f300	۱٫۱۶	۳٫۲۴	۸٫۹۰	۴٫۱۲
E	*vidyo1 1280x720p_60_f600*	۱٫۰۳	-۰٫۶۶	-۳٫۵۶	-۵٫۷۷
	vidyo3 1280x_720p_60_f600	۰٫۸۹	-۶٫۸۸	-۷٫۸۶	-4.73
	vidyo4 1280x720p_60_f600	-۳٫۸۰	-۱٫۹۶	-۳٫۵۳	-۲٫۵۵
	Average	2.71	-۰٫۱۸	-۲٫۵۷	-۱٫۶۷

میانگین زمان اجرا در تعداد زیادی از توالی‌ها کاهش یافته است. هرچند محاسبه JND دارای سربار محاسباتی است، اما اثر آن در مقایسه با کاهش زمان اجرای حاصل از افزایش QP بسیار کوچک بوده و در نهایت منجر به کاهش یا ثابت ماندن زمان کل اجرا شده است؛ به‌ویژه در توالی‌های پرجزئیات و پرتحرک که صرفه‌جویی نرخ‌بیت چشمگیرتر است.

در مجموع، نتایج عینی نشان می‌دهند که روش پیشنهادی در مجموعه‌ای متنوع از محتواهای ویدئویی رفتاری پایدار و سازگار از خود نشان داده است؛ به‌گونه‌ای که در بسیاری از توالی‌ها کاهش قابل‌توجهی در نرخ‌بیت حاصل شده و در عین حال کیفیت ادراکی حفظ شده است. اختلاف‌های جزئی مشاهده‌شده در برخی توالی‌ها عمدتاً به ویژگی‌های ساختاری یا حرکتی خاص آن‌ها مربوط است و تأثیر معناداری بر نتیجه‌گیری کلی ندارد. این نتایج تأیید می‌کنند که اعمال ΔQP به‌صورت ادراکی و مبتنی بر مدل *JND* می‌تواند بدون افزایش زمان اجرای کلی کدگذاری، بازده فشرده‌سازی را در بخش وسیعی از محتوای ویدئویی بهبود دهد.



شکل ۴: منحنی D–R بر اساس PSNR و MS-SSIM (الف) توالی PeopleOnStreet ، (ب) توالی BasketballDrive و (ج) توالی BlowingBubbles.

٤-۲ نتایج ذهنی

ارزیابی ذهنی[1] بر پایه آزمون I-MOS انجام شد که با الهام از توصیه ITU-T P.۹۱۰ طراحی شده است. در این ارزیابی، ۱۰ مشاهده‌گر با دید طبیعی یا اصلاح‌شده به طبیعی شرکت کردند. پیش از آغاز آزمون، نحوه انجام ارزیابی برای تمامی شرکت‌کنندگان توضیح داده شد و ترتیب نمایش توالی‌ها به‌صورت تصادفی تعیین گردید. در هر آزمون، ابتدا نسخه اصلی و بدون فشرده‌سازی هر توالی به‌عنوان مرجع بصری نمایش داده شد و سپس نسخه فشرده‌شده (نسخه مرجع HEVC یا روش پیشنهادی) به‌صورت تصادفی ارائه شد..

مشاهده‌گران میزان اعوجاج قابل‌تشخیص در نسخه دوم نسبت به مرجع را بر اساس یک مقیاس پنج‌درجه‌ای امتیازدهی کردند؛ به‌گونه‌ای که امتیاز ۱ بیانگر اعوجاج بسیار آزاردهنده و امتیاز ۵ نشان‌دهنده اعوجاج غیرقابل‌تشخیص بود. امتیازهای ثبت‌شده برای هر توالی میانگین‌گیری شده و به‌عنوان مقدار I-MOS گزارش شدند.

آزمون‌های ذهنی در یک محیط مشاهده کنترل‌شده انجام گرفتند. نمایش ویدئوها بر روی یک نمایشگر با وضوح Full HD (۱۹۲۰×۱۰۸۰) و نرخ نوسازی ثابت صورت پذیرفت. شرایط نوری محیط نیز در سطحی متوسط و یکنواخت نگه داشته شد.

فاصله مشاهده‌گران از نمایشگر مطابق با توصیه استاندارد ITU-T P.۹۱۰ و برابر با حدود سه برابر ارتفاع تصویر تنظیم شد تا شرایط مشاهده با سناریوهای کاربردی واقعی هم‌خوانی داشته باشد. تمامی شرکت‌کنندگان پیش از آغاز آزمون با روند ارزیابی آشنا شدند و ترتیب نمایش توالی‌ها برای هر مشاهده‌گر به‌صورت تصادفی تعیین شد تا اثر خستگی یا یادگیری به حداقل برسد.

1. Subjective Results

جدول ۴: نتایج MOS نسخه مرجع و نسخه پیشنهادی در ΔQPهای ۲۷ و ۳۷.

Class	Sequence	QP	I-MOS-HM	I-MOS-Prop	ΔI-MOS
A	*PeopleOnStreet*	۲۷	۴٫۷۵	۴٫۳۸	۰٫۱۳
		۳۷	۴٫۱۳	۴٫۳۸	۰٫۲۵
	Traffic	۲۷	۴٫۷۵	۴٫۷۵	۰٫۰۰
		۳۷	۴٫۲۵	۴٫۲۵	۰٫۰۰
B	*BasketballDrive*	۲۷	۴٫۳۸	۴٫۵۰	۰٫۱۳
		۳۷	۳٫۲۵	۳٫۲۵	۰٫۰۰
	Cactus	۲۷	۴٫۶۳	۴٫۸۸	۰٫۲۵
		۳۷	۴٫۱۳	۴٫۲۵	۰٫۱۳
	Kimono	۲۷	۴٫۷۵	۴٫۳۸	-۰٫۲۵
		۳۷	۳٫۲۵	۳٫۵۰	۰٫۲۵
	ParkScene	۲۷	۴٫۵۰	۴٫۳۸	-۰٫۱۳
		۳۷	۳٫۵۰	۳٫۳۸	-۰٫۱۳
C	*BQMall*	۲۷	۴٫۲۵	۴٫۶۳	۰٫۳۸
		۳۷	۳٫۵۰	۳٫۶۳	۰٫۱۳
	BasketballDrill	۲۷	۴٫۳۸	۴٫۵۰	۰٫۱۳
		۳۷	۳٫۱۳	۳٫۱۳	۰٫۰۰
	PartyScene	۲۷	۴٫۵۰	۴٫۸۸	۰٫۳۸
		۳۷	۳٫۳۸	۳٫۷۵	۰٫۳۸
	RaceHorses_c	۲۷	۴٫۶۳	۴٫۷۵	۰٫۱۳
		۳۷	۳٫۲۵	۳٫۵۰	۰٫۲۵
D	*BQSquare*	۲۷	۴٫۲۵	۴٫۵۰	۰٫۲۵
		۳۷	۳٫۵۰	۳٫۶۳	۰٫۱۳
	BasketballPass	۲۷	۳٫۸۸	۴٫۲۵	۰٫۳۸
		۳۷	۲٫۸۸	۳٫۲۵	۰٫۳۸
	BlowingBubbles	۲۷	۴٫۲۵	۴٫۲۵	۰٫۰۰
		۳۷	۳٫۵۰	۳٫۷۵	۰٫۲۵
	RaceHorses_d	۲۷	۴٫۰۰	۴٫۳۸	۰٫۳۸
		۳۷	۳٫۳۸	۳٫۶۳	۰٫۲۵
E	*vidyo1*	۲۷	۴٫۳۸	۴٫۵۰	۰٫۱۳
		۳۷	۳٫۰۰	۳٫۰۰	۰٫۰۰
	vidyo3	۲۷	۴٫۰۰	۴٫۳۸	۰٫۳۸
		۳۷	۳٫۰۰	۳٫۱۳	۰٫۱۳
	vidyo4	۲۷	۴٫۰۰	۴٫۲۵	۰٫۲۵
		۳۷	۳٫۱۳	۳٫۳۸	۰٫۲۵
Average			۳٫۸۹	۴٫۰۵	۰٫۱۷

مقادیر I-MOS به‌دست‌آمده برای تمامی توالی‌های آزمایشی در جدول ۴ گزارش شده است. نتایج نشان می‌دهد که روش پیشنهادی از نظر کیفیت ادراکی عملکردی بسیار نزدیک به نسخه مرجع دارد و در بسیاری از موارد، اندکی بهتر نیز ظاهر شده است. میانگین I-MOS نسخه مرجع برابر با ۳٫۸۹ و برای نسخه پیشنهادی برابر با ۴٫۰۵ است؛ به‌عبارت دیگر، نسخه پیشنهادی به‌طور میانگین ΔI-MOS=+۰٫۱۷ بهبود کیفیت ادراکی داشته است. تحلیل جزئی‌تر نتایج نشان می‌دهد که در ۱۲ توالی از ۱۷ توالی، نسخه پیشنهادی امتیاز بالاتری کسب کرده و در ۳ توالی امتیاز دو نسخه برابر بوده است. بیشترین بهبود کیفیت مربوط به توالی‌های BQMall، PartyScene، BasketballPass، RaceHorses_d و vidyo3 است که در آن‌ها ΔI-MOS حدود ۰٫۳۸+ گزارش شده است.

نکته قابل توجه آن است که در برخی توالی‌ها، از جمله PartyScene و RaceHorses، نسخه پیشنهادی نرخ‌بیت بالاتری نسبت به نسخه مرجع مصرف کرده است، در حالی‌که مقادیر MOS نیز افزایش یافته‌اند. این رفتار نشان می‌دهد که مدل JND در نواحی بصری حساس، افزایش ΔQP را به‌صورت مؤثر محدود کرده و بدین ترتیب از تضعیف جزئیات مهم تصویر جلوگیری نموده است. در نتیجه، افزایش محدود نرخ‌بیت در این موارد به افت کیفیت ادراکی منجر نشده و حتی به بهبود تجربه بصری مشاهده‌گران انجامیده است.

از سوی دیگر، افت‌های محدود مشاهده‌شده در دو توالی (در حد چند دهم واحد I-MOS) بسیار کوچک بوده و در محدوده نوسانات معمول ارزیابی‌های ذهنی قرار می‌گیرند. افزون بر این، این افت‌ها در مقادیر مختلف QP پایدار نبوده‌اند. در مقابل، در تعداد بیشتری از توالی‌ها بهبود I-MOS مشاهده شده است که نشان می‌دهد روش پیشنهادی در مجموع کیفیت ادراکی را حفظ کرده و در برخی محتواها به بهبود جزئی نیز منجر شده است.

به‌طور خلاصه، ترکیب نتایج عینی و ذهنی نشان می‌دهد که روش پیشنهادی توانسته است بدون ایجاد تغییر محسوس در کیفیت ادراکی، کاهش قابل‌توجهی در نرخ‌بیت و زمان اجرا ایجاد کند. اگرچه معیار PSNR در برخی توالی‌ها افزایش BD-Rate را گزارش می‌کند، معیار MS-SSIM و به‌ویژه نتایج I-MOS نشان می‌دهند که این افزایش از نظر ادراکی قابل‌تشخیص نبوده و بر کیفیت بصری اثر منفی ندارد. در بسیاری از توالی‌ها، نسخه پیشنهادی حتی امتیاز بالاتری نسبت به نسخه مرجع کسب کرده است؛ موضوعی که بیانگر آن است که مدل JND در سطح فریم توانسته نواحی کم‌حساس را به‌درستی شناسایی کرده و با تنظیم کنترل‌شده ΔQP، بازده فشرده‌سازی را بهبود دهد. در مجموع، روش پیشنهادی رفتاری پایدار، با پیچیدگی(زمان اجرا) کمتر و سازگار با HEVC ارائه می‌دهد و می‌تواند بدون ایجاد تغییر در ساختار کدگذار، بهبودهای مؤثری در کارایی ادراکی و محاسباتی فراهم سازد. همچنین، مقایسه مستقیم با روش‌های جدید مبتنی بر JND و توسعه روش پیشنهادی با بهره‌گیری از ایده‌های آن‌ها، از موضوعات قابل بررسی در پژوهش‌های آینده خواهد بود.

۵- جمع‌بندی و نتیجه‌گیری

در این پژوهش یک روش ادراکی برای تنظیم تطبیقی ΔQP مبتنی بر مدل آستانه تفاوت قابل‌تشخیص (JND) در کدگذار HEVC ارائه شد. هدف روش پیشنهادی آن بود که کیفیت بصری حفظ شود و هم‌زمان بازده فشرده‌سازی و کارایی محاسباتی بهبود یابد. نتایج شبیه‌سازی نشان داد که الگوریتم در بخش قابل‌توجهی از توالی‌ها کاهش نرخ‌بیت ایجاد کرده و در بسیاری از موارد، کیفیت ادراکی بالاتر یا برابر با نسخه مرجع ارائه داده است. این امر ناشی از تنظیم ΔQP بر اساس حساسیت ادراکی فریم‌هاست؛ به‌طوری‌که فریم‌های حساس‌تر با QP ملایم‌تر و فریم‌های کم‌حساسیت‌تر با QP بزرگ‌تر کدگذاری می‌شوند. همچنین نتایج I-MOS تأیید کرد که حتی در مواردی که بهبود عینی محدود بوده، نسخه پیشنهادی از دید مشاهده‌گران کیفیت بهتری داشته است. علاوه بر این، روش پیشنهادی به‌طور متوسط ۱٫۶۷٪ کاهش زمان اجرا را بدون

اعمال تغییر در ساختار اصلی HEVC به همراه دارد. مجموع نتایج عینی و ذهنی نشان می‌دهد که روش پیشنهادی می‌تواند رویکردی با پیچیدگی پایین، پایدار و سازگار برای بهبود بازده فشرده‌سازی و کیفیت بصری باشد.

مراجع

[1] Cisco, *Cisco Annual Internet Report* (2018–2023), White Paper, Mar. 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/executive-perspectives/annual-internet-report/index.html

[2] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[3] G. Wang et al., "A survey on just noticeable distortion estimation and its applications in video coding," *Journal of Visual Communication and Image Representation*, vol. 98, Article ID: 104034, Feb. 2024.

[4] K. McCann, *et al.*,, "High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) encoder description," JCT-VC R1002, *18th Meeting of JCT-VC*, San DiegoCA, USA, 22-26 Feb. 2014. Available: https://mpeg.chiariglione.org

[5] Y. Xu, S. Yi, L. Lin, W. Chen, and T. Zhao, "GOP structure-independent quantization parameter cascading in video coding," *IEEE Access*, vol. 7, pp. 76274-76282, 2019.

[6] Y. Gong, *et al.*, "Quantization parameter cascading for surveillance video coding considering all inter reference frames," *IEEE Trans. Image Processing*, vol. 30, pp. 5692-5707, 2021.

[7] X. Huang, D. Li, and H. Yin, "HEVC quantization parameter selection algorithm based on inter-frame dependency," *Multimedia Tools and Applications*, vol. 79, no. 19, pp. 13951-13966, 2020.

[8] H. Amer and E. H. Yang, "Adaptive quantization parameter selection for low-delay HEVC via temporal propagation length estimation," *Signal Processing: Image Communication*, vol. 84, Article ID: 115826, May 2020.

[9] L. Jia, K. Jia, and X. Fan, "Adaptive Lagrangian multiplier for quantization parameter cascading in HEVC hierarchical coding," *IEEE Signal Processing Letters*, vol. 27, pp. 1220=1224, 2020.

[10] I. Marzuki and D. Sim, "Perceptual adaptive quantization parameter selection using deep convolutional features for HEVC encoder," *IEEE Access*, vol. 8, pp. 37052-37065, 2020.

[11] Y. Wu, *et al.*, "Perceptual VVC quantization refinement with ensemble learning," *Displays*, vol. 70, Article ID: 102103, 2021.

[12] H. Wang *et al.,* "Surprise-based JND estimation for perceptual quantization in H.265/HEVC codecs," *Signal Processing: Image Communication*, vol. 118, Article ID: 117019, Oct. 2023.

[13] H. Wang *et al.*, "Learning-based JNCD prediction for quality-wise perceptual quantization in HEVC," *Journal of Visual Communication and Image Representation*, vol. 95, Article ID: 103877, Sept.2023.

[14] A. J. Ruiz, G. O. López, M. M. Pérez, and M. O. Martínez-Rach, "Perceptual QP optimization for VVC with dual hybrid neural networks," *The Journal of Supercomputing*, vol. 81, no. 3, Article ID: 464, 2025.

[15] W. Lim and D. Sim, "A perceptual rate control algorithm based on luminance adaptation for HEVC encoders," *Signal Image Video Processin*g, vol. 14, no. 5, pp. 887-895, 2020.

[16] M. Zhou, X. Wei, S. Kwong, W. Jia, and B. Fang, "Just noticeable distortion-based perceptual rate control in HEVC," *IEEE Trans. Image Process*ing, vol. 29, pp. 7603-7614, 2020.

[17] K. C. Liu, "Color video JND model using compound spatial masking and structure-based temporal masking," *IEEE Access*, vol. 8, pp. 136760-136768, 2020.

[18] H. Jo, S. Lee, Y. Ahn, Y. Kim, and D. Sim, "Perceptual intra-frame coding for HEVC still picture profile based on invisible signal suppression," *Signal Image Video Processin*g, vol. 14, pp. 1053-1061, 2020.

[19] J. Guo, "An HEVC-compliant perceptual video coding using just noticeable difference," *Multimedia Tools and Applications*, vol. 81, pp. 1257-1286, 2022.

[20] S. Nami, F. Pakdaman, and M. R. Hashemi, "Juniper: A JND-based perceptual video coding framework to jointly utilize saliency and JND," in *Proc. IEEE Int. Conf.* Multimedia Expo Workshops, 6 pp., Londom, UK, 6-10 Jul. 2020.

[21] S. Ki, J. Do, and M. Kim, "Learning-based JND-directed HDR video preprocessing for perceptually lossless compression with HEVC," *IEEE Access*, vol. 8, pp. 228605-228618, 2020.

[22] J. Kim, D. Y. Lee, S. Jeong, and S. Cho, "Perceptual video coding using deep neural network based JND model," in *Proc. Data Compression Conf.*, pp. 375-375, Snowbird, UT, USA, 24-27 Mar. 2020.

[23] D. Fu, Y. Wang, B. Fan, and N. Ding, "HEVC/H.265 intra coding based on the human visual system," *IEEE Access*, vol. 7, pp. 186587-186599, 2019.

[24] H. Wang, *et al.*, "JND-based multi-module cooperative perceptual optimization for HEVC," *Displays*, vol. 84, Article ID: 102783, Sept. 2024.

[25] Y. Zhang, *et al.*, "SJ-PVC: An efficient perceptual video compression scheme based on adaptive QP and rate-distortion optimization," *IEEE Trans. Consumer Electronics,* col. 71, no. 1, pp. 706-719, Feb. 2025.

[26] ITU-T, *Common Test Conditions for HEVC*, JCT-VC Document, 2016.

[27] Z. Y. Wei and K. N. Ngan, "Spatio-temporal just noticeable distortion profile for gray-scale image/video in DCT domain," *on Circuits and Systems for Video Technology*, vol. 19, no. 3, pp. 337-346, Mar. 2009.

[28] G. Bjøntegaard, *Calculation of Average PSNR Differences between RD-Curves*, VCEG-M33, Apr. 2001.

[29] Z. Wang, *et al.*, "Multiscale structural similarity for image quality assessment," in *Proc. The Thrity-Seventh Asilomar Conf. on Signals, Systems & Computers*, pp. 1398-1402.Pacific Grove, CA, USAm 9-12 Nov. 2003.

[30] ITU-T, *Subjective Video Quality Assessment Methods for Multimedia Applications*, ITU-T Rec. P.910, 2008, rev. 2021.

مریم کلانتری در سال ۱۳۹۶ مقطع کارشناسی خود را در رشته مهندسی برق الکترونیک در دانشگاه سیستان و بلوچستان به پایان رسانده است. وی مدرک کارشناسی ارشد را نیز در رشته مهندسی برق مخابرات سیستم از دانشگاه فردوسی مشهد در سال ۱۴۰۵ اخذ کرده است. پژوهش‌های مورد علاقه ایشان در زمینه فشرده سازی ویدئو است.

عباس ابراهیمی مقدم مدرک کارشناسی برق گرایش الکترونیک خود را از دانشگاه صنعتی شریف در سال ۱۳۷۰ و در سال ۱۳۷۴ مدرک کارشناسی ارشد را از دانشگاه صنعتی خواجه نصیر اخذ کرده است. ایشان مدرک دکتری خود را از دانشگاه McMaster کانادا دریافت کرده است و از سال ۱۳۹۰ به عنوان استادیار در دانشگاه فردوسی مشهد فعالیت علمی می‌نماید. زمینه‌های تحقیقاتی مورد علاقه ایشان پردازش گفتار، پردازش تصویر و ویدیو، بینایی ماشین و پردازش سیگنال‌های حیاتی می‌باشد.

مرتضی خادمی درح تحصیلات خود را در مقاطع کارشناسی و کارشناسی ارشد مهندسی برق به‌ترتیب در سال‌های ۱۳۶۴ و ۱۳۶۶ در دانشگاه صنعتی اصفهان به پایان رسانده است. نام‌برده از سال ۱۳۶۶ الی ۱۳۷۰ به عنوان عضو هیات علمی (مربی) در دانشگاه فردوسی مشهد به کار مشغول بود. پس از آن به دوره دکترای مهندسی برق در دانشگاه ولونگونگ (استرالیا) وارد گردیده و در سال ۱۳۷۴ موفق به اخذ درجه دکترا در مهندسی برق از دانشگاه مذکور گردید. دکتر خادمی از سال ۱۳۷۴ مجددا در دانشکده مهندسی دانشگاه فردوسی مشهد مشغول به فعالیت گردید و اینک نیز استاد این دانشکده است. زمینه‌های علمی مورد علاقه ایشان شامل موضوعاتی مانند مخابرات ویدئویی، فشرده‌سازی ویدئو، پردازش تصویر، پردازش سیگنال‌های پزشکی و پنهان‌سازی اطلاعات در ویدئو می‌باشد.