



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده مهندسی برق و کامپیوتر

چارچوبی انطباق‌پذیر برای کدکننده مقیاس‌پذیر ویدئوی چندنمایی با استاندارد H.264/AVC

نگارش:
هدا رودکی لواسانی

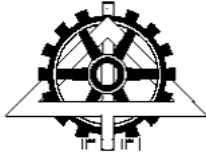
استاد راهنما:
دکتر محمودرضا هاشمی

استاد مشاور:
دکتر شروین شیرمحمدی

رساله برای دریافت درجه دکتری در رشته
مهندسی کامپیوتر – گرایش معماری کامپیوتر

شهریور ۱۳۹۳

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده مهندسی برق و کامپیوتر

چارچوبی انطباق‌پذیر برای کدکننده مقیاس‌پذیر ویدئوی چندنمایی با استاندارد H.264/AVC

نگارش:
هدا رودکی لواسانی

استاد راهنما:
دکتر محمودرضا هاشمی

استاد مشاور:
دکتر شروین شیرمحمدی

رساله برای دریافت درجه دکتری در رشته
مهندسی کامپیوتر – گرایش معماری کامپیوتر

شهریور ۱۳۹۳



دانشگاه تهران

پردیس دانشکده های فنی

دانشکده مهندسی برق و کامپیوتر

پایان نامه برای دریافت درجه دکتری در رشته مهندسی کامپیوتر

عنوان: چارچوبی انطباق پذیر برای کدکننده مقیاس پذیر ویدئوی چندنمایی با استاندارد H.264/AVC

نگارش: هدا رودکی لواسانی

این پایان نامه در تاریخ ۱۳۹۳/۶/۲۴ در مقابل هیأت داوران دفاع گردید و مورد تصویب قرار گرفت.

معاون آموزشی و تحصیلات تکمیلی پردیس دانشکده های فنی: دکتر علی افضلی کوشا

رئیس دانشکده مهندسی برق و کامپیوتر: دکتر شاهرخ فرهنگی

معاون پژوهشی و تحصیلات تکمیلی دانشکده مهندسی برق و کامپیوتر: دکتر سید کمال الدین ستاره دان

استاد راهنما: دکتر محمودرضا هاشمی

استاد مشاور: دکتر شروین شیرمحمدی

عضو هیأت داوران: دکتر حمیدرضا ربیعی

عضو هیأت داوران: دکتر شهره کسایی

عضو هیأت داوران: دکتر محمد قنبری

عضو هیأت داوران: دکتر سید امید فاطمی

تعهد نامه اصالت اثر

اینجانب هدا رودکی لواسانی تأیید می‌کنم که مطالب مندرج در این پایان نامه حاصل کار پژوهشی اینجانب است و به دستاوردهای پژوهشی دیگران که در این نوشته از آنها استفاده شده است مطابق مقررات ارجاع گردیده است. این پایان نامه قبلاً برای احراز هیچ مدرک هم سطح یا بالاتر ارائه نشده است.

کلیه حقوق مادی و معنوی این اثر متعلق به دانشکده فنی دانشگاه تهران می‌باشد.

نام و نام خانوادگی دانشجو : هدا رودکی لواسانی

امضای دانشجو :

تشکر و قدردانی:

لازم می‌دانم از زحمات جناب آقای دکتر محمودرضا هاشمی که راهنمایی بنده را در این دوره پنج ساله بر عهده داشتند، صمیمانه تشکر و سپاسگزاری نمایم. در مراحل مختلف رساله، از تعیین موضوع گرفته تا انجام تحقیقات و ارزیابی آنها تا تهیه مقالات و نوشتن رساله از راهنمایی‌های ایشان استفاده شایانی نمودم.

استاد مشاور این رساله جناب آقای دکتر شروین شیرمحمدی نیز کمک‌های شایانی در انجام مراحل مختلف نمودند.

از اساتید داور چه در مرحله پیشنهاد رساله و چه در مرحله دفاع نهایی که با راهنمایی‌های خود مسیر مناسب‌تری پیش روی اینجانب گذاشتند نیز کمال تشکر و امتنان را دارم.

نهایتاً از دوستان خود در آزمایشگاه سیستم‌های چندرسانه‌ای در دانشگاه تهران که با همفکری‌های خود در مسائل مختلف مساعدت نمودند، تشکر می‌نمایم.

چکیده

با گسترش درخواست برای ویدئوی چندنمایی سه بعدی در حوزه‌های گوناگون، تحقیقات در زمینه استفاده از این فناوری در کاربردهای مختلف مورد توجه قرار گرفته است. مقیاس پذیری یک راه حل مناسب برای کاهش حجم بالای داده‌های ویدئوی چندنمایی سه بعدی است، اما استفاده از آن برای ویدئوهای چندنمایی هنوز مشکلات حل نشده زیادی دارد.

هدف از انجام این رساله، ارائه یک چارچوب برای استفاده انطباق پذیر از مقیاس پذیری در ویدئوی چندنمایی سه بعدی است. در چارچوب مورد نظر، بر خلاف کارهای قبلی که تلاشی برای استخراج انواع مناسب مقیاس پذیری برای ویدئوی چندنمایی سه بعدی بسته به کاربردهای جدید و به صورت روش مند انجام نشده است، مقیاس پذیری‌های مناسب برای ویدئوی چندنمایی مقیاس پذیر تعریف و بسته به نیازهای کاربرد مورد نظر انتخاب می‌شود. در ادامه، یک روش شناسی برای تخصیص داده‌های موجود به لایه پایه و لایه‌های افزایشی پیشنهاد شده است. برخلاف روش‌های قبلی تخصیص داده‌ها، روش ارائه شده با در نظر گرفتن مفاهیم اختلاف بین لایه‌ای و درون لایه‌ای که بر اساس نوع مقیاس پذیری و کاربرد مورد نظر تعریف می‌شوند، به تخصیص داده‌ها به لایه‌های مختلف پرداخته است. با استفاده از این روش پهنای باند مورد نیاز برای نماهای مختلف به طور میانگین 0.5% کاهش یافته و کیفیت تجربی نیز به طور میانگین 2% افزایش یافته است. قسمت بعدی چارچوب پیشنهادی به اختصاص بهینه نرخ بیت به لایه‌های مختلف و محتوای آنها می‌پردازد. به این منظور و به عنوان پیش نیاز ابتدا یک مدل نرخ بیت - اعوجاج مناسب در سطح نما برای ویدئوی چندنمایی سه بعدی به صورت روش مند استخراج شده است. مدل‌های پیشنهاد شده قبلی اثر ویژگی‌های خاص ویدئوی چندنمایی مانند مفاهیم اختلاف بین نمایی و درون نمایی را در نظر نمی‌گیرند. نتایج بررسی‌ها نشان می‌دهد که با در نظر گرفتن این مفاهیم، مدل نرخ بیت - اعوجاج پیشنهادی می‌تواند به انواع مختلف مقیاس پذیری تعمیم داده شده و نرخ بیت هر نما را با میانگین خطای 12% پیش بینی کند. از آنجا که معیار کارائی الگوریتم اختصاص نرخ بیت، کیفیت نهایی ویدئوی حاصل است و نیز اینکه تا کنون معیار مناسبی برای ارزیابی کمی کیفیت ویدئوی چندنمایی مقیاس پذیر ارائه نشده است، در این بخش از چارچوب پیشنهادی، یک روش شناسی به منظور استخراج معیار کمی ارزیابی کیفیت برای این ویدئوها نیز ارائه شده است. نتایج ارزیابی‌ها نشان می‌دهد که معیار ارزیابی کمی استخراج شده با نتایج ارزیابی بصری همبستگی بالایی داشته و ضریب همبستگی آنها به طور میانگین برای ویدئوهای آزمایش شده 0.8 است. در نهایت و با توجه به عدم وجود الگوریتم اختصاص بهینه نرخ بیت برای ویدئوی چندنمایی سه بعدی مقیاس پذیر، الگوریتمی برای تخصیص نرخ بیت موجود به نماها و لایه‌های مختلف با استفاده از مفاهیم اختلاف بین نمایی و درون نمایی که بر اساس مقیاس پذیری انتخابی و کاربرد مورد نظر تعریف شده‌اند، ارائه می‌شود. این الگوریتم می‌تواند نرخ بیت موجود را به ترتیب با کارایی فشرده سازی بالای 24% و 42% نسبت به حالت مرجع استاندارد H.264، به نماهای لایه پایه و افزایشی تخصیص دهد.

فهرست مطالب

۱- مقدمه	۱۹
۱-۱- مقیاس پذیری در ویدئوی چندنمایی	۲۳
۱-۲- بیان مسأله	۲۴
۱-۲-۱- استخراج مقیاس پذیری مناسب بسته به کاربرد موردنظر	۲۶
۱-۲-۲-۱- اختصاص مناسب داده‌ها به لایه‌های پایه و افزایشی	۲۷
۱-۲-۳-۱- مدل نرخ بیت - اعوجاج مناسب برای ویدئوی چندنمایی	۳۰
۱-۲-۴- تخمین کیفیت در ویدئوی چندنمایی مقیاس پذیر	۳۰
۱-۲-۵- اختصاص بهینه نرخ بیت به لایه‌های پایه و افزایشی و محتوای آنها	۳۱
۱-۳- ساختار رساله	۳۲
۱-۴- جمع بندی	۳۲
۲- پیشینه تحقیق	۳۴
۲-۱- مقیاس پذیری در ویدئوی چندنمایی	۳۵
۲-۲- روش‌های ارائه شده برای اختصاص داده‌ها به لایه‌های پایه و افزایشی در ویدئوی چندنمایی	۳۶
۲-۳- مدل‌های نرخ بیت - اعوجاج برای ویدئوی چندنمایی	۳۷
۲-۴- روش‌های ارائه شده برای تخمین کیفیت در ویدئوی چندنمایی	۳۹
۲-۵- روش‌های ارائه شده برای بهینه‌سازی نرخ بیت - اعوجاج در ویدئوی چندنمایی	۴۳
۲-۶- جمع بندی	۴۵
۳- روش‌شناسی ارائه شده برای استخراج مقیاس پذیری مناسب برای ویدئوی چندنمایی بسته به کاربرد موردنظر	۴۷
۳-۱- نظریه مبنایی	۴۸
۳-۱-۱- مرحله بازکردن اطلاعات	۴۹
۳-۱-۲- مرحله کشف ارتباطات بین داده‌ها	۵۰
۳-۱-۳- مرحله انتخابی	۵۰

۲-۳- اعمال نظریه مبنایی به منظور استخراج انواع مقیاس‌پذیری‌های مناسب برای ویدئوی چندنمایی سه‌بعدی	۵۰
۱-۲-۳- مرحله بازکردن اطلاعات	۵۱
۲-۲-۳- مرحله کشف ارتباطات بین داده‌ها	۵۲
۳-۲-۳- مرحله انتخابی	۵۲
۳-۳- تعاریف استخراج‌شده برای مقیاس‌پذیری در ویدئوی چندنمایی و لزوم استفاده از آنها در کاربردهای واقعی	۵۳
۱-۳-۳- مقیاس‌پذیری عمق	۵۴
۱-۱-۳-۳- مثالی از کاربرد مقیاس‌پذیری عمق	۵۵
۲-۳-۳- مقیاس‌پذیری پیچیدگی	۵۷
۱-۲-۳-۳- مثالی از کاربرد مقیاس‌پذیری پیچیدگی	۵۸
۳-۳-۳- مقیاس‌پذیری سطح	۵۸
۱-۳-۳-۳- مثالی از کاربرد مقیاس‌پذیری سطح	۵۹
۴-۳-۳- مقیاس‌پذیری عمق-وضوح تصویر/عمق-کیفیت تصویر	۵۹
۱-۴-۳-۳- مثالی از کاربرد مقیاس‌پذیری عمق-وضوح تصویر/عمق-کیفیت تصویر	۶۰
۵-۳-۳- مقیاس‌پذیری اطلاعات جانبی	۶۰
۱-۵-۳-۳- مثالی از کاربرد مقیاس‌پذیری اطلاعات جانبی	۶۱
۶-۳-۳- مقیاس‌پذیری کیفیت ادراکی با استفاده از نظریه محوشدگی	۶۱
۴-۳- جمع‌بندی	۶۲
۴- روش‌شناسی ارائه‌شده برای اختصاص داده‌های مناسب به لایه‌های پایه و افزایشی در ویدئوی چندنمایی بر اساس کاربرد مورد نظر	۶۳
۱-۴- روش‌شناسی پیشنهادی	۶۴
۱-۱-۴- تعاریف و مشاهدات	۶۴
۲-۱-۴- مراحل روش‌شناسی پیشنهادی	۶۵

۲-۴- اعمال روش‌شناسی پیشنهادی برای کاربردهای موبایل با محدودیت پهنای باند و توان پردازشی	۶۷
۴-۲-۱- ارزیابی روش‌شناسی ارائه‌شده برای اختصاص داده‌ها به لایه‌های پایه و افزایشی برای کاربردهای موبایل	۶۹
۴-۳- اعمال روش‌شناسی پیشنهادی برای کاربردهای نمای آزاد با قابلیت کنترل تعاملی محدوده دید توسط کاربر	۷۴
۴-۳-۱- ارزیابی روش‌شناسی ارائه‌شده برای اختصاص داده‌ها با لایه‌های پایه و افزایشی برای کاربردهای ویدئوی نمای آزاد	۷۵
۴-۴- جمع‌بندی	۷۷
۵- روش‌شناسی ارائه‌شده برای استخراج مدل نرخ بیت - اعوجاج برای ویدئوی چندنمایی	۷۸
۵-۱- لزوم ارائه مدل نرخ بیت - اعوجاج برای ویدئوی چندنمایی سه‌بعدی و نوآوریهای روش ارائه‌شده	۷۹
۵-۲- گام‌های روش‌شناسی پیشنهادی	۸۲
۵-۲-۱- گام اول: استخراج مدل نرخ بیت - اعوجاج بر اساس اختلاف درون‌نمایی	۸۴
۵-۲-۱-۱- استخراج پارامترهای مؤثر برای توصیف اختلاف درون‌نمایی	۸۴
۵-۲-۱-۱-۱- استخراج ارتباط بین نرخ بیت هر نما و پارامترهای شاخص اختلاف درون‌نمایی	۸۵
۵-۲-۲- گام دوم: استخراج مدل نرخ بیت - اعوجاج بر اساس اختلاف بین‌نمایی	۸۵
۵-۲-۲-۱- استخراج پارامترهای مؤثر برای توصیف اختلاف بین‌نمایی	۸۵
۵-۲-۲-۱-۱- استخراج ارتباط بین نرخ بیت هر نما و پارامترهای شاخص اختلاف بین‌نمایی	۸۵
۵-۲-۳- مرحله سوم: مدل نرخ بیت - اعوجاج نهایی در سطح نما برای ویدئوی چندنمایی سه‌بعدی	۸۶
۵-۳- استخراج مدل نرخ بیت - اعوجاج در سطح نما برای ویدئوی چندنمایی سه‌بعدی بر اساس روش‌شناسی ارائه‌شده	۸۶

۵-۳-۱- مرحله اول: استخراج مدل نرخ بیت - اعوجاج بر اساس اختلاف درون‌نمایی	۸۷
۵-۳-۱-۱- استخراج پارامترهای مؤثر برای توصیف اختلاف درون‌نمایی	۸۷
۵-۳-۱-۱-۱- استخراج ارتباط بین نرخ بیت هر نما و پارامترهای شاخص اختلاف درون‌نمایی	۸۸
۵-۳-۲- مرحله دوم: استخراج مدل نرخ بیت - اعوجاج بر اساس اختلاف بین‌نمایی	۹۶
۵-۳-۲-۱- استخراج پارامترهای مؤثر برای توصیف اختلاف بین‌نمایی	۹۶
۵-۳-۲-۱-۱- استخراج ارتباط بین نرخ بیت هر نما و پارامترهای شاخص اختلاف درون‌نمایی	۹۹
۵-۳-۳- مرحله سوم: مدل نهایی نرخ بیت - اعوجاج در سطح نما برای ویدئوی چندنمایی	۱۰۶
۵-۴- ارزیابی مدل پیشنهادی	۱۰۷
۵-۵- جمع‌بندی	۱۱۴
۶- روش‌شناسی ارائه‌شده برای استخراج معیار ارزیابی کمی کیفیت در ویدئوی چندنمایی مقیاس‌پذیر	۱۱۵
۶-۱- روش‌شناسی پیشنهادی	۱۱۶
۶-۱-۱- تعاریف و مشاهدات	۱۱۶
۶-۱-۲- مراحل روش‌شناسی پیشنهادی	۱۱۷
۶-۱-۲-۱- ارزیابی کیفیت برای هر لایه در ویدئوی چندنمایی مقیاس‌پذیر	۱۱۹
۶-۱-۲-۱-۱- ارزیابی کیفیت برای هر لایه در مقیاس‌پذیری نامایی	۱۲۰
۶-۱-۲-۱-۲- ارزیابی کیفیت برای هر لایه در مقیاس‌پذیری عمق	۱۲۲
۶-۲-۱-۲- ارزیابی کیفیت نهایی برای ویدئوی چندنمایی مقیاس‌پذیر	۱۲۳
۶-۲- ارزیابی روش ارائه‌شده برای تخمین کمی کیفیت ویدئوی چندنمایی مقیاس‌پذیر	۱۲۶

چندنمایی مقیاس پذیر	۱۲۷
۶-۲-۱- ارزیابی روش ارائه شده برای تخمین کمی کیفیت هر یک از لایه های ویدئوی	
چندنمایی مقیاس پذیر برای مقیاس پذیری نمایی	۱۲۸
۶-۲-۱-۲- ارزیابی روش ارائه شده برای تخمین کمی کیفیت هر یک از لایه های	
ویدئوی چندنمایی مقیاس پذیر برای مقیاس پذیری عمق	۱۳۱
۶-۲-۲- ارزیابی روش ارائه شده برای تخمین کمی کیفیت نهایی ویدئوی چندنمایی	
مقیاس پذیر	۱۳۲
۶-۲-۲-۱- ارزیابی روش ارائه شده برای تخمین کمی کیفیت نهایی ویدئوی	
چندنمایی مقیاس پذیر برای مقیاس پذیری نمایی	۱۳۳
۶-۲-۲-۲- ارزیابی روش ارائه شده برای تخمین کمی کیفیت نهایی ویدئوی	
چندنمایی مقیاس پذیر برای مقیاس پذیری عمق	۱۳۶
۶-۳- فرآیند ارزیابی بصری	۱۳۷
۶-۴- جمع بندی	۱۳۹
۷- روش ارائه شده برای بهینه سازی نرخ بیت - اعوجاج در ویدئوی چندنمایی مقیاس پذیر	
.....	۱۴۰
۷-۱- بهینه سازی نرخ بیت - اعوجاج در ویدئوی چندنمایی مقیاس پذیر	۱۴۱
۷-۲- روش پیشنهادی	۱۴۲
۷-۲-۱- بهینه سازی نرخ بیت - اعوجاج برای لایه های پایه و افزایشی ویدئوی	
چندنمایی مقیاس پذیر در مقیاس پذیری نمایی	۱۴۳
۷-۲-۲- بهینه سازی نرخ بیت - اعوجاج برای هر یک از نماهای لایه های پایه و	
افزایشی در مقیاس پذیری نمایی	۱۴۶
۷-۳- ارزیابی روش پیشنهادی	۱۴۸
۷-۴- جمع بندی	۱۵۴
۸- نتیجه گیری و کارهای آینده	۱۵۵
۸-۱- کارهای آینده	۱۵۷

۹- مراجع ۱۵۹

فهرست اشکال

- شکل ۱-۱- مثالی از یک سامانه تلویزیون نمای آزاد [4]..... ۲۱
- شکل ۲-۱ - مثالی از یک سامانه ویدئوی چندنمایی مقیاس پذیر. (الف) تبلت با ۴ نمای دریافتی (ب) دستگاه موبایل با دریافت ویدئوی سه بعدی برجسته نما با یک نما به اضافه اطلاعات عمق ۲۴
- شکل ۳-۱ - ساختار کلی چارچوب پیشنهادی در این رساله ۲۶
- شکل ۴-۱ - انواع مختلف مقیاس پذیری (الف) مقیاس پذیری مکانی/زمانی/کیفیت برای ویدئوی تک‌نمایی (ب) مقیاس پذیری نمایی برای ویدئوی چندنمایی (ج) مقیاس پذیری سازگار با فریم برای ویدئوی برجسته نما (د) مقیاس پذیری نمای آزاد برای ویدئوی چندنمایی سه بعدی ۲۸
- شکل ۵-۱ - ساختار پیش‌بینی معمول در استاندارد H.264/AVC [15] ۲۹
- شکل ۱-۳ - دسته‌ها، زیردسته‌ها و هسته اصلی که در گام‌های مختلف نظریه مبنایی به دست آمده‌اند. ۵۲
- شکل ۲-۳ - مقیاس‌پذیری عمق برای کلاس آموزشی (الف) لایه پایه (ب) لایه‌های پایه و افزایشی ۵۶
- شکل ۳-۳ - ساختار پیش‌بینی مناسب برای (الف) لایه پایه و (ب) لایه‌های افزایشی در مقیاس‌پذیری پیچیدگی ۵۷
- شکل ۴-۳ - (الف) لایه‌های پایه و افزایشی در ویدئوی برجسته در مقیاس‌پذیری عمق - وضوح تصویر (ب) (لایه‌های پایه و افزایشی در ویدئوی برجسته در مقیاس‌پذیری عمق - کیفیت تصویر ۶۰
- شکل ۱-۴ - مثالی از ساختار پیش‌بینی استفاده شده برای هر لایه ۶۶
- شکل ۲-۴ - مراحل مختلف روش‌شناسی ارائه شده برای تخصیص داده‌ها به لایه‌های پایه و افزایشی ۶۷
- شکل ۳-۴ - مقایسه کیفیت بصری نماهای بازتولید شده در (الف) روش پیشنهادی و (ب) روش مرجع که نماهای مجاور به‌عنوان لایه پایه انتخاب شده‌اند. ۷۷

- شکل ۱-۵ - ساختار کلی روش شناسی پیشنهادی برای استخراج مدل نرخ بیت - اعوجاج در سطح نما برای ویدئوی چندنمایی سه بعدی ۸۴
- شکل ۲-۵ - پیچیدگی محتوای واقعی به صورت تابعی از پارامترهای ارائه شده برای اندازه گیری پیچیدگی ویدئو (C, M) ۹۲
- شکل ۳-۵ - ضرایب a, b و c مربوط به رابطه (۵-۱۱) بر حسب مقدار عکس پارامتر کوانتیزاسیون ۹۵
- شکل ۴-۵ - نرخ بیت واقعی ویدئوهای گذشته به صورت تابعی از پارامتر X برای (الف) ۱۵ QP = (ب) ۲۰ QP = ۱۰۳
- شکل ۷-۵ - مقایسه توزیع نرخ بیت برای انواع مختلف نما نسبت به نرخ بیت کل برای مدل پیشنهادی و مدل ارائه شده در [25] ۱۱۱
- شکل ۱-۶. مراحل مختلف روش شناسی ارائه شده برای استخراج معیار ارزیابی کمی کیفیت برای ویدئوی چندنمایی مقیاس پذیر ۱۱۸
- شکل ۲-۶ - ساختار پیش بینی انتخابی برای کد کردن ویدئوی چندنمایی ۱۲۱
- شکل ۳-۶ - نماهای اصلی از ویدئوی Ballet ۱۲۹
- شکل ۴-۶ - حالت های مختلف اعمال نويز به نماهای مهم و غیرمهم (الف) حالت ۱ (ب) حالت ۲ (ج) حالت ۳ (د) حالت ۴ ۱۲۹
- شکل ۵-۶ - واحدهای مرتبه پایین مربوط به قسمت هایی از صحنه که نزدیک به بیننده هستند. (الف) ویدئوی Balloons (ب) ویدئوی Kendo ۱۳۱
- شکل ۶-۶ - نماهای انتخاب شده ویدئوی Break-dancer برای لایه های پایه و افزایشی با اختلاف بین لایه های کم برای دو حالت مختلف تخصیص دادن نماها به لایه های مختلف ۱۳۴
- شکل ۷-۶ - لایه های پایه و افزایشی در مقیاس پذیری نمایی برای ویدئوی برجسته نما، فریم های مختلف ویدئوی Tunnel ۱۳۵
- شکل ۸-۶ - ساختار پیش بینی استفاده شده برای کد کردن ویدئوی برجسته نما ۱۳۶

فهرست جداول

- جدول ۱-۳- ارتباطات بین هسته اصلی و زیردسته‌ها ۵۴
- جدول ۲-۳- تعاریف استخراج شده برای مقیاس‌پذیری در ویدئوی سه‌بعدی چندنمایی با استفاده از نظریه مبنایی و ویژگی‌های متناظر آنها ۵۵
- جدول ۱-۴- مشخصات ویدئوهای آزمایش‌شده ۶۹
- جدول ۲-۴- متوسط اختلاف بین‌نمایی بین نماهای مختلف ویدئوهای آزمایش‌شده ۷۱
- جدول ۳-۴- متوسط اختلاف درون‌نمایی برای نماهای مختلف از ویدئوهای آزمایش‌شده ۷۲
- جدول ۴-۴- مقایسه کیفیت و نرخ بیت (کیلو بیت بر ثانیه) میانگین نماهای انتخابی روش ارائه‌شده و نماهای مجاور برای ویدئوی Exit ۷۲
- جدول ۵-۴- میزان کارایی روش ارائه‌شده از لحاظ BD-PSNR و BD-Bitrate در مقایسه با حالتی که نماهای مجاور انتخاب شوند. (اعداد منفی بهبود کارایی را نشان می‌دهند). ۷۳
- جدول ۶-۴- مشخصات ویدئوهای آزمایش‌شده ۷۵
- جدول ۷-۴- نماهای اختصاص داده شده به لایه پایه و افزایشی و میانگین اختلاف بین لایه-ای برای ویدئوهای آزمایش‌شده ۷۵
- جدول ۸-۴- میزان کارایی روش ارائه‌شده از لحاظ BD-Bitrate در مقایسه با حالتی که نماهای مجاور انتخاب شوند ۷۶
- جدول ۱-۵- پارامترهای تأثیرگذار بر نرخ بیت هر نما در ویدئوی چندنمایی [104] ۸۸
- جدول ۲-۵- مشخصات ویدئوهای کد شده برای استخراج پارامترهای مدل نرخ بیت - اعوجاج ۹۱
- جدول ۳-۵- مقادیر پارامترهای C و M و حاصل ضرب نرخ بیت در پارامتر کوانتیزاسیون برای نماهای مختلف ویدئوی Ballet و به ازای پارامتر کوانتیزاسیون ۱۵ ۹۳
- جدول ۴-۵- مقادیر ضرایب استخراج‌شده a, b و c رابطه (۵-۱۱) و پارامترهای میانگین مربعات خطا و R-square مربوط به فرآیند برازش منحنی ۹۴
- جدول ۵-۵- توزیع نرخ بیت در سطح نما برای ویدئوی Ballet در دو حالت مختلف از لحاظ میانگین اختلاف بین‌نمایی ۹۷

- جدول ۵-۶- مقادیر انتخابی برای پارامتر کوانتیزاسیون، میانگین اختلاف بین‌نمایی و توان پردازشی و نرخ بیت (کیلو بیت بر ثانیه) و کیفیت نهایی برای ویدئوی Ballet ۱۰۲
- جدول ۵-۷ - مقادیر ضرایب استخراج شده e, d, f و رابطه (۵-۲۱) و پارامترهای میانگین مربعات خطا و R-square مربوط به فرآیند برازش منحنی ۱۰۴
- جدول ۵-۸ - میزان حساسیت نرخ بیت پیش‌بینی شده به تغییرات وزن‌های مدل نرخ بیت - اعوجاج ۱۰۷
- جدول ۵-۹ - مشخصات ویدئوهای کد شده برای ارزیابی مدل نرخ بیت - اعوجاج پیشنهادی ۱۰۸
- جدول ۵-۱۰ - میانگین خطای تخمین مدل نرخ بیت - اعوجاج پیشنهادی برای نماهای مختلف ویدئوهای آزمایش شده و برای مقادیر مختلف پارامتر کوانتیزاسیون ۱۰۹
- جدول ۵-۱۱ - مقایسه میانگین خطای تخمین برای مدل نرخ بیت - اعوجاج پیشنهادی و مدل نرخ بیت - اعوجاج خطی [111] برای ۴ نما از ویدئوهای آزمایش شده ۱۰۹
- جدول ۵-۱۲ - میانگین خطای تخمین برای نماهای مختلف عمق برای ویدئوی Pantomim ۱۱۲
- جدول ۶-۱ - خصوصیات ویدئوهای آزمایش شده ۱۲۷
- جدول ۶-۲ - خصوصیات نویزهای اعمال شده برای مقیاس‌پذیری نمایی ۱۲۸
- جدول ۶-۳ - اختلاف بین‌نمایی برای نماهای مختلف ویدئوی Ballet ۱۳۰
- جدول ۶-۴ - مقایسه بین معیار ارزیابی کمی ارائه شده برای هر لایه و روش جمع ساده برای مقیاس‌پذیری نمایی برای حالت‌های مختلف اعمال نویز ۱۳۱
- جدول ۶-۵ - مقایسه بین معیار ارزیابی کمی ارائه شده برای هر لایه و روش جمع ساده برای مقیاس‌پذیری عمق برای حالت‌های مختلف اعمال نویز ۱۳۲
- جدول ۶-۶ - نماهای انتخاب شده برای لایه‌های پایه و افزایشی برای حالت‌های اختلاف بین-لایه‌ای کم و زیاد ۱۳۳
- جدول ۶-۷ - مقایسه نتایج معیار ارزیابی کمی ارائه شده و نتایج ارزیابی بصری برای مقیاس‌پذیری نمایی ۱۳۵

- جدول ۶-۸ - ضریب همبستگی بین معیار ارزیابی کمی کیفیت و معیار ارزیابی بصری کیفیت برای مقیاس پذیری نمایی ۱۳۵
- جدول ۶-۹ - مقایسه نتایج ارزیابی کمی و نتایج ارزیابی بصری برای ویدئوی Tunnel برای مقیاس پذیری نمایی ۱۳۶
- جدول ۶-۱۰ - ضریب همبستگی بین معیار ارزیابی کمی و معیار ارزیابی بصری برای ویدئوی Tunnel برای مقیاس پذیری نمایی ۱۳۶
- جدول ۶-۱۱ - مقایسه نتایج ارزیابی کمی و بصری برای مقیاس پذیری عمق ۱۳۷
- جدول ۶-۱۲ - ضریب همبستگی بین معیار ارزیابی کمی و بصری کیفیت برای مقیاس پذیری عمق ۱۳۷
- جدول ۷-۱ - خصوصیات ویدئوهای آزمایش شده ۱۴۸
- جدول ۷-۲ - نماهای اختصاص داده شده به لایه‌های پایه و افزایشی و مقادیر پارامتر کوانتیزاسیون برای روش پیشنهادی و روش مرجع برای ویدئوی Ballet ۱۴۹
- جدول ۷-۳ - کارایی فشرده‌سازی روش بهینه‌سازی نرخ بیت - اعوجاج پیشنهادی در مقایسه با حالت مرجع استاندارد H.264 ۱۴۹
- جدول ۷-۴ - نماهای اختصاص داده شده به لایه‌های پایه و افزایشی و مقادیر اختلاف بین- لایه‌ای و درون لایه‌ای متناظر ۱۵۳
- جدول ۷-۵ - مقایسه کارایی فشرده‌سازی برای روش بهینه‌سازی نرخ بیت - اعوجاج پیشنهادی نسبت به حالت مرجع استاندارد H.264 برای چهار نمای متفاوت اختصاص داده شده به لایه‌های پایه و افزایشی ۱۵۴

فصل اول

مقدمه

در سال‌های اخیر، ویدئوی چندنمایی سه‌بعدی در حوزه‌های تحقیقاتی و صنعتی بسیار مورد توجه قرار گرفته است. با پیشرفت فناوری‌های¹ ارسال و نمایش ویدئوی سه‌بعدی، افزایش چشمگیری نیز در تولید محصولات سه‌بعدی در صنایع الکترونیک، فیلم‌سازی و بازی‌های کامپیوتری اتفاق افتاده است. ویدئوی سه‌بعدی با در نظر گرفتن ویژگی عمق² نسبت به حالت دوبعدی دید واقعی‌تری از صحنه نمایش داده شده را در اختیار بیننده قرار می‌دهد. انواع مختلفی از قالب‌های³ ویدئوی سه‌بعدی در مقالات این حوزه ارائه شده است [1][2][3][4][5]. ساده‌ترین قالب ارائه شده، ویدئوی سه‌بعدی برجسته‌نما⁴ است که دو نمای⁵ مختلف برای هر یک از دو چشم بیننده تولید می‌کند [1]. در این قالب، تجسم سه بعدی از صحنه، با استفاده از ارسال سیگنال‌هایی با تفاوت کم، برای چشم‌های راست و چپ بیننده ایجاد می‌شود [2]. روش‌های گوناگونی برای تولید ویدئوهای سه‌بعدی برجسته نما وجود دارد. می‌توان هر یک از نماهای راست و چپ را به صورت جداگانه تولید و ارسال کرد یا یک نما را تولید و ارسال کرده و برای نمای دیگر تنها اختلاف آن با نمای اول را ارسال کرد. اطلاعات عمق و اختلاف⁶ نماهای راست و چپ به همراه اطلاعات رنگ هر نما نیز می‌تواند برای تولید و نمایش یک ویدئوی سه‌بعدی برجسته‌نما استفاده شود [3]. با پیشرفت فناوری‌های ارسال و نمایش ویدئوی سه‌بعدی، نسخه جدیدتری از این قالب ویدئویی با عنوان ویدئوهای سه‌بعدی خودبرجسته‌نمای⁷ چندنمایی نیز معرفی شده است که امکان دیدن یک صحنه سه‌بعدی برجسته‌نما (با دو نما برای چشم‌های راست و چپ) را از جهات مختلف فراهم می‌کند [6].

یکی دیگر از کاربردهای ویدئوی چندنمایی سه‌بعدی، ویدئوی نمای آزاد⁸ است که به کاربر این امکان را می‌دهد که جهت مورد نظر خود برای دیدن صحنه سه‌بعدی موجود را به دلخواه انتخاب کند. به منظور فراهم کردن این امکان، ویدئوی سه‌بعدی چندنمایی که با نام تلویزیون نمای آزاد⁹ (FTV) نیز شناخته می‌شود، به عنوان قالب دیگری از ویدئوی چندنمایی سه‌بعدی معرفی شده است [4]. یک سامانه¹⁰ تلویزیون نمای آزاد می‌تواند با استفاده از چند دوربین که به صورت مستقل از هم، از یک صحنه تصویربرداری می‌کنند،

¹ Technology

² Depth

³ Format

⁴ Stereoscopic

⁵ View

⁶ Disparity

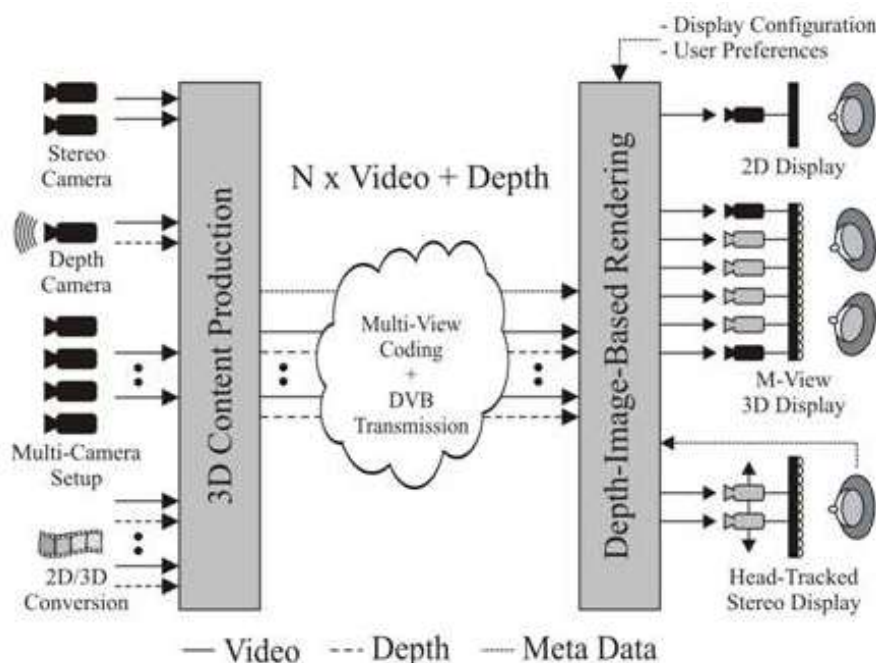
⁷ Autostereoscopic

⁸ Free viewpoint video

⁹ Free viewpoint television

¹⁰ System

ساخته شود. علاوه بر این، یک یا چند دوربین عمق^۱ و چند دوربین مستقل از هم برای تصویربرداری نیز می‌توانند در یک سامانه تلویزیون نمای آزاد استفاده شوند. از فرآیند تبدیل ویدئوی دوبعدی به ویدئوی سه‌بعدی نیز می‌توان به این منظور استفاده کرد. شکل ۱-۱ مثالی از یک سامانه تلویزیون نمای آزاد را نشان می‌دهد.



شکل ۱-۱- مثالی از یک سامانه تلویزیون نمای آزاد [4]

علاوه بر موارد ذکر شده، ویدئوی چندنمایی می‌تواند در کاربردهای قدیمی‌تر مربوط به ویدئوی دوبعدی هم استفاده شده و کاربر را قادر سازد که زاویه دید خود را تغییر داده و صحنه مورد نظر را از زوایای مختلف ببیند [1]. این عملکرد می‌تواند در کاربردهایی مانند سامانه‌های نظارتی، آموزشی یا تصویربرداری از مسابقات ورزشی مورد توجه قرار گیرد. وجه اشتراک تمام قالب‌های اشاره شده برای ویدئوی سه‌بعدی، وجود یک رشته ویدئویی چندنمایی است که توسط چندین دوربین و به صورت هم‌زمان از یک صحنه واحد تصویربرداری می‌کنند. از آنجا که در سال‌های اخیر کاربردهای گوناگون ویدئوی سه‌بعدی و چندنمایی مانند تله‌کنفرانس چندنمایی^۲، تلویزیون سه‌بعدی^۳ و تلویزیون نمای آزاد رشد چشم‌گیری داشته است، و نیز اینکه در همه این کاربردها حجم داده‌هایی که باید ذخیره و ارسال شود، متناسب با تعداد دوربین‌ها افزایش می‌یابد، فشرده‌سازی بهینه ویدئوی

¹ Depth camera

² Immersive teleconferencing

³ 3DTV

چندنمایی ضروری به نظر می‌رسد. استاندارد کد کردن ویدئوی چندنمایی¹ (MVC) به عنوان یکی از بخش‌های اضافه شده به استاندارد کد کردن ویدئوی H.264/AVC [7] به منظور بهبود فشرده‌سازی و کیفیت ویدئوی چندنمایی طراحی شده است. علاوه بر قابلیت این استاندارد در کد کردن بهینه ویدئوی چندنمایی با کیفیت بالا، پروفایل دو نمایی² از این استاندارد نیز برای کد کردن ویدئوی سه‌بعدی دو نمایی با کیفیت بالا استفاده می‌شود [8]. در سال‌های اخیر، دستگاه‌های موبایل و تبلت‌ها³ با داشتن امکاناتی مانند پردازنده‌های سریع، حافظه‌های با ظرفیت بالا، قابلیت اتصال به شبکه‌های بی‌سیم و امکان به‌روز رسانی سریع برای فناوری‌های جدید، به سرعت جایگزین کامپیوترهای معمولی در بسیاری از کاربردها شده‌اند. در حالی که کاربردهای چند رسانه‌ای قدیمی‌تر مانند ویدئوهای دوبعدی، همچنان برای دستگاه‌های موبایل مورد توجه هستند، با رشد استفاده از دستگاه‌های موبایل و گسترش استفاده از ویدئوهای سه‌بعدی، انتظار می‌رود که امکان استفاده از این فناوری جدید نیز بر روی این دستگاه‌ها فراهم شود.

بعضی ویژگی‌های خاص دستگاه‌های موبایل، آنها را برای استفاده در صنعت ویدئوهای سه‌بعدی بسیار مناسب می‌سازد. عینک‌های سه‌بعدی، که برای دیدن تصاویر سه‌بعدی در تلویزیون و سینما ضروری است، یکی از موانع بزرگ در گسترش استفاده از این صنعت به شمار می‌رود. اما دستگاه‌های موبایل مانند تلفن‌های هوشمند⁴ معمولاً در حین استفاده، در فاصله و زاویه خاصی از صورت کاربر قرار می‌گیرند، به این ترتیب موقعیت ایده‌آلی برای نمایش ویدئوهای خودبرجسته‌نما بدون نیاز به عینک فراهم می‌شود [9]. با وجود این ویژگی و نیز قابلیت جایگزینی سریع که برای دستگاه‌های موبایل وجود دارد، استفاده از ویدئوی سه‌بعدی می‌تواند برای این دستگاه‌ها حتی سریع‌تر از تلویزیون‌های با صفحه نمایش بزرگ، رشد کند [9].

کارهای انجام شده توسط شرکت‌های سازنده دستگاه‌های موبایل، نشان‌دهنده میزان اهمیت گسترش فناوری ویدئوهای سه‌بعدی در نسل‌های بعدی دستگاه‌های موبایل است. گروه تحقیقاتی نوکیا یک نسخه از کد کننده ویدئوی سه‌بعدی برای تبلت نوکیا پیاده‌سازی کرده و آن را به عنوان نمونه اولیه ویدئوی سه‌بعدی موبایل ارائه کرده است [10]. HTC EVO 3D می‌تواند از صحنه مورد نظر به صورت سه‌بعدی عکس و فیلم تهیه کرده و آن را بدون نیاز به عینک به صورت سه‌بعدی نمایش دهد [11]. Nintendo 3DS یک محیط بازی قابل حمل است

¹ Multi-view video coding

² Stereo

³ Tablet

⁴ Smartphone

که توسط شرکت Nintendo ساخته شده و تصاویر سه‌بعدی را بدون نیاز به تجهیزات اضافی نمایش می‌دهد [12]. Movidius Myriad 3D قادر به تصویربرداری دو نمایی با کمک دو دوربین مجزا با قدرت تفکیک‌پذیری بالا است. این تصاویر سپس به صورت بی‌درنگ پردازش شده و به تصاویر و فیلم‌های سه‌بعدی تبدیل می‌شوند [13].

علی‌رغم مزایای مطرح شده، چالش‌های مهمی نیز در استفاده از ویدئوهای سه‌بعدی در دستگاه‌های موبایل وجود دارد. محدودیت توان پردازشی، پهنای باند محدود، تأخیر بالای شبکه به دلیل ماهیت شبکه‌های بی‌سیم و نامتجانس بودن دستگاه‌های گیرنده موبایل بعضی از مشکلاتی هستند که استفاده از ویدئوهای سه‌بعدی را برای این دستگاه‌ها محدود می‌کنند. حتی با وجود شبکه‌های بی‌سیم نسل چهارم^۱ با پهنای باند ۱۰۰ مگا بیت در ثانیه، نمی‌توان به صورت مداوم از این پهنای باند استفاده کرد. استفاده از پهنای باند بالا به مصرف انرژی بیشتر و تخلیه‌شدن سریع‌تر باتری منجر می‌شود. به دلیل این محدودیت‌ها، امروزه ویدئوهای سه‌بعدی برای دستگاه‌های موبایل معمولاً با تفکیک‌پذیری و کیفیت پایین ارائه می‌شوند. بنابراین سامانه‌های چند رسانه‌ای نسل بعدی باید قابلیت تطبیق با امکانات دستگاه‌های موبایل را داشته باشند.

۱-۱- مقیاس‌پذیری در ویدئوی چندنمایی

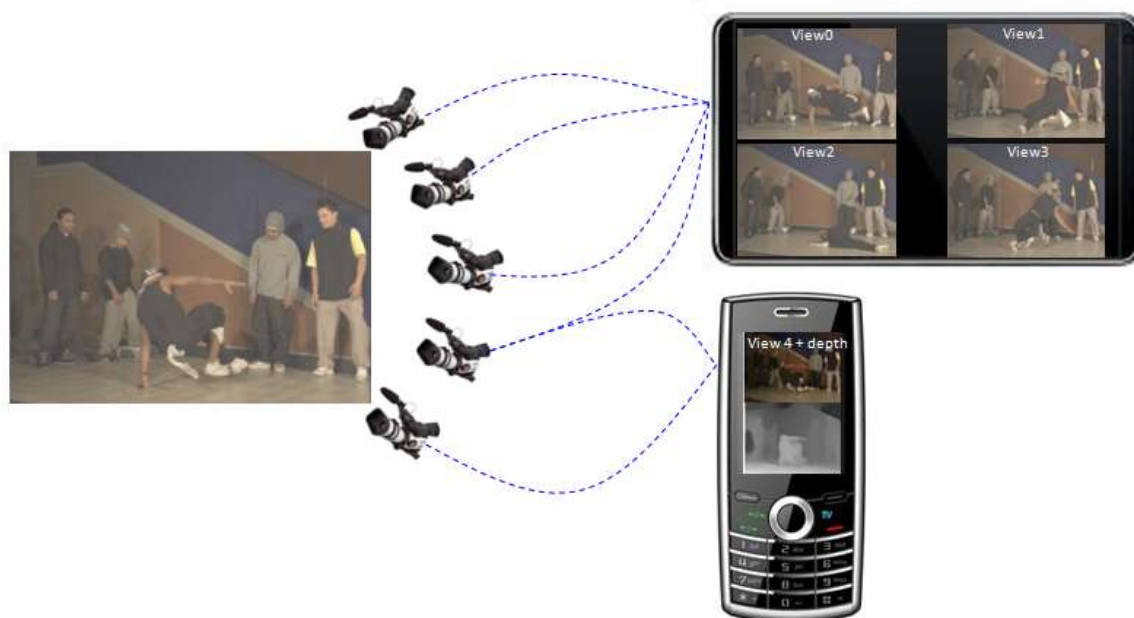
در حال حاضر، حتی با در نظر گرفتن قابلیت‌های فشرده‌سازی استاندارد کدکردن ویدئوی چندنمایی، ویدئوی چندنمایی و سایر کاربردهای آن به دلیل حجم بالای داده‌های تولیدی، قابل استفاده در بسیاری از دستگاه‌های موبایل با قابلیت‌های محدود پردازشی و پهنای باند نیستند. یک روش مؤثر برای کاهش حجم داده‌ها در ویدئوی چندنمایی، استفاده از مقیاس‌پذیری^۲ در رشته ویدئویی تولید شده است. این ویژگی به‌خصوص برای ارسال ویدئو در محیط‌های ناهمگون که گیرنده‌ها قابلیت‌های متفاوتی از لحاظ پهنای باند، قدرت تفکیک‌پذیری، اندازه صفحه نمایش و توان پردازشی و شرایط شبکه ارتباطی [14] دارند، می‌تواند بسیار مفید باشد. برای مثال، یک رشته ویدئو با ۵ نمای متفاوت می‌تواند به یک ویدئوی سه‌بعدی برجسته‌نمای مناسب برای دستگاه‌های موبایل تبدیل شود [15]. شکل ۱-۲ مثالی از یک سامانه ویدئوی چندنمایی مقیاس‌پذیر را نشان می‌دهد.

در این شکل، یک صحنه توسط چندین دوربین تصویر برداری شده و یک رشته ویدئویی چندنمایی به دست آمده است. سپس این رشته ویدئویی بر اساس قابلیت‌های گیرنده‌ها

¹ Fourth generation wireless network

² Scalability

تعدیل شده است. برای دستگاه موبایل با توان پردازشی و پهنای باند محدود، نسخه سه-بعدی برجسته‌نما و برای تبلت با امکانات و قابلیت‌های بیشتر، ۴ نما از ۵ نمای موجود آماده شده است.



شکل ۱-۲ - مثالی از یک سامانه ویدئوی چندنمایی مقیاس‌پذیر. (الف) تبلت با ۴ نمای دریافتی (ب) دستگاه موبایل با دریافت ویدئوی سه‌بعدی برجسته‌نما با یک نما به اضافه اطلاعات عمق

۲-۱- بیان مسأله

تمرکز این رساله بر روی عملی ساختن استفاده از مقیاس‌پذیری در ویدئوی چندنمایی سه-بعدی است. در کنار مزایای مطرح شده برای ویدئوی چندنمایی مقیاس‌پذیر، محدودیت‌هایی نیز برای استفاده از این نوع ویدئو وجود دارد که به‌منظور عملی ساختن استفاده از آن باید به آنها پرداخته شود. در این رساله یک چارچوب برای استفاده انطباق‌پذیر از مقیاس‌پذیری در ویدئوی چندنمایی ارائه می‌شود. منظور از چارچوب ساختاری است که شامل قسمت‌های لازم برای استفاده عملی از مقیاس‌پذیری در کاربردهای متنوع ویدئوی چندنمایی است. انطباق‌پذیری نیز به انطباق روش استفاده شده برای مقیاس‌پذیری با کاربرد مورد نظر اشاره دارد. قسمت اول چارچوب پیشنهادی، نوع مقیاس‌پذیری مورد استفاده را از روی ویژگی‌های کاربرد مورد نظر انتخاب می‌کند. سپس بر اساس نوع مقیاس‌پذیری، مفاهیم اختلاف بین‌لایه‌ای و درون‌لایه‌ای تعریف می‌شوند. این مفاهیم که در قسمت‌های بعدی چارچوب پیشنهادی مورد استفاده قرار می‌گیرند، بر روی کارایی بخش‌های مختلف از جهت کیفیت تجربی اثر خواهد داشت. به‌این‌ترتیب، روش‌های استفاده شده در قسمت‌های مختلف

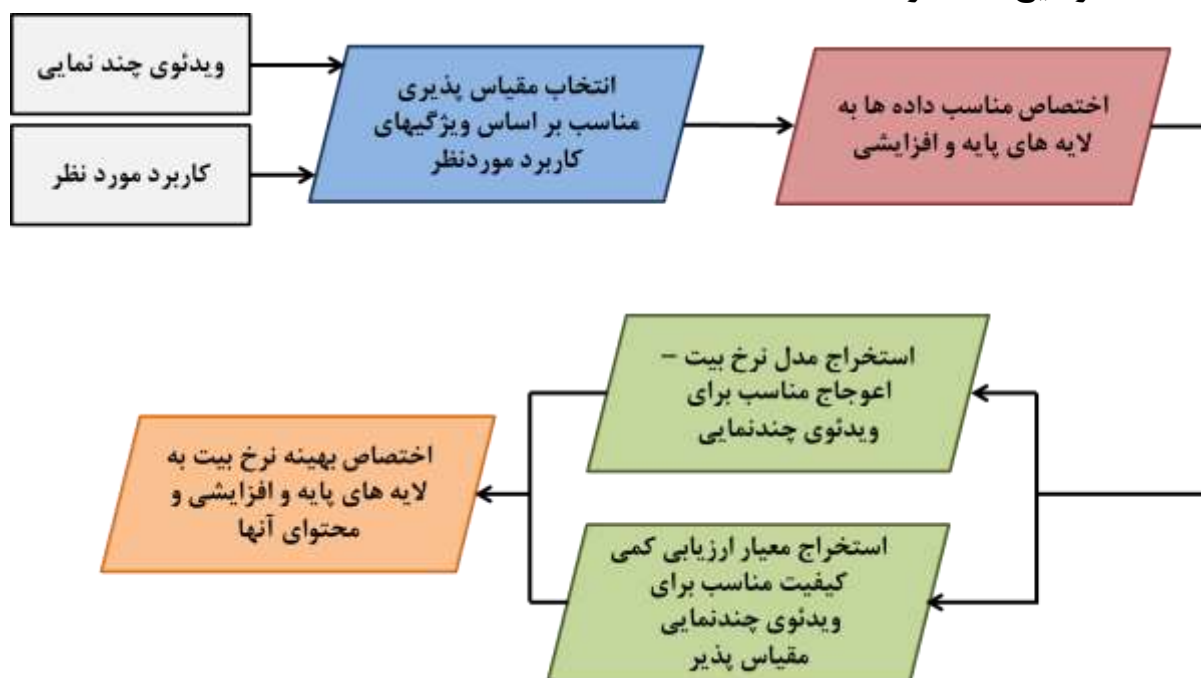
چارچوب پیشنهادی با در نظر گرفتن ویژگی‌های کاربرد موردنظر، نیازهای آنها از جهت کیفیت تجربی را مرتفع می‌کند. این مسأله لزوم و اهمیت تطبیق مقیاس‌پذیری با کاربرد مورد نظر را نشان می‌دهد. پس از انتخاب نوع مقیاس‌پذیری مناسب، در قسمت بعدی چارچوب پیشنهادی، یک روش‌شناسی برای تخصیص داده‌های مناسب به لایه‌های پایه و افزایشی تخصیص ارائه می‌شود. روش‌شناسی رویه‌ای است که با تجزیه و تحلیل داده‌های موجود، روشی کلی برای حل مسأله مورد بررسی ارائه می‌دهد. به این ترتیب، روش‌شناسی می‌تواند در برگیرنده همه روش‌های موجود برای حل این مسأله در حالت کلی باشد. سپس، یک روش‌شناسی برای استخراج مدل نرخ بیت - اعوجاج در سطح نما برای ویدئوی چندنمایی به صورت روش‌مند استخراج می‌شود. در ادامه، به منظور ارزیابی کارایی کیفیت ویدئوی چندنمایی مقیاس‌پذیر، معیاری برای ارزیابی کمی کیفیت ویدئوی چندنمایی مقیاس‌پذیر پیشنهاد می‌شود. در نهایت، با استفاده از مدل نرخ بیت - اعوجاج ارائه‌شده و معیار ارزیابی کیفیت، روشی برای اختصاص بهینه نرخ بیت به لایه‌های مختلف و محتوای آنها در ویدئوی چندنمایی مقیاس‌پذیر ارائه خواهد شد.

در مقالات ارائه‌شده در این حوزه، تاکنون روش‌شناسی برای تخصیص داده‌ها به لایه‌های مختلف، و بهینه‌سازی نرخ بیت - اعوجاج این لایه‌ها و محتوای آنها که ویژگی‌های ویدئوی چندنمایی و نیز محدودیت‌ها و نیازمندی‌های کاربردهای خاص آن را در نظر بگیرد، ارائه نشده است. به این ترتیب، چارچوب پیشنهادی ارائه شده در این رساله، با در نظر گرفتن موارد فوق، امکان استفاده از ویدئوی چندنمایی با حجم بالای داده‌ها را برای محیط‌هایی با قابلیت‌های محدود مانند دستگاه‌های موبایل و تبلت، فراهم می‌کند. به دلیل استفاده از ویژگی‌های کاربرد موردنظر در انتخاب مقیاس‌پذیری مناسب در چارچوب پیشنهادی، این چارچوب به راحتی می‌تواند با کاربردهای مختلف ویدئوی چندنمایی سه‌بعدی انطباق داده شود. علاوه بر این، استخراج مدل نرخ بیت - اعوجاج برای ویدئوی چندنمایی در سطح نما و ارائه معیار ارزیابی کیفیت کمی ویدئوی چندنمایی مقیاس‌پذیر، به عنوان پیش‌نیاز برای استفاده در روش ارائه‌شده برای بهینه‌سازی نرخ بیت - اعوجاج در این چارچوب در نظر گرفته شده و از آنجایی که در مقالات و مطالعات قبلی روش‌شناسی کارآمد با در نظر گرفتن ویژگی‌های خاص ویدئوی چندنمایی و مقیاس‌پذیر برای این منظور وجود ندارد، در این رساله، برای هر قسمت یک روش‌شناسی ارائه شده است.

نکته قابل ذکر اینکه، اکثر ویدئوهای چندنمایی استاندارد که به صورت عمومی در دسترس هستند و در مقالات و رساله‌ها مورد استفاده قرار می‌گیرند، ویدئوهای دوبعدی چندنمایی هستند.

به دلیل عدم وجود ویدئوهای سه‌بعدی چندنمایی استاندارد، برای ارزیابی بخش‌های مختلف چارچوب پیشنهادی در این رساله نیز از ویدئوهای چندنمایی دوبعدی استفاده شده است. اما تمامی روش‌شناسی‌های ارائه‌شده در این رساله قابل تعمیم به ویدئوهای سه‌بعدی چندنمایی نیز می‌باشند. البته در بعضی قسمت‌های چارچوب پیشنهادی مانند مدل نرخ بیت - اعوجاج و نیز معیار ارزیابی کیفیت کمی ویدئوی چندنمایی که برداشت بیننده از ویژگی سه‌بعدی ویدئوها در نشان دادن کارایی روش ارائه‌شده اثر مهمی دارد، به‌طور خاص از ویدئوهای سه‌بعدی چندنمایی یا ویدئوهای برجسته‌نما نیز برای ارزیابی‌ها استفاده شده است.

براین اساس، ساختار کلی چارچوب پیشنهادی در این رساله به‌صورت شکل ۱-۳ خواهد بود. قسمت‌های مختلف این چارچوب در زیربخش‌های ۱-۲-۱، ۱-۲-۲، ۱-۲-۳، ۱-۲-۴ و ۱-۲-۵- توضیح داده خواهند شد.



شکل ۱-۳ - ساختار کلی چارچوب پیشنهادی در این رساله

۱-۲-۱ استخراج مقیاس‌پذیری مناسب بسته به کاربرد مورد نظر

انواع مختلف مقیاس‌پذیری برای ویدئوی تک‌نمایی و چندنمایی سه‌بعدی معرفی شده است. این موارد شامل، مقیاس‌پذیری زمانی^۱، مکانی^۲ و کیفیت^۳ و ترکیبات مختلف آنها، مقیاس-

^۱ Temporal scalability

^۲ Spatial scalability

^۳ SNR scalability

پذیری ناحیه مورد علاقه^۱ و مقیاس پذیری مبتنی بر شیء^۲، مقیاس پذیری نمایی^۳ و مقیاس-پذیری نمای آزاد^۴ است. شکل ۱-۴ بعضی از انواع مقیاس پذیری های اشاره شده را نمایش می دهد.

انواع مقیاس پذیرهای موجود برای ویدئوی چندنمایی سه بعدی، به صورت غیر روش مند^۵ استخراج شده، یا از حالت تک نمایی به حالت چندنمایی تعمیم داده شده [16] و یا تنها برای کاربردهای خاص ویدئوی چندنمایی سه بعدی تعریف شده است. این روش نمی تواند ضمانت کند که همه مقیاس پذیری های مناسب برای همه کاربردهای ویدئوی چندنمایی و سه بعدی در نظر گرفته شده است. علاوه بر این، با رشد فناوری های مربوط به ویدئوهای چندنمایی، کاربردهای جدیدتری نیز برای این ویدئوها ارائه شده است که تعاریف فعلی مقیاس پذیری ممکن است به راحتی قابل تعمیم به آنها نباشد و یا اینکه در آن موارد خاص کارایی چندانی نداشته باشد. بنابراین، نیاز داریم که انواع مختلف مقیاس پذیری که ویژگی های اصلی ویدئوی چندنمایی سه بعدی را در نظر بگیرد، استخراج کنیم. استخراج انواع مختلف مقیاس-پذیری برای کاربردهای ویدئوی چندنمایی سه بعدی به صورت روش مند یکی از قسمت های چارچوب پیشنهادی این رساله است.

۱-۲-۲- اختصاص مناسب داده ها به لایه های پایه و افزایشی

ویدئوی مقیاس پذیر از یک لایه پایه^۶ و یک یا چند لایه افزایشی^۷ تشکیل شده است. نماهایی که برای لایه پایه انتخاب می شوند، اثر مهمی روی کیفیت و کارایی فشرده سازی خواهد داشت. در محیط های موبایل، احتمال اینکه لایه های افزایشی به دلیل محدودیت پهنای باند یا توان پردازشی به طور کامل حذف شوند، زیاد است. در این حالت، به دلیل وجود وابستگی بین لایه های موجود در مقیاس پذیری های تعریف شده برای ویدئوی چندنمایی ممکن است حتی داده های لایه پایه هم قابل واکد کردن نباشند. به عنوان مثال، در کد کردن ویدئوی چندنمایی، از ساختار پیش بینی شکل ۱-۵ استفاده می شود. به دلیل وجود وابستگی های بین نمایی در این شکل، واضح است که همه نماهای مرجع باید در سمت گیرنده دریافت شوند، تا سایر نماها قابل واکد کردن باشند.

¹ Region of Interest (RoI) scalability

² Object-based scalability

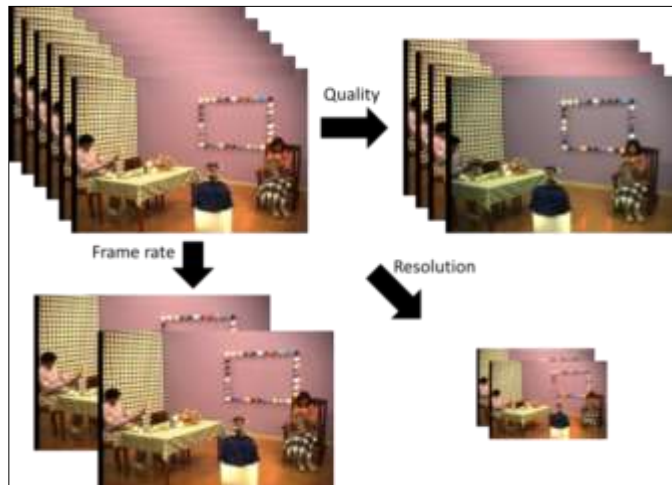
³ View scalability

⁴ Free view-point scalability

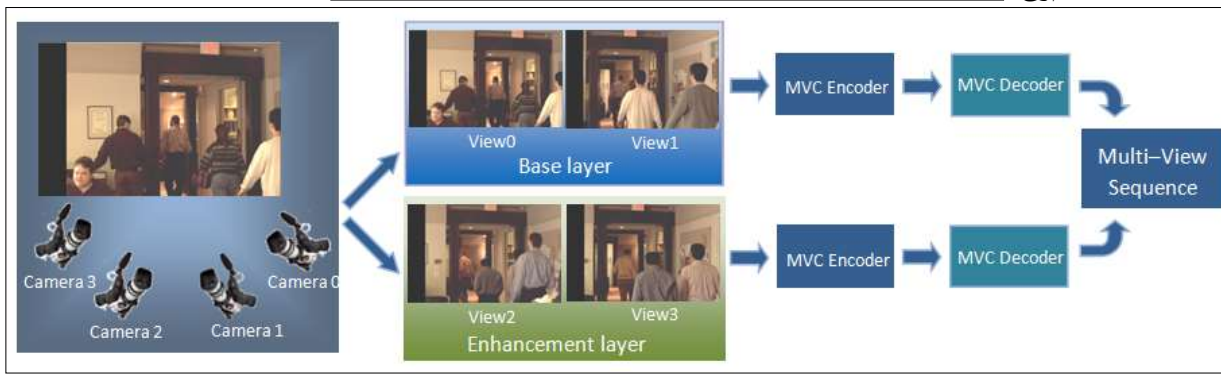
⁵ Ad-hoc

⁶ Base layer

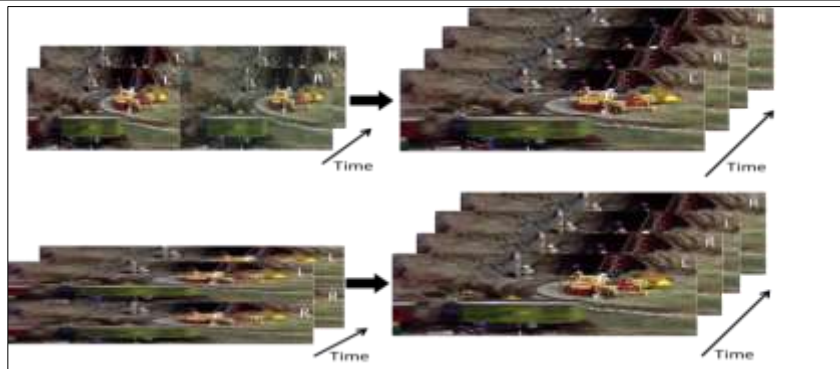
⁷ Enhancement layer



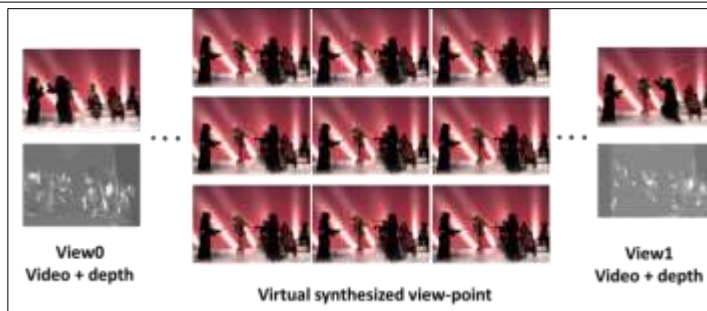
الف



ب

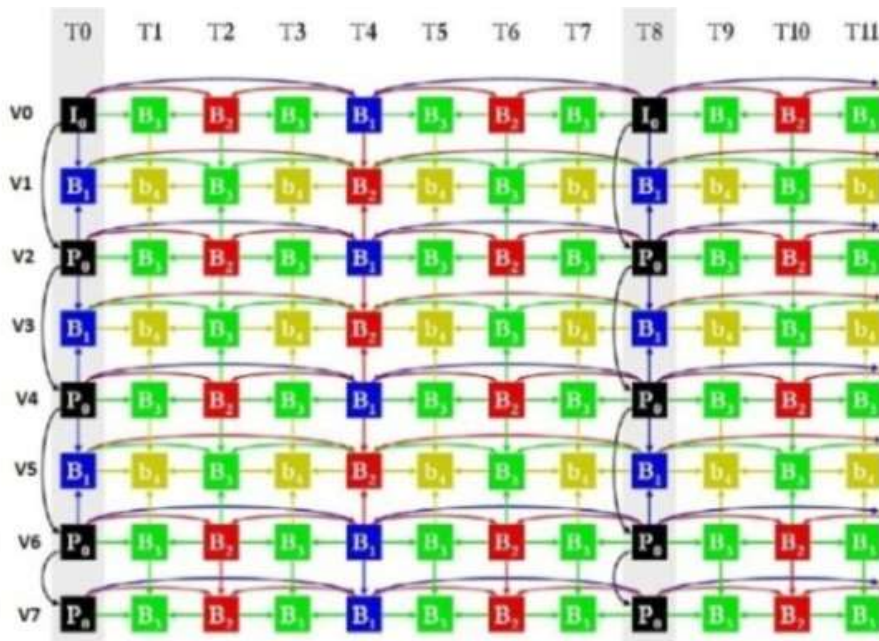


ج



د

شکل ۴-۱ - انواع مختلف مقیاس پذیری (الف) مقیاس پذیری مکانی/زمانی/کیفیت برای ویدئوی تک‌نمایی (ب) مقیاس پذیری نمایی برای ویدئوی چندنمایی (ج) مقیاس پذیری سازگار با فریم برای ویدئوی برجسته‌نما (د) مقیاس-پذیری نمای آزاد برای ویدئوی چندنمایی سه‌بعدی



شکل ۱-۵ - ساختار پیش‌بینی معمول در استاندارد H.264/AVC [15]

اما همان‌طور که می‌دانیم مثلاً در مقیاس‌پذیری نمایی، هدف این است که تعداد نماهای ارسالی با توجه به قابلیت‌های گیرنده‌ها محدود شود. روش کلی برای این کار انتخاب تعدادی از نماها برای لایه پایه و تخصیص سایر نماها به لایه‌های افزایشی است. بسته به قابلیت هر گیرنده، لایه پایه و تعدادی از لایه‌های افزایشی می‌تواند توسط هر گیرنده دریافت شود.

اما به دلیل وابستگی‌های بین‌نمایی در ساختار پیش‌بینی بر اساس شکل ۱-۵، لایه پایه هم می‌تواند به لایه‌های افزایشی وابسته باشد. در مقاله‌های پیشین، از روش‌های بازتولید^۱ نماها به منظور بازیابی نماهای از دست رفته در سمت گیرنده استفاده می‌کنند. این روش‌ها در واقع این قابلیت را فراهم می‌کند که در صورت از دست رفتن نماهای لایه افزایشی، نماهای لایه پایه قابل واکردن باشند. اما الگوریتم‌های بازتولید نماها پیچیدگی محاسباتی بالایی دارند، علاوه بر این، ارسال اطلاعات جانبی^۲ که برای بازتولید نماها مورد نیاز است (مانند اطلاعات عمق)، نیازمند پهنای باند اضافی است. اکثر دستگاه‌ها و محیط‌های موبایل قابلیت تأمین کردن این نیازمندی‌ها را ندارند.

در قسمت دوم چارچوب پیشنهادی این رساله، یک روش‌شناسی^۳ برای انتخاب مناسب داده‌ها برای لایه پایه و افزایشی برای ویدئوی مقیاس‌پذیر ارائه می‌شود. در روش‌شناسی

^۱ Synthesize

^۲ Side information

^۳ Methodology

ارائه شده نحوه انتخاب داده‌های مناسب برای لایه پایه بر اساس نوع مقیاس‌پذیری انتخاب شده در قسمت قبل خواهد بود. در واقع این روش، از مفاهیم اختلاف بین لایه‌ای و درون-لایه‌ای که بر اساس نوع مقیاس‌پذیری تعریف می‌شوند، به عنوان معیاری برای این انتخاب استفاده می‌شود.

۱-۲-۳ - مدل نرخ بیت - اعوجاج مناسب برای ویدئوی چندنمایی

قسمت بعدی چارچوب پیشنهادی به استخراج مدل نرخ بیت - اعوجاج مناسب در سطح نما برای ویدئوی چندنمایی سه بعدی می‌پردازد. در واقع، به منظور حل مسأله بهینه‌سازی نرخ بیت - اعوجاج برای ویدئوی چندنمایی مقیاس‌پذیر نیاز داریم که الگوریتم‌های کنترل نرخ بیت را در ویدئوهای چندنمایی بررسی کنیم. الگوریتم‌های کنترل نرخ بیت معمولاً از یک مدل نرخ بیت - اعوجاج به منظور توصیف ارتباط بین نرخ بیت و کیفیت ویدئو استفاده می‌کنند. به این دلیل، چارچوب پیشنهادی در این رساله، به استخراج یک مدل نرخ بیت - اعوجاج مناسب در سطح نما برای ویدئوی چندنمایی می‌پردازد.

انتخاب پارامترهای مناسب برای مدل نرخ بیت - اعوجاج اثر مهمی روی دقت مدل دارد. بنابراین ویژگی‌های مهم ویدئوی چندنمایی سه بعدی باید در انتخاب پارامترها مورد توجه قرار گیرد. در این قسمت از چارچوب پیشنهادی رساله، از مفاهیم اختلاف‌های بین لایه‌ای و درون لایه‌ای که بر اساس نوع مقیاس‌پذیری انتخاب شده تعریف شده‌اند، به منظور استخراج این پارامترها استفاده می‌شود. بنابراین مدل پیشنهادی علاوه بر در نظر گرفتن ویژگی‌های ویدئوی چندنمایی، با کاربرد خاص مورد نظر نیز قابل تطبیق خواهد بود.

۱-۲-۴ - تخمین کیفیت در ویدئوی چندنمایی مقیاس‌پذیر

به منظور اندازه‌گیری اعوجاج در مدل نرخ بیت - اعوجاج پیشنهادی و نیز بررسی کارایی الگوریتم بهینه‌سازی نرخ بیت - اعوجاج، نیاز داریم کیفیت ویدئوی چندنمایی مقیاس‌پذیر به طور مستمر ارزیابی کنیم. این کار دست‌یابی مستمر به کیفیت قابل قبول در محیط‌های با خطای بالا را امکان‌پذیر می‌سازد. بر اساس دانسته‌های ما معیاری برای اندازه‌گیری کیفیت به صورت کمی برای ویدئوی چندنمایی مقیاس‌پذیر در کارهای انجام شده قبلی وجود ندارد. ارزیابی کیفیت در ویدئوی چندنمایی مقیاس‌پذیر چالش‌های زیادی دارد. اگرچه معیارهای مختلفی برای ارزیابی کمی کیفیت ویدئو برای ویدئوهای چندنمایی دو بعدی و سه بعدی ارائه شده است، اما هیچ‌یک از آنها نمی‌تواند به سادگی و با کارایی بالا به حالت مقیاس‌پذیر تعمیم داده شود. روش‌های ارزیابی بصری کیفیت نیز بسیار زمان‌بر و پرهزینه هستند و در همه

کاربردهای از جمله کاربردهای بی‌درنگ قابل استفاده نخواهند بود. در حقیقت، هدف ارزیابی کمی کیفیت فراهم کردن یک راه قابل اطمینان برای تخمین کیفیت دریافتی ویدئو توسط گیرنده است. به این منظور، مفهوم کیفیت تجربی^۱ باید در ارزیابی کمی کیفیت به‌طور گسترده مدنظر قرار گیرد.

در این رساله، یک روش‌شناسی مشخص برای ارزیابی کمی کیفیت ویدئوی چندنمایی مقیاس‌پذیر ارائه می‌شود. این روش‌شناسی از دو مرحله تشکیل شده است: ارزیابی کمی کیفیت هر لایه، و سپس ارزیابی کیفیت نهایی ویدئوی چندنمایی مقیاس‌پذیر. برای هر مرحله روش جمع وزن‌دار پیشنهاد شده است، چون می‌تواند کیفیت دریافتی بیننده را به‌طور عمده در معیار کمی کیفیت منعکس کند. علاوه بر این، مفاهیم اختلاف بین‌لایه‌ای و اختلاف درون‌لایه‌ای نیز در ارزیابی کیفیت ویدئو مورد استفاده قرار خواهند گرفت. این مفاهیم بر اساس هر نوع مقیاس‌پذیری خاص به‌صورت جداگانه تعریف شده و سپس به‌منظور یافتن وزن مناسب برای جمع وزن‌دار کیفیت نماها و محتوای آنها مورد استفاده قرار می‌گیرند.

۱-۲-۵ - اختصاص بهینه نرخ بیت به لایه‌های پایه و افزایشی و محتوای آنها

اگرچه کدکردن ویدئوی چندنمایی به‌صورت مقیاس‌پذیر می‌تواند برای دستگاه‌های موبایل با قابلیت‌های پهنای باند و توان پردازشی متفاوت مفید باشد، اما اختصاص پهنای باند موجود به لایه‌های مختلف و محتوای آنها به‌نحوی که بیشترین کیفیت قابل دسترسی حاصل شود، یکی از چالش‌های این حوزه است. بهینه‌سازی نرخ بیت - اعوجاج، به این معنی که مقدار اعوجاج در مقابل محدودیت نرخ بیت، بهینه شود، یک روش برای بهبود کیفیت ویدئو در فشرده‌سازی ویدئو است. حاصل این بهینه‌سازی، انتخاب پارامترهای مناسب برای کدکردن، شامل نوع ماکروبلوک، گام کوانتیزاسیون و غیره است، به‌نحوی که بیشترین کارایی فشرده‌سازی ممکن حاصل شود.

روش‌های مختلفی برای بهینه‌سازی نرخ بیت - اعوجاج برای ویدئوی تک‌نمایی و چندنمایی در تحقیقات قبلی پیشنهاد شده است. در قسمت بعدی چارچوب پیشنهادی این رساله کمی از تحقیقات موجود فراتر رفته و روشی برای بهینه‌سازی نرخ بیت - اعوجاج برای ویدئوی چندنمایی سه‌بعدی مقیاس‌پذیر ارائه می‌دهیم. روش پیشنهادی برای بهینه‌سازی نرخ بیت - اعوجاج ویدئوی چندنمایی سه‌بعدی مقیاس‌پذیر باید ویژگی‌های مرتبط با این نوع ویدئوها را در نظر بگیرد. از آنجایی که حل مسائل بهینه‌سازی معمولاً پیچیدگی

^۱ Quality of Experience (QoE)

محاسباتی بالایی دارد. استفاده از ویژگی‌های مرتبط ویدئوهای چندنمایی سه‌بعدی، می‌تواند به ساده‌سازی فرآیند بهینه‌سازی از لحاظ محاسباتی کمک کند. در روش پیشنهادی در این رساله، خروجی مسأله بهینه‌سازی شامل پارامترهای کوانتیزاسیون برای کدکردن محتوای لایه‌های پایه و افزایشی است.

۱-۳- ساختار رساله

ساختار کلی رساله بر اساس چارچوب پیشنهادی از چند قسمت تشکیل شده است. در فصل دوم رساله به پیشینه تحقیق قسمت‌های مختلف چارچوب پیشنهادی خواهیم پرداخت و کارهایی که در زمینه هر یک از قسمت مذکور وجود دارند را معرفی کرده، ضمن برشمردن مزایا و معایب هر یک، نوآوری‌های خود را در مقایسه با هر یک از آنها عنوان خواهیم کرد. در فصل سوم، روش‌شناسی پیشنهادی برای استخراج مقیاس‌پذیری مناسب برای ویدئوی چندنمایی بسته به کاربرد موردنظر مورد بررسی قرار خواهد گرفت. در فصل چهارم، روش اختصاص مناسب داده‌ها به لایه‌های مختلف پایه و افزایشی ارائه شده و برای دو کاربرد خاص شامل کاربردهای موبایل و تلویزیون نمای آزاد مورد بررسی قرار خواهد گرفت. فصل پنجم به معرفی روش پیشنهادی برای استخراج مدل نرخ بیت - اعوجاج برای ویدئوی چندنمایی سه‌بعدی می‌پردازد. در فصل ششم نیز معیار پیشنهادی در این رساله برای ارزیابی کمی کیفیت ویدئوی چندنمایی مقیاس‌پذیر معرفی می‌شود. فصل هفتم به روش پیشنهادی برای اختصاص نرخ بیت بهینه به لایه‌های مختلف ویدئوی چندنمایی مقیاس‌پذیر و محتوای آنها بر اساس مدل نرخ بیت - اعوجاج و معیار ارزیابی کیفیت ارائه شده در فصل‌های پنج و شش خواهد پرداخت. در نهایت، در فصل هشتم به جمع‌بندی رساله خواهیم پرداخت و کارهای آینده ممکن در رابطه با این رساله پیشنهاد خواهند شد.

۱-۴- جمع‌بندی

در این فصل فضای کلی این رساله که در ارتباط با یک چارچوب انطباق‌پذیر برای ویدئوی چندنمایی سه‌بعدی است، تبیین شد. تعریف انواع مناسب مقیاس‌پذیری برای کاربردهای مختلف ویدئوی چندنمایی و انتخاب مقیاس‌پذیری مناسب بسته به کاربرد موردنظر، روش اختصاص مناسب داده‌های به لایه‌های مختلف ویدئوی چندنمایی مقیاس‌پذیر، ارائه رویه‌ای روش‌مند برای استخراج مدل مناسب نرخ بیت - اعوجاج برای ویدئوی چندنمایی بر اساس ویژگی‌های آن، ارزیابی کیفیت ویدئوی چندنمایی مقیاس‌پذیر و اختصاص بهینه نرخ بیت به

لایه‌های مختلف ویدئوی چندنمایی مقیاس‌پذیر و محتوای آنها از بخش‌های اصلی این چارچوب است که در این فصل مورد بررسی اجمالی قرار گرفت. در بخش دیگری از این فصل به ساختار رساله و روند ارائه مطالب اشاره شد که در ادامه به هر یک از آنها خواهیم پرداخت. در ابتدا در فصل بعدی به پیشینه تحقیق و کارهای مشابه در زمینه مسائل مطرح شده در این فصل خواهیم پرداخت.

فصل دوم

پیشینه تحقیق

در این فصل به ارائه پیشینه تحقیق خواهیم پرداخت. مقالات مربوط به هر یک از قسمت‌های چارچوب پیشنهادی شامل انواع مختلف مقیاس‌پذیری، روش‌های تخصیص داده به لایه‌های پایه و افزایشی، مدل‌های نرخ بیت - اعوجاج برای ویدئوی تک‌نمایی و چندنمایی، معیارهای ارزیابی کیفیت ویدئوی چندنمایی و روش‌های بهینه‌سازی نرخ بیت - اعوجاج به ترتیب در زیر فصل‌های ۱-۲، ۲-۲، ۳-۲، ۴-۲ و ۵-۲ مورد بررسی قرار خواهند گرفت.

۱-۲ - مقیاس‌پذیری در ویدئوی چندنمایی

در ویدئوی تک‌نمایی، مقیاس‌پذیری زمانی، مکانی و کیفیت و ترکیبات مختلف آنها برای کاربردهای متفاوت بسیار مورد استفاده قرار گرفته است [17]. مقیاس‌پذیری زمانی اجازه می‌دهد که لایه‌های مختلف، ویدئوهایی با نرخ فریم متفاوت داشته باشند. در مقیاس‌پذیری مکانی، ویدئو در لایه‌های مختلف با وضوح تصویر متفاوت کد می‌شود. مقیاس‌پذیری کیفیت نیز لایه‌هایی با سطوح متفاوت کیفیت تولید می‌کند. مقیاس‌پذیری ناحیه مورد علاقه و مقیاس‌پذیری مبتنی بر شیء انواع دیگر مقیاس‌پذیری برای ویدئوی تک‌نمایی هستند، که در کاربردهای خاص ویدئوی تک‌نمایی استفاده شده‌اند [18].

علاوه بر مقیاس‌پذیری‌های ارائه‌شده برای ویدئوی تک‌نمایی، انواع مختلف مقیاس‌پذیری برای ویدئوی چندنمایی سه‌بعدی نیز ارائه شده است. برای مثال، مقیاس‌پذیری نمایی که گیرنده را قادر می‌سازد که براساس قابلیت پهنای باند و توان پردازشی خود، تنها تعدادی از نماها را دریافت کند [19]. در مقیاس‌پذیری نمای آزاد رشته ویدئوی مقیاس‌پذیر به‌نحوی تولید می‌شود که گیرنده بتواند با دریافت لایه‌های خاص به نماهای موردنظر خود دسترسی داشته باشد [15]. در ویدئوی برجسته‌نما، مقیاس‌پذیری این‌گونه تعریف می‌شود که رشته بیت غیربرجسته‌نما به‌عنوان لایه پایه ارسال می‌شود و داده‌های باقی‌مانده که شامل سیگنال برجسته‌نما است در یک یا چند لایه افزایشی ارسال می‌شود. سرانجام، فرمت ویدئویی سازگار با فریم^۱ یک فرمت خاص برای ویدئوی سه‌بعدی است که در آن فریم‌های نماهای راست و چپ با هم در یک فریم قرار می‌گیرند، به‌نحوی که وضوح تصویر آنها نصف وضوح تصویر فریم اصلی باشد [8]. این ساختار برای نمایش ویدئوی سه‌بعدی، امکان تعمیم-دادن مقیاس‌پذیری‌های حالت تک‌نمایی به حالت چندنمایی سه‌بعدی را فراهم می‌کند [20].

^۱ Frame compatible video format

انتخاب نوع مقیاس‌پذیری مورد نظر در کاربردهای مختلف یکی از مسائل مهم است. برای مثال، اگر یک دستگاه موبایل پهنای باندی کمتر از نرخ بیت ویدئو ارسالی داشته باشد، کدام نوع مقیاس‌پذیری (زمانی، مکانی، کیفیت یا ترکیبی از آنها) باید به‌منظور منطبق کردن نرخ بیت ویدئوی مقیاس‌پذیر با پهنای باند دستگاه موبایل مورد استفاده قرار گیرد؟ این موضوع معمولاً مسأله پیچیده‌ای است که تحقیقاتی [21] برای حل آن در ویدئوهای دوبعدی نیز انجام شده است.

کد کردن ویدئوی چندنمایی به‌صورت مقیاس‌پذیر ابعاد این مسأله را گسترش می‌دهد، چون انواع مقیاس‌پذیری ارائه شده برای ویدئوی چندنمایی به مراتب بیشتر است. با این حال، تاکنون تلاشی برای استخراج انواع مقیاس‌پذیری برای ویدئوی چندنمایی سه‌بعدی به‌صورت روش‌مند انجام نشده است. بدون بررسی روش‌مند این مسأله، نمی‌توان گفت که آیا تعاریف موجود برای مقیاس‌پذیری برای کاربردهای جدید ویدئوی چندنمایی و سه‌بعدی نیز مناسب است یا نه. قسمت اول چارچوب پیشنهادی این رساله به ارائه مقیاس‌پذیری‌های مناسب به‌صورت روش‌مند و بسته به کاربردهای مختلف ویدئوی چندنمایی می‌پردازد.

۲-۲- روش‌های ارائه‌شده برای اختصاص داده‌ها به لایه‌های پایه و افزایشی در

ویدئوی چندنمایی

همان‌طور که پیش‌تر توضیح داده شد، مقیاس‌پذیری در ویدئوی چندنمایی روشی برای انطباق ویدئوی چندنمایی سه‌بعدی با شرایط گیرنده‌های مختلف است. اما یکی از سؤال‌های اساسی در این زمینه، نحوه تخصیص مناسب داده‌های موجود به لایه‌های مختلف پایه و افزایشی است. در [22] روشی ارائه شده است که مقیاس‌پذیری را بر اساس اطلاعات عمق و فاصله تا نمای مرکزی تعیین می‌کند. به‌این ترتیب، نمای مرکزی به‌عنوان لایه پایه در نظر گرفته می‌شود و داده‌های مرتبط با بقیه نماها بر حسب فاصله آنها تا نمای مرکزی در لایه‌های مختلف افزایشی قرار می‌گیرند. در این حالت، فاصله تا نمای مرکزی و اطلاعات عمق مبنای تخصیص داده‌ها به لایه‌های مختلف افزایشی است.

در این روش انتخاب نما برای لایه‌های پایه و افزایشی بدون توجه به ویژگی‌های کاربرد موردنظر انجام می‌شود. روش‌شناسی ارائه‌شده در این رساله، اختصاص داده‌ها را بر اساس نوع مقیاس‌پذیری انتخاب‌شده انجام می‌دهد. از آنجایی‌که در چارچوب پیشنهادی این رساله، برای انتخاب مقیاس‌پذیری، ویژگی‌های کاربرد موردنظر در نظر گرفته شده است،

روش پیشنهادی در این رساله می‌تواند محدودیت‌های کاربرد موردنظر مانند محدودیت پهنای باند یا توان پردازشی را مدنظر قرار دهد.

۲-۳- مدل‌های نرخ بیت – اعوجاج برای ویدئوی چندنمایی

الگوریتم‌های مختلفی برای کنترل نرخ بیت در کاربردهای متفاوت ویدئوی چندنمایی سه-بعدی ارائه شده است. همان‌طور که پیش‌تر اشاره شد، بیشتر این الگوریتم‌ها از نوعی مدل نرخ بیت – اعوجاج برای توصیف ارتباط بین نرخ بیت و کیفیت استفاده می‌کنند. در [23] یک الگوریتم کنترل نرخ بیت برای ویدئوی چندنمایی سه‌بعدی بر اساس مدل نرخ بیت – اعوجاج درجه دوم و مدل ترافیک جریان سیال^۱ پیشنهاد شده است. مدل نرخ بیت – اعوجاج درجه دو در [24] نیز برای ویدئوی چندنمایی استفاده شده است. این مدل برای تخمین نرخ بیت با دقت بالاتر از سه سطح تشکیل شده است: گروهی از گروه تصاویر^۲، گروه تصاویر^۳ و فریم. در [25] تکنیکی برای تخمین نرخ بیت در سطح نما برای ویدئوی چندنمایی به‌علاوه عمق برای کاربردهای بی‌درنگ پیشنهاد شده است. روش پیشنهادی بر اساس تحلیل تصادفی مودهای پیش‌بینی استفاده شده در انواع مختلف نماها است. در [26] بحث شده است که در ویدئوی چندنمایی نماهای نوع B داریم که تنها از فریم‌های نوع B تشکیل شده‌اند. بنابراین، مقدار پارامتر کوانتیزاسیون فریم‌های B به‌عنوان یک ویژگی خاص برای ویدئوهای چندنمایی باید به‌منظور اختصاص نرخ بیت مناسب به نماهای B در نظر گرفته شوند. در [27] یک روش کنترل نرخ بیت پیشنهاد شده است که از خصوصیات سامانه بصری انسان به‌منظور توزیع نرخ بیت به نواحی مورد توجه فریم‌ها استفاده می‌کند. سپس پارامتر کوانتیزاسیون نواحی مورد توجه به‌منظور کدکردن آنها با کیفیت بهتر، کاهش می‌یابد. در [28] یک تکنیک کنترل نرخ بیت برای ویدئوی چندنمایی به‌علاوه عمق معرفی شده است که در سه سطح انجام می‌شود: سطح نما، سطح ویدئو/عمق و سطح فریم. یک الگوریتم کنترل نرخ بیت برای کدکردن ویدئوی چندنمایی در [29] پیشنهاد شده است که مدل نرخ بیت – اعوجاج درجه دو را بر اساس نوع فریم‌ها تغییر می‌دهد. الگوریتم دیگری برای کنترل نرخ بیت در [30] پیشنهاد شده است که نرخ بیت را در سه سطح نما، گروهی از تصاویر و سطح فریم تخصیص می‌دهد. کنترل نرخ بیت بر اساس نوع نما و با استفاده از روش تخصیص نرخ بیت پیش‌آماری به هر فریم اختصاص می‌یابد. در [31] یک مدل نرخ بیت

¹ Fluid-flow traffic

² Group of GOP (GGOP)

³ Group of Picture (GOP)

– اعوجاج برای ویدئوی چندنمایی پیشنهاد شده است که در آن کیفیت هر نما نسبت به نرخ بیت آن از یک تابع لگاریتمی افزایشی پیروی می‌کند. در [32] یک الگوریتم کنترل نرخ بیت برای کدکردن ویدئوی چندنمایی پیشنهاد شده است که از مدل تخصیص نرخ بیت بر اساس تئوری لاگرانژ استفاده می‌کند. در نهایت، [33] یک کنترل‌کننده نرخ بیت برای کدکردن ویدئوی چندنمایی ارائه می‌کند که وابستگی بین گروه‌های تصاویر نماهای مختلف را به منظور پیش‌بینی نرخ بیت فریم‌های آینده با در نظر گرفتن وابستگی بین گروه‌های تصاویر همان نما استخراج می‌کند. در [34] بحث می‌شود که از آنجایی که دریافت‌کننده ویدئو انسان است، خصوصیات سامانه بینایی انسان می‌تواند به منظور فشرده‌سازی ویدئو چندنمایی مورد استفاده قرار گیرد. این مقاله یک مدل جدید پیشنهاد می‌کند که خصوصیات سامانه بینایی انسان را در دریافت سه‌بعدی مورد توجه قرار می‌دهد. این مدل علاوه بر اطلاعات زمانی و مکانی، اطلاعات عمق را نیز در نظر می‌گیرد. در [35] یک الگوریتم کنترل نرخ بیت بر مبنای مدل نرخ بیت – اعوجاج درجه دو پیشنهاد شده که به منظور دقت بیشتر، نرخ بیت را در چهار سطح متفاوت تخصیص می‌دهد، سطح نما، سطح فریم، سطح واحد پایه¹ و سطح ماکروبلوک. مقاله [36] به تحلیل نرخ بیت و اعوجاج در زمینه کدکردن ویدئوی چندنمایی می‌پردازد. این مقاله یک مدل تحلیلی برای واریانس خطای باقی‌مانده‌ها در فریم‌های پیش‌بینی شده با استفاده از اطلاعات عمق ارائه می‌کند. مقاله [37] توضیح می‌دهد که در کدکردن ویدئوی سه‌بعدی، توزیع نرخ بیت بین ویدئوی عمق و بافت روی کیفیت نماهای مجازی بازتولید شده اثر مهمی خواهد داشت. بنابراین، در یک نرخ بیت معین، اختصاص مناسب نرخ بیت بین ویدئوی عمق و بافت مهم خواهد بود. در این مقاله یک روش جدید برای اختصاص و کنترل نرخ بیت برای ویدئوی چندنمایی به علاوه عمق پیشنهاد شده است. به این منظور، ابتدا یک مدل اعوجاج برای مشخص کردن اثر اعوجاج ویدئوی بافت و عمق روی نماهای مجازی بازتولید شده پیشنهاد شده است. سپس، بر اساس این مدل، یک مدل نرخ بیت – اعوجاج برای مشخص کردن ارتباط بین نرخ بیت و اعوجاج نماهای مجازی بازتولید شده ارائه شده است. در نهایت، با استفاده از این مدل، نسبت توزیع نرخ بیت بهینه بین ویدئوی بافت و عمق با حل مسأله بهینه‌سازی استخراج شده است. مقاله [38] از خصوصیات دریافت بصری برای بینندگان ویدئوی سه‌بعدی استفاده کرده و نواحی مورد علاقه برای بینندگان را در همه نماها و در سطوح متفاوتی از میزان اهمیت تعیین می‌کند. میزان اهمیت این نواحی از روی وجود ماکروبلوک‌هایی با میزان حرکت بالا یا اشیایی با شکل یا بافت مشخص تعیین می‌شوند. سپس پارامتر کوانتیزاسیون برای این نواحی به-

¹ Basic unit

صورت مناسب تنظیم می‌شود به نحوی که کیفیت نواحی مورد علاقه در سطح مناسبی باقی بماند. در [39] یک الگوریتم کنترل نرخ بیت برای ویدئوی چندنمایی براساس دریافت بصری ارائه می‌شود که از چهار سطح تشکیل شده است. در سطح نما، گروهی از تصویر پیش‌کد شده و سهم نرخ بیت آنها از بین نرخ بیت نماهای مختلف به دست می‌آید. در سطح گروهی از تصاویر، پارامتر کوانتیزاسیون اولیه و نرخ بیت برای گروهی از تصاویر تعیین می‌شود. در سطح فریم نیز تخصیص نرخ بیت براساس پیش‌بینی پیچیدگی فریم انجام می‌شود. در نهایت، در سطح ماکروبلوک، مدل نرخ بیت - اعوجاج براساس دریافت بصری تنظیم می‌شود. در [40] یک الگوریتم کنترل نرخ بیت سلسله‌مراتبی برای ویدئوی چندنمایی و برای کاربردهایی با پهنای باند و کیفیت بالا ارائه می‌شود. این روش در سطح فریم و واحد پایه طراحی شده است. تعیین پارامتر کوانتیزاسیون هر فریم به وسیله فرآیند تصمیم‌مارکوف که در سطح واحدهای پایه طراحی شده است، می‌تواند نواحی مهم و غیرمهم را در تعیین مناسب پارامتر کوانتیزاسیون در نظر بگیرد.

علی‌رغم مزایای روش‌های اشاره شده در بالا، هیچ‌یک از آنها ویژگی‌های خاص ویدئوی چندنمایی را در نظر نمی‌گیرند. روش‌شناسی ارائه‌شده در این رساله، ویژگی‌های مرتبط با ویدئوی چندنمایی مانند اختلاف بین‌نمایی و درون‌نمایی را در استخراج مدل نرخ بیت - اعوجاج در نظر می‌گیرد.

۲-۴- روش‌های ارائه‌شده برای تخمین کیفیت در ویدئوی چندنمایی

هدف ارزیابی کیفیت ویدئو تخمین کیفیت دریافتی توسط بیننده و میزان رضایت او از ویدئوی موردنظر است. این کار می‌تواند به دو صورت انجام شود: ارزیابی بصری و ارزیابی کمی. در ارزیابی بصری کیفیت، بیننده بر حسب کیفیت دریافتی خود به ویدئوی موردنظر امتیازدهی می‌کند. بنابراین این روش تخمین دقیقی از کیفیت دریافتی بیننده است. متأسفانه، ارزیابی بصری کیفیت نمی‌تواند به‌طور عملی به‌صورت بی‌درنگ و به‌صورت محاسبات اتوماتیک در کاربردهای مختلف ویدئو استفاده شود. بنابراین، ارزیابی کمی کیفیت به‌منظور تخمین کیفیت ویدئو با در نظر گرفتن مدل‌های ریاضی که نتایج ارزیابی بصری را تخمین می‌زند، استفاده می‌شود.

معیارهای ارزیابی کمی کیفیت در سه گروه اصلی دسته‌بندی می‌شوند، مرجع کامل^۱، بدون مرجع^۲ و مرجع کاهش‌یافته^۱ [41]. در روش‌های ارزیابی کمی با مرجع کامل، به‌منظور ارزیابی

¹ Full-reference

² No-reference

کیفیت، باید ویدئوی اصلی در سمت گیرنده موجود باشد. در روش‌های ارزیابی کیفیت بدون مرجع، ارزیابی کیفیت نهایی تنها بر اساس ویدئوی دریافتی انجام می‌شود و نیازی به ویدئوی اصلی نیست. در نهایت، در معیارهای ارزیابی کیفیت با مرجع کاهش یافته، کیفیت ویدئو تنها با استفاده از اطلاعات جزئی از ویدئوی مرجع انجام می‌شود [42]. روش‌های معمول برای ارزیابی کیفیت ویدئوی دیجیتال در معیارهای ارزیابی با مرجع کامل نسبت سیگنال به نویز (PSNR) بین ویدئوی اصلی و ویدئوی پردازش شده را محاسبه می‌کنند. اگرچه این روش در کاربردهای زیادی مورد استفاده قرار گرفته، اما نتایج تجربی نشان می‌دهد که به دلیل رفتار غیرخطی سامانه بینایی انسان، این معیار همیشه با دریافت واقعی انسان هم‌خوانی ندارد [43]. بنابراین، روش‌های ارزیابی کمی کیفیت دیگری مانند شباهت ساختاری (SSIM) که اعوجاج ساختاری بین ویدئوی پردازش شده و ویدئوی مرجع را نشان می‌دهد و تقریب بهتری از کیفیت دریافتی است، معرفی شده‌اند [44].

روش‌های مختلفی برای ارزیابی بصری کیفیت و به‌طور خاص برای ویدئوی چندنمایی سه‌بعدی پیشنهاد شده است. در [45] در مورد اثر اعوجاج ناشی از ارسال ویدئو و روش‌های مختلف پنهان کردن خطا در کیفیت بصری بحث شده است. مقاله [46] اثر کد کردن نامتقارن در کیفیت بصری ویدئوی برجسته‌نمای سه‌بعدی را مورد بررسی قرار داده است. مقاله [47] روش تعاملی دیگری برای اندازه‌گیری کیفیت دریافتی برای کد کردن ویدئوی نامتقارن ارائه می‌کند. در [48] ارزیابی بصری کیفیت به‌منظور معتبر ساختن کیفیت ویدئوی سه‌بعدی و دریافت عمق انجام شده و نتیجه این آزمایش‌ها به‌منظور تعیین معیارهای ارزیابی کمی کیفیت دقیق‌تر مورد استفاده قرار گرفته است. مقاله [49] ادعا می‌کند که استفاده از منابع و افراد توزیع شده^۲ روشی کم‌هزینه برای ارزیابی کمی کیفیت ویدئو است. در این مقاله دو روش برای بررسی کیفیت ویدئو به‌علاوه عمق روی نمایشگر دوبعدی با استفاده از این منابع پیشنهاد شده است: با استفاده از نمای مجازی و با استفاده از ویدئوی نمای آزاد.

به‌طور مشابه، روش‌های اندکی برای ارزیابی کمی کیفیت برای ویدئوهای چندنمایی سه‌بعدی پیشنهاد شده است. مقاله [50] یک معیار کیفیت بر اساس نقشه عمق برای ویدئو برجسته‌نما و برجسته‌نمای سه‌بعدی پیشنهاد می‌کند. مقاله [51] اثر کیفیت عمق را روی کیفیت ویدئوی سه‌بعدی با ارزیابی کمی و بصری نشان می‌دهد. در [52] یک معیار ارزیابی کمی کیفیت برای ویدئوی سه‌بعدی پیشنهاد می‌شود که به تغییرات کم در کاهش کیفیت ویدئو و تصویر و میزان خطای آنها حساس است. در [53] یک الگوریتم ارزیابی کمی کیفیت

¹ Reduced-reference

² Crowdsourcing

پیشنهاد شده است که از اطلاعات عمق به‌همراه نماهای برجسته استفاده می‌کند. این روش برمبنای تشخیص لبه و کاهش کیفیت رنگ است. مقاله [54] مدل استاندارد شده معیار کیفیت ویدئو¹ را با استفاده از نور محیط، ویژگی‌های مربوط به محتوا مانند میزان حرکت، ویژگی‌های ساختاری و تبیین² روشنایی برای اندازه‌گیری کیفیت ویدئوی سه‌بعدی تعمیم می‌دهد.

مقاله [55] یک روش ارزیابی کیفیت برای ویدئوی برجسته براساس تبدیل DCT سه‌بعدی ارائه می‌کند. در این روش، بلوک‌های مشابه از نماهای چپ و راست از روش تطبیق بلوک‌ها و به‌وسیله تحلیل DCT سه‌بعدی استخراج می‌شوند. سپس میانگین مربعات خطا در حوزه DCT سه‌بعدی محاسبه می‌شود. در [56] یک معیار ارزیابی کمی کیفیت پیشنهاد می‌شود که دریافت چشم انسان از عمق و ویدئو، مانند واریانس زمانی، تغییرات اختلاف بین‌نمایی در فریم‌های intra، تغییرات اختلاف بین‌نمایی در فریم‌های چپ مرز فریم‌ها را در نظر می‌گیرد. در [57] بررسی شده است که اختلاف بین کیفیت نماهای چپ و راست چگونه می‌تواند روی کیفیت نهایی ویدئوی سه‌بعدی اثر بگذارد. در [58] یک روش ارزیابی کیفیت ویدئوی چندنمایی سه‌بعدی براساس اختلاف بین‌نمایی و معیار شباهت ساختاری پیشنهاد می‌شود. در این مقاله کیفیت ویدئو از دو جنبه مورد بررسی قرار گرفته است، دریافت سه‌بعدی و کیفیت تصویر. دریافت سه‌بعدی بر اساس اختلاف بین‌نماها و کیفیت تصویر بر اساس شباهت نماها ارزیابی شده است.

درنهایت، کیفیت تجربی نیز در سال‌های اخیر در ویدئوی چندنمایی سه‌بعدی مورد توجه قرار گرفته است. در [59] ادعا می‌شود که کیفیت تجربی در ویدئوی سه‌بعدی به فرمت‌های کدکردن ویدئو مثل سازگار با فریم و فرمت فریم‌های متوالی بستگی دارد. این مقاله ارزیابی‌های بصری مختلفی برای ویدئوهایی با فرمت فریم‌های کنار هم³ و فرمت فریم‌های متوالی⁴ پیشنهاد می‌کند. مقاله [60] اثر خطای ارسال را روی کیفیت تجربی سامانه‌های تلویزیون سه‌بعدی بررسی می‌کند، با این فرض که ویدئو با فرمت فریم‌های کنار هم و روی شبکه برمبنای بسته⁵ منتقل می‌شود. در [61] کیفیت تجربی ویدئوی چندنمایی و ارسال صوت روی شبکه‌های آی پی بررسی می‌شود. این مقاله حالتی که در آن کاربر تنها یک نما را مشاهده می‌کند را با حالتی که امکان انتخاب نماها وجود داشته باشد، مقایسه می‌کند. در [62] دو معیار ارزیابی کمی کیفیت برای پیش‌بینی کیفیت تجربی ویدئوی برجسته‌نما

¹ Video Quality Metric (VQM)

² Contrast

³ Side-by-Side

⁴ Frame-sequential

⁵ Packet-based network

پیشنهاد می‌شود. یکی از آنها از عوامل مهم تأثیرگذار در کیفیت تجربی مانند محتوای صحنه، محل دوربین، اندازه صفحه و سطح تداخل^۱ استفاده می‌کند. در حالی که دومی از ویژگی‌های ادراکی شامل ادراک تداخل^۲ درک شده و عمق درک شده استفاده می‌کند. مقاله [63] کیفیت تجربی را در ارسال ویدئوهای چندنمایی و صوت روی شبکه‌های آی پی بررسی می‌کند که در آن کاربران این امکان را دارند که نمای موردنظر خود را انتخاب کنند. در [64] یک روش تخمین کیفیت تجربی به صورت بی‌درنگ برای ویدئوی چندنمایی و صوت روی شبکه‌های آی پی ارائه شده است. این روش از پارامترهای کیفیت تجربی برای ویدئو در سطح کاربردی مانند میانگین بار روی شبکه، زمان بافر کردن و تأخیر شبکه استفاده می‌کند که می‌توانند به صورت بی‌درنگ اندازه‌گیری شوند. در [65] کیفیت تجربی و کیفیت خدمات در سطح کاربردی ارزیابی می‌شود. تمرکز این مقاله بر روی اثرات ترتیب دوربین‌ها و محتوا روی کیفیت تجربی و کیفیت خدمات است. مقاله [66] بر روی ویدئوی چندنمایی و انتخاب صوت روی شبکه‌های آی پی بحث می‌کند. در این کاربرد، کاربران نه تنها ویدئو بلکه صوت را نیز بر اساس زاویه موردنظر خود درخواست می‌دهند. کیفیت تجربی در این کاربرد به صورت بصری ارزیابی شده است.

از لحاظ دریافت محتوای سه‌بعدی، ارزیابی کیفیت می‌تواند به عنوان بازخورد برای تنظیم ارسال ویدئو و پارامترهای سامانه مورد استفاده قرار گیرد. در این حالت، باید امکان ارزیابی کیفیت ویدئو در سمت گیرنده حتی وقتی که ویدئوی سه‌بعدی اصلی در دسترس نیست، وجود داشته باشد. معیارهای ارزیابی کیفیت با مرجع کاهش یافته یا بدون مرجع برای حل این مشکلات لازم است [67]. مقاله [68] یک معیار ارزیابی کیفیت بدون مرجع پیشنهاد می‌دهد که ویژگی‌های آماری را از اطلاعات اختلاف‌های بین‌نمایی و گرادیان آنها و نیز فعالیت‌های مکانی تصویر استخراج می‌کند. سپس، این ویژگی‌های مکانی و اختلاف‌های بین‌نمایی به منظور پیش‌بینی کیفیت تجربی مورد استفاده قرار می‌گیرد. در [69] یک معیار ارزیابی کمی کیفیت بدون مرجع پیشنهاد می‌شود که از تجزیه و بولت، ارتباط بین فریمی و وضوح لبه‌ها در فریم‌های متوالی استفاده می‌کند. در [70] یک معیار ارزیابی کیفیت با مرجع کاهش یافته برای نقشه عمق برای ویدئوهای با فرمت عمق به علاوه رنگ پیشنهاد می‌شود که از تشخیص لبه استفاده می‌کند. مقاله [71] یک معیار ارزیابی کیفیت با مرجع کاهش یافته برای ارسال نقشه عمق ویدئوی سه‌بعدی پیشنهاد می‌کند که اطلاعات استخراج شده برای لبه‌ها را نیز در نظر می‌گیرد.

¹ Crosstalk level

² Crosstalk perception

اگرچه روش‌های بالا می‌تواند برای ارزیابی کیفیت ویدئوی چندنمایی سه‌بعدی مورد استفاده قرار گیرد، اما بر اساس دانسته‌های ما، معیاری برای ارزیابی کیفیت نهایی ویدئوی چندنمایی مقیاس‌پذیر وجود ندارد. اشاره به این نکته لازم است که روش‌های بالا به راحتی قابل تعمیم به حالت مقیاس‌پذیر نیستند. کیفیت نهایی در ویدئوی مقیاس‌پذیر بر اساس کاربرد مورد نظر، شرایط شبکه و محدودیت‌های گیرنده تعیین می‌شود. برای مثال، در مقیاس‌پذیری نمایی، کیفیت دریافتی ویدئوی مقیاس‌پذیر ممکن است تحت تأثیر کیفیت نماهای با کیفیت پایین از لایه افزایشی قرار نگیرد، زیرا احتمال اینکه آن نماها به دلیل محدودیت‌های شبکه و گیرنده به طور کامل حذف شوند، وجود دارد. روش پیشنهادی در این رساله برای ارزیابی کمی کیفیت ویدئوی چندنمایی مقیاس‌پذیر همه موارد فوق را در نظر می‌گیرد.

۲-۵- روش‌های ارائه شده برای بهینه‌سازی نرخ بیت - اعوجاج در ویدئوی چندنمایی

هدف بهینه‌سازی نرخ بیت - اعوجاج حداقل کردن اعوجاج با در نظر گرفتن محدودیت نرخ بیت نهایی است. روش‌های مختلفی برای بهینه‌سازی نرخ بیت - اعوجاج در ویدئوی مقیاس‌پذیر ارائه شده است. معروف‌ترین الگوریتم بهینه‌سازی نرخ بیت - اعوجاج برای کد کردن تک‌گذری^۱ در نرم‌افزار مرجع استاندارد H.264/SVC ارائه شده است. این الگوریتم، بهینه‌سازی را بر اساس فرآیند کوانتیزاسیون انجام می‌دهد، در حالیکه ویژگی‌های ویدئوی ورودی را در نظر نمی‌گیرد [72]. در [73] ویدئوی تک‌نمایی به صورت مجموعه‌ای از ویدئوهای کوچک‌تر در نظر گرفته می‌شود. سپس تحلیل نرخ بیت - اعوجاج به منظور اختصاص نرخ بیت مناسب به هر یک از این ویدئوهای کوچک انجام می‌شود. یک تکنیک برای توزیع بهینه نرخ بیت برای زیرگروه‌ها نیز در [74] پیشنهاد و بر روی کدکننده مقیاس‌پذیر دو بعدی پیاده‌سازی شده است. این تکنیک از مدل نرخ بیت - اعوجاجی استفاده می‌کند که بین نرخ بیت‌های بالا و پایین سوئیچ کرده و بهترین نرخ بیت - اعوجاج موجود را انتخاب می‌کند. در [75] یک مدل اتفاقی برای کد کردن ویدئوی مقیاس‌پذیر پیشنهاد شده است که وابستگی بین لایه‌ها را مدل می‌کند. سپس روشی برای انتخاب پارامتر کوانتیزاسیون بهینه به صورت توأم برای لایه‌های مختلف و برای منبع گاوسی بدون حافظه استخراج می‌کند. مقاله [76] یک روش بهینه‌سازی نرخ بیت - اعوجاج برای ویدئوی مقیاس‌پذیر با مقیاس-پذیری مکانی پیشنهاد می‌دهد که از وابستگی بین باقی‌مانده‌های حاصل از پیش‌بینی بین

^۱ One-pass

لایه‌ای برای انتخاب مقادیر مناسب برای ضرایب تبدیل برای لایه‌هایی با وضوح تصویر پایین‌تر استفاده می‌کند. یک روش انتخاب مود کد کردن ماکروبلوک بر اساس بهینه‌سازی نرخ بیت - اعوجاج نیز در [77] معرفی شده که اثر کاهش کیفیت ناشی از خطای ارسال و انتشار را حداقل می‌کند. در [78] مدلی برای بهینه‌سازی نرخ بیت - اعوجاج به‌منظور بهبود انعطاف‌پذیری خطا برای ارسال ویدئو روی کانال‌های در معرض خطا پیشنهاد شده است. در [79] بحث شده است که در همه روش‌های قبلی بهینه‌سازی نرخ بیت - اعوجاج، معیار اعوجاج از منظر پردازش سیگنال مورد توجه قرار گرفته است، که همه خصوصیات دریافت بصری را منعکس نمی‌کند، زیرا برای ویدئوهایی با محتوای مختلف، سامانه بینایی انسان حساسیت‌های متفاوتی دارد. در این مقاله به‌منظور فراهم کردن یک مدل بهینه‌سازی نرخ بیت - اعوجاج که با سامانه بینایی انسان مطابقت بیشتری داشته باشد، از معیاری برای اعوجاج استفاده شده است که شباهت ساختاری و اطلاعات محتوا را بیشتر در نظر می‌گیرد. مدل بهینه‌سازی نرخ بیت - اعوجاج ارائه شده، نرخ بیت موجود را به‌نحوی تخصیص می‌دهد که حداکثر تعادل بین نرخ بیت و کیفیت بصری فراهم شود. در [80] نیز یک روش برای بهینه‌سازی نرخ بیت - اعوجاج پیشنهاد شده است که بر اساس شباهت ساختاری است و کیفیت دریافتی را بهتر از معیار میانگین مربعات خطا نشان می‌دهد. این الگوریتم در سطح فریم، بر اساس الگوریتم تخمین شباهت ساختاری¹ آماری بدون مرجع و مدل نرخ بیتی عمل می‌کند که از اطلاعات جانبی و آنتروپی باقی‌مانده‌های تبدیل استفاده می‌کند. در سطح ماکروبلوک، این الگوریتم پیشنهادی بر اساس اطلاعات تئوری که اطلاعات حرکت محتوا را در نظر می‌گیرد تنظیم شده و در نهایت مود مناسب برای کد کردن ماکروبلوک بر اساس این اطلاعات استخراج می‌شود. در [81] یک روش بهینه‌سازی نرخ بیت - اعوجاج بر اساس دریافت بصری ارائه شده است که کارایی فشرده‌سازی را به‌خصوص از جهت سطح کیفیت بالا می‌برد.

در [82] یک روش بهینه‌سازی نرخ بیت - اعوجاج آنلاین برای ویدئوی برجسته‌نما معرفی شده است. در [83] روشی برای بهینه‌سازی نرخ بیت - اعوجاج - پیچیدگی برای کدکننده ویدئوی برجسته‌نما و برای کاربردهای موبایل ارائه شده است. در [84] با تحلیل خصوصیات انتقال خطا در ویدئوی چندنمایی، یک مدل تخمینی پیوندی برای انتقال اعوجاج معرفی می‌شود. سپس یک الگوریتم انتخاب مود به‌منظور بهینه‌سازی نرخ بیت - اعوجاج برای منبع و کانال به‌صورت توأم و براساس این مدل پیشنهاد می‌شود. اختصاص نرخ بیت ناحیه-ای و الگوریتم بهینه‌سازی نرخ بیت - اعوجاج برای ویدئوی چندنمایی به‌اضافه ویدئوی عمق

¹ Structural SIMilarity (SSIM)

نیز در [85] پیشنهاد شده است که نرخ بیت بیشتر را به نواحی رنگی دارای بافت و نرخ بیت کمتر را به نواحی رنگی مسطح اختصاص می‌دهد. روش پیشنهادی در [86] فرستنده را قادر می‌سازد که مجموعه‌ای از نماهای کدشده و نرخ کدشدن آنها را انتخاب کند، به نحوی که مجموع اعوجاج روی زنجیره‌ای از نماهای بازتولیدشده حداقل شود. مقاله [87] یک چارچوب برای اختصاص نرخ بیت بر اساس عمومیت نماها در کدکردن ویدئوی چندنمایی در استاندارد H.264 و براساس ارتباطات ویدئوهای چندنمایی وقتی که محدودیت نرخ بیت داریم، ارائه می‌کند. در [88] توضیح داده شده است که اطلاعات ویدئوی عمق تفاوت زیادی با ویدئوی بافت دارد. بنابراین پارامتر λ که برای بهینه‌سازی نرخ بیت - اعوجاج برای نماهای بافت استفاده می‌شود، برای کدکردن اطلاعات عمق مناسب نیست. در این مقاله، روشی پیشنهاد شده است که بر اساس آن مود بهینه کد کردن ماکروبلوک‌ها بر اساس نواحی غیرمتصل اطلاعات عمق و نوع فریم تعیین می‌شود. در نهایت، [89] توضیح داده است که فرمت درهم‌بافته¹ در ویدئوی سه‌بعدی در سال‌های اخیر بسیار مورد توجه قرار گرفته و با همه استانداردهای ویدئوی دوبعدی نیز سازگار است. اما روش‌های بهینه‌سازی نرخ بیت - اعوجاج برای کدکردن این ویدئوها سناریوی خاص کاربردهای این ویدئو را که در آنها نمونه‌برداری رو به بالا² یک نیاز اساسی برای نمایش ویدئو است، در نظر نمی‌گیرند. در روش ارائه شده در این مقاله این موضوع مدنظر قرار گرفته شده است.

اگرچه این روش‌ها می‌تواند برای بهینه‌سازی نرخ بیت - اعوجاج برای ویدئوی تک‌نمایی و چندنمایی استفاده شود، ولی بر اساس دانسته‌های ما، یک روش جامع برای بهینه‌سازی نرخ بیت - اعوجاج برای ویدئوی چندنمایی مقیاس‌پذیر وجود ندارد. اهمیت این روش از آن جهت است که الگوریتم‌های ارائه‌شده برای ویدئوی چندنمایی به‌سادگی قابل تعمیم به حالت مقیاس‌پذیر نیست، زیرا در محیط مقیاس‌پذیر، شرایط شبکه و محدودیت‌های گیرنده می‌تواند برای هر گیرنده خاص متفاوت باشد. روش پیشنهادی برای بهینه‌سازی نرخ بیت - اعوجاج ویدئوی چندنمایی مقیاس‌پذیر در این رساله، این موارد را در نظر می‌گیرد.

۲-۶- جمع‌بندی

در این فصل مقیاس‌پذیری‌های ارائه‌شده برای ویدئوی تک‌نمایی و چندنمایی، روش‌های ارائه‌شده برای تخصیص داده‌ها به لایه‌های مختلف پایه و افزایشی، مدل‌های موجود برای نرخ بیت - اعوجاج ویدئوهای چندنمایی سه‌بعدی، روش‌های ارزیابی کمی و بصری کیفیت

¹ Interleave

² Upsampling

در ویدئوهای چندنمایی و سه‌بعدی و روش‌های بهینه‌سازی نرخ بیت - اعوجاج مورد بررسی قرار گرفته‌اند. تاکنون رویه‌ای روش‌مند برای استخراج مقیاس‌پذیری بسته به کاربرد مورد نظر ارائه نشده است. از سوی دیگر، روش‌های ارائه‌شده موجود برای نحوه تخصیص داده‌ها به لایه‌های مختلف، ویژگی‌های مقیاس‌پذیری مورد استفاده را در نظر نمی‌گیرند. در زمینه مدل نرخ بیت - اعوجاج، معیار ارزیابی کمی کیفیت و اختصاص بهینه نرخ بیت به لایه‌های مختلف و محتوای آنها نیز کارهای ارائه شده تنها برای ویدئوهای چندنمایی سه‌بعدی است و به حالت مقیاس‌پذیر تعمیم داده نشده است. چارچوب ارائه شده برای کدکردن ویدئوی چندنمایی مقیاس‌پذیر در این رساله، کاستی‌های مطرح‌شده را در نظر می‌گیرد.

فصل سوم

روش‌شناسی ارائه‌شده برای استخراج مقیاس‌پذیری

مناسب برای ویدئوی چندکانالی بسته به کاربرد مورد نظر

معمولاً ویژگی‌هایی از ویدئوهای یک‌نمایی، چندنمایی و سه‌بعدی به صورت غیر روش‌مند به منظور تعریف مقیاس‌پذیری‌های مختلف مورد استفاده قرار می‌گیرد. به این ترتیب نمی‌توان گفت که انواع مختلف مقیاس‌پذیری‌های موجود برای همه کاربردهای ویدئوهای چندنمایی و سه‌بعدی مناسب و کافی باشند. در نتیجه، نیاز به یافتن تعاریف جدیدی برای مقیاس‌پذیری با توجه به خصوصیات اصلی ویدئوی چندنمایی سه‌بعدی داریم. در قسمت اول چارچوب پیشنهادی در این رساله از نظریه مبنایی^۱ به منظور یافتن انواع جدید مقیاس‌پذیری برای کاربردهای مختلف ویدئوی چندنمایی سه‌بعدی استفاده شده است.

در این فصل ابتدا در بخش ۳-۱ و زیربخش‌های آن نظریه مبنایی را بررسی می‌کنیم. سپس روش اعمال نظریه مبنایی و گام‌های مختلف آن به منظور استخراج مقیاس‌پذیری‌های مورد نیاز در بخش ۳-۲ و زیربخش‌های آن بررسی می‌شود. مقیاس‌پذیری‌های استخراج شده از این روش و لزوم استفاده از آنها در کاربردهای واقعی ویدئوی چندنمایی سه‌بعدی نیز در بخش ۳-۳ و زیربخش‌های آن توضیح داده می‌شود.

۳-۱- نظریه مبنایی

نظریه مبنایی یک روش تحقیقی کیفی است که درک ما از یک پدیده را به صورت استقرایی توسعه می‌دهد [90]. اصل غالب این نظریه این است توسعه یافتن درک ما از پدیده مورد نظر با جمع‌آوری و تحلیل هم‌زمان داده‌ها در مورد آن پدیده میسر می‌شود. نظریه مبنایی از گام‌های مختلفی تشکیل شده است که اجرای دقیق آنها رسیدن به یک نتیجه قابل اطمینان را تضمین می‌کند. ایده اصلی این نظریه خواندن حجم زیادی از منابع مرتبط و پیدا کردن مفاهیم، خصوصیات، دسته‌بندی‌ها و ارتباطات بین آنها است [90]. در این نظریه، جمع‌آوری داده‌ها و تحلیل آنها در سه مرحله انجام می‌گیرد: مرحله بازکردن اطلاعات^۲، مرحله کشف ارتباطات بین داده‌ها^۳، مرحله انتخابی^۴.

هدف مرحله بازکردن اطلاعات، تشخیص مفاهیم اصلی در مورد موضوع مورد بررسی است. هدف اصلی منابع مورد مطالعه، مشکلی که در آن منابع مورد بحث قرار گرفته است و راه-حل‌های پیشنهادی در هر یک از منابع، برای یافتن مفاهیم مورد استفاده قرار می‌گیرند. سپس از تفاوت بین مفاهیم مختلف به عنوان معیاری برای دسته‌بندی آنها در دسته‌های

¹ Grounded Theory

² Open coding step

³ Axial coding step

⁴ Selective coding step

متفاوت استفاده می‌شود. مفاهیمی که به یک ایده یکسان اشاره می‌کنند، در دسته‌های یکسان قرار داده شده و مفاهیم جدید و متفاوت در دسته‌ای جدید جای می‌گیرند [90]. در مرحله کشف ارتباطات بین داده‌ها، به دنبال روابط بین دسته‌های متفاوت خواهیم بود. در نهایت در مرحله انتخابی، یکی از گروه‌های موجود به‌عنوان هسته اصلی انتخاب شده و ارتباط بین بقیه گروه‌ها با آن تعیین می‌شود [91]. هسته اصلی، موضوع اصلی مورد بحث در تحقیق موجود است [90].

به این ترتیب، نظریه مبنایی رویه‌ای برای استخراج نتایج با در نظر گرفتن محدوده وسیعی از منابع مطالعاتی و جستجو کردن موضوع مورد مطالعه در حوزه‌های مرتبط است. در مقالات مختلفی از این نظریه برای حل مسائل مهندسی استفاده شده است. برای مثال در [90] یک چارچوب اصولی^۱ پیشنهاد شده است که با استفاده از نظریه مبنایی به تعریف دامنه و حوزه اصلی داده‌کاوی^۲ و کشف دانش^۳ می‌پردازد. به‌عنوان مثالی دیگر، [92] ویژگی‌های مهندسی روش^۴ را با نظریه مبنایی ترکیب کرده است. مهندسی روش در حوزه‌های سامانه‌های اطلاعاتی، روشی برای ساختن رویه‌های جدید از روی رویه‌های موجود است. تأکید این روش روی طراحی، ساختن و ارزیابی رویه‌ها، تکنیک‌ها و ابزارهای حمایتی برای توسعه سامانه‌های اطلاعاتی است. مهندسی روش معمولاً رویه‌ای استقرایی بر مبنای مطالعه پیشینه تحقیق را دنبال می‌کند. استفاده از تکنیک‌های ارائه‌شده در نظریه مبنایی، منجر به توسعه رویه‌ای روش‌مند برای استخراج دانش در حوزه مورد بررسی می‌شود. در ادامه مراحل مختلف نظریه مبنایی به تفصیل توضیح داده خواهد شد.

۳-۱-۱- مرحله بازکردن اطلاعات

در این مرحله، باید مفاهیم اصلی و خصوصیات زمینه تحقیقی مورد نظر را پیدا کنیم. مفاهیم می‌توانند در یک فرآیند تکراری و با مراجعه مستمر به منابع استخراج شوند. هدف اصلی هر یک از این منابع، مسأله اصلی که در آنها مورد توجه قرار می‌گیرد و نیز راه‌حل پیشنهادی برای حل آن مسأله، می‌توانند به‌منظور استخراج مفاهیم اصلی مورد استفاده قرار گیرند. وقتی مفاهیم استخراج شدند، باید با استفاده از شباهت‌ها و تفاوت‌های موجود بین مفاهیم استخراج شده، این مفاهیم در سطوح بالاتر دسته‌بندی شوند. در اکثر مواقع، مفاهیم نام‌های متفاوتی دارند ولی به یک ایده یکسان اشاره می‌کنند. ایده اصلی هر یک از

¹ Systematic

² Data mining

³ Knowledge discovery

⁴ Method Engineering

این مفاهیم برای دسته‌بندی آنها در سطوح بالاتر مورد استفاده قرار می‌گیرد. این فرآیند تا زمانی که متقاعد شویم دسته‌بندی‌ها همه مفاهیم را دربرمی‌گیرند، ادامه پیدا می‌کند.

۳-۱-۲ - مرحله کشف ارتباطات بین داده‌ها

در این مرحله، داده‌های به‌دست آمده از مرحله قبل، مجدداً در سطحی بالاتر دسته‌بندی می‌شوند. در حالت کلی، دسته‌هایی که در مرحله بازکردن اطلاعات استخراج شده‌اند، ارتباطاتی ذاتی با یکدیگر دارند. در این مرحله این ارتباطات استخراج شده و زیردسته‌هایی مرتبط با دسته‌های اصلی ایجاد می‌شوند. سپس ارتباط بین دسته‌ها و زیردسته‌های آنها مشخص می‌شود. به‌منظور استخراج این ارتباطات ممکن است نیاز به بررسی مجدد منابع اولیه باشد. استخراج این ارتباطات توصیفی سلسله‌مراتبی از موضوع مورد بررسی فراهم می‌کند.

۳-۱-۳ - مرحله انتخابی

هدف این مرحله کامل‌کردن منطقی مطالعات با جمع‌کردن همه اطلاعات موجود و فراهم کردن توضیحات نظری برای موضوع مورد مطالعه است. به این منظور ابتدا هسته اصلی مطالعه انتخاب می‌شود. هسته اصلی، مسأله اساسی و زمینه مرکزی بحث است. سپس باید ارتباط بین هسته اصلی و سایر دسته‌ها تعیین شود. زیردسته‌ها در واقع به نوعی توصیف‌کننده هسته اصلی خواهند بود که خصوصیات، اهمیت، اثرمتقابل و روشی برای فهم هسته اصلی را فراهم می‌کنند. ارتباطات بین هسته اصلی و زیردسته‌ها نشان می‌دهند که مفاهیم مرتبط با یک دسته چگونه می‌توانند هسته اصلی را توصیف کنند. نکته قابل توجه اینکه، روند استفاده‌شده در نظریه‌مبنایی ایجاد می‌کند که به منظور ارزیابی نتایج حاصل از این نظریه، لزوم و اهمیت این نتایج در حوزه مورد نظر بررسی شود [93].

۳-۲ - اعمال نظریه‌مبنایی به‌منظور استخراج انواع مقیاس‌پذیری‌های مناسب برای ویدئوی چندنمایی سه‌بعدی

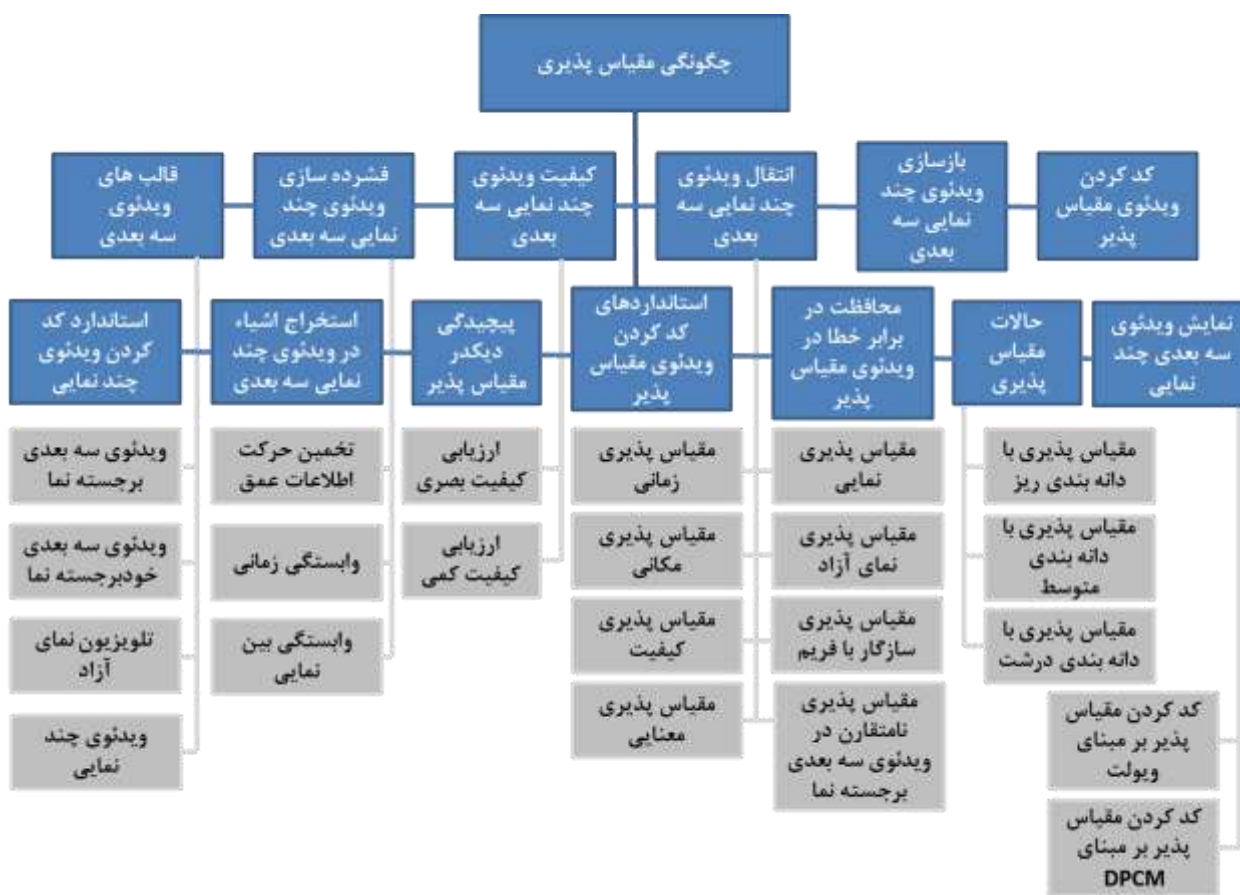
منابع مورد استفاده در این رساله برای نظریه‌مبنایی برای استخراج مقیاس‌پذیری‌های مناسب برای ویدئوی چندنمایی سه‌بعدی، ۱۱۴ مقاله از مجلات و کنفرانس‌ها و رساله‌های دکتری در ارتباط با "مقیاس‌پذیری ویدئو"، "ویدئوی سه‌بعدی" و "کدکردن ویدئوی چندنمایی" است که فعالیت‌های تحقیقاتی در این زمینه‌ها را در ده سال گذشته نشان می‌-

دهد. در ادامه به شرح نحوه اعمال گام‌های مختلف نظریه مبنایی برای استخراج انواع مناسب مقیاس‌پذیری برای ویدئوی چندنمایی سه‌بعدی خواهیم پرداخت.

۳-۲-۱- مرحله بازکردن اطلاعات

همان‌طور که اشاره شد در این مرحله باید مفاهیم مرتبط با موضوع مورد بررسی و دسته‌بندی‌های مرتبط با آنها از منابع موجود استخراج شود. برای موضوع مورد مطالعه، می‌توان یک دسته با عنوان "قالب ویدئوی سه‌بعدی"^۱ داشت که مفاهیم قالب‌های ویدئوی سه‌بعدی بر حسب نوع نمایشگر، روش‌هایی که برای تولید و نمایش آنها استفاده می‌شود و روش‌هایی که برای غلبه بر محدودیت‌های ویدئوی دوبعدی وجود دارد، را شامل شود. همه این مفاهیم به تفصیل در پیشینه تحقیق مورد بحث قرار گرفته‌اند. همچنین، مفاهیم فشردسازی کارآمد ویدئوی چندنمایی و فشردسازی و تخمین حرکت نقشه عمق ویدئوی سه‌بعدی، تحت عنوان "فشردسازی ویدئوی چندنمایی سه‌بعدی" دسته‌بندی می‌شوند. شکل ۱-۳ مفاهیم استخراج شده از منابع و دسته‌بندی‌های آنها را نشان می‌دهد.

^۱ 3D video format



شکل ۳-۱ - دسته‌ها، زیردسته‌ها و هسته اصلی که در گام‌های مختلف نظریه مبنایی به دست آمده‌اند.

۳-۲-۲ - مرحله کشف ارتباطات بین داده‌ها

در این مرحله، برای هر دسته تعدادی زیردسته پیدا می‌کنیم، به نحوی که همه جنبه‌های موضوع مورد مطالعه شامل "مقیاس‌پذیری ویدئو"، "ویدئوی سه‌بعدی" و "کد کردن ویدئوی چندنمایی" را شامل شود. دسته‌هایی که در مرحله بازکردن اطلاعات مشخص شده‌اند، به‌طور ذاتی با هم ارتباط دارند. این ارتباطات باید در این مرحله تشخیص داده شوند. زیردسته‌ها در واقع دسته‌هایی هستند که خصوصیات مشخص و مخصوصی دارند و می‌توانند اطلاعات مفیدی برای دسته رده بالاتر فراهم کنند. زیردسته‌های استخراج شده در مورد موضوع مورد مطالعه نیز در شکل ۳-۱ نشان داده شده‌است.

۳-۲-۳ - مرحله انتخابی

در این مرحله به منظور کامل کردن منطقی مطالعات، ابتدا هسته اصلی مطالعه انتخاب می‌شود که هسته اصلی، مسأله اساسی و زمینه مرکزی بحث است. سپس ارتباط بین هسته اصلی و سایر دسته‌ها تعیین می‌شود.

بنابراین، در موضوع مورد مطالعه در این رساله، هسته اصلی "تعاریف مختلف مقیاس-پذیری" خواهد بود که موضوع اصلی بحث است. جدول ۳-۱ مروری مفهومی از ارتباطات استخراج‌شده و مفاهیم مرتبط با آنها را نشان می‌دهد. ستون اول ارتباط بین "تعاریف مقیاس‌پذیری" به عنوان هسته اصلی و بقیه دسته‌ها را نشان می‌دهد که در مراحل مختلف تحقیق به دست آمده‌اند.

نحوه استخراج این ارتباطات را با ذکر مثالی توضیح می‌دهیم. قالب‌های ویدئوی سه‌بعدی در مقایسه با ویدئوی تک‌نمایی اولیه، خصوصیات جدیدی دارند، مانند قالب ویدئوی سازگار با فریم که در فصل ۱- توضیح داده شده‌است. این ویژگی باید در تعاریف مقیاس‌پذیری ویدئوی سه‌بعدی چندنمایی، به‌منظور استفاده از مقیاس‌پذیری در کاربردهای جدید مرتبط منعکس شود. به این ترتیب، "فرمت ویدئوی سه‌بعدی" به عنوان یکی از دسته‌های استخراج شده، نیازمندی‌هایی را به هسته اصلی اضافه می‌کند. یکی از ارتباطات تعریف شده در جدول ۳-۱ با عنوان "اضافه کردن تعاریف مقیاس‌پذیری به هسته اصلی" بر اساس این بحث استخراج شده‌است. به‌طور مشابه، ویژگی‌هایی از هر دسته به‌منظور استخراج روابط مناسب بین این دسته‌ها و هسته اصلی استخراج شده‌اند. این ویژگی‌ها در ستون دوم جدول ۳-۱ نشان داده شده‌اند.

در ادامه و به‌منظور ارزیابی نتایج حاصل از نظریه مبنایی، ضرورت استفاده از مقیاس-پذیری‌های استخراج‌شده و قابلیت‌های آنها را در کاربردهای واقعی ویدئوی چندنمایی و سه‌بعدی بررسی کنیم.

۳-۳- تعاریف استخراج شده برای مقیاس‌پذیری در ویدئوی چندنمایی و لزوم استفاده از آنها در کاربردهای واقعی

تعاریف جدید استخراج شده برای مقیاس‌پذیری با استفاده از نظریه مبنایی و ویژگی‌های مرتبط با آنها به ترتیب در ستون‌های اول و دوم جدول ۳-۲ آورده شده‌اند. بعضی از تعاریف ارائه‌شده در این جدول، مانند مقیاس‌پذیری نمایی، مقیاس‌پذیری نمای آزاد، مقیاس‌پذیری مکانی نامتقارن و مقیاس‌پذیری سازگار با فریم قبلاً در فصل دوم معرفی شده‌اند. بقیه موارد، ویژگی‌های مرتبط با آنها و لزوم استفاده از آنها در کاربردهای واقعی، در ادامه توضیح داده خواهند شد.

جدول ۳-۱- ارتباطات بین هسته اصلی و زیردسته‌ها

دسته ها و ارتباطات آنها با هسته اصلی	مفاهیم متناظر برای استخراج ارتباطات
کد کردن ویدئوی مقیاس پذیر (دنبال کردن)	<ul style="list-style-type: none"> • کد کردن مبتنی بر اشیاء ویدئو • انعطاف پذیری در انتقال ویدئو
استانداردهای ویدئوی مقیاس پذیر و چندنمایی قالب های ویدئوی سه بعدی چندنمایی (اضافه کردن انواع جدید مقیاس پذیری به هسته اصلی)	<ul style="list-style-type: none"> • قالب ویدئوی سه بعدی چندنمایی با وضوح متفاوت
کیفیت ویدئوی سه بعدی چندنمایی بازسازی ویدئوی سه بعدی چندنمایی محافظت در برابر خطا در ویدئوی مقیاس پذیر حالت‌های مقیاس پذیری (پشتیبانی شده توسط هسته اصلی)	<ul style="list-style-type: none"> • میزان دریافت بیننده از اطلاعات عمق • نظریه محو شدگی • اطلاعات عمق • ازلاعات جانبی برای بازسازی نماهای مجازی
انتقال ویدئوی سه بعدی چندنمایی (بهره گیری از هسته اصلی)	<ul style="list-style-type: none"> • انتقال روی شبکه های مستعد خطا
استخراج اشیاء در ویدئوی سه بعدی چندنمایی پیچیدگی دیکدر مقیاس پذیر فشرده سازی ویدئوی سه بعدی چندنمایی نمایش ویدئوی سه بعدی چندنمایی (ساده سازی شده با استفاده از هسته اصلی)	<ul style="list-style-type: none"> • اطلاعات عمق • قابلیت انطباق رشته بیت با گیرنده های نامتناجس • تجزیه به ویدئوهای با پیچیدگی پایین

۳-۳-۱- مقیاس پذیری عمق

اطلاعات عمق یکی از پارامترهای مهم ویدئوی سه بعدی است که فاصله بین اشیاء موجود در صحنه را از دوربین نشان می دهد. در این نوع مقیاس پذیری، رشته بیت های مقیاس پذیر، منطبق با قسمت های مختلف صحنه براساس فاصله شان از دوربین هستند. به عنوان مثال، لایه پایه می تواند شامل نواحی از صحنه باشد که فاصله کمتری تا دوربین دارند و لایه های افزایشی بقیه نواحی را شامل می شوند. به این ترتیب، همه گیرنده ها، نواحی نزدیک به دوربین را دریافت کرده و قادر خواهند بود این نواحی را بازسازی کنند. همانند سایر تعاریف مقیاس پذیری، هر لایه از مقیاس پذیری عمق می تواند با سایر مقیاس پذیری ها مانند مقیاس - پذیری زمانی، مکانی و غیره ترکیب شده و مقیاس پذیری پیوندی^۱ معرفی کند.

^۱ Hybrid scalability

جدول ۳-۲ - تعاریف استخراج شده برای مقیاس‌پذیری در ویدئوی سه بعدی چندنمایی با استفاده از نظریه مبنایی و ویژگی‌های متناظر آنها

ویژگی‌های مرتب	مقیاس‌پذیری‌های استخراج شده
انعطاف‌پذیری در انتقال ویدئو	مقیاس‌پذیری نمایی
قابلیت انطباق رشته بیت با گیرنده‌های نامتناجس	مقیاس‌پذیری نمای آزاد
قالب ویدئوی سه بعدی چندنمایی با وضوح متفاوت	مقیاس‌پذیری مکانی غیرمتقارن (برای ویدئوی برجسته نما)
قالب ویدئوی سه بعدی چندنمایی با وضوح متفاوت	مقیاس‌پذیری سازگار با فریم
اطلاعات عمق	مقیاس‌پذیری عمق
قابلیت انطباق رشته بیت با گیرنده‌های نامتناجس تجزیه به ویدئوهای با پیچیدگی پایین	مقیاس‌پذیری پیچیدگی
قابلیت انطباق رشته بیت با گیرنده‌های نامتناجس تجزیه به ویدئوهای با پیچیدگی پایین	مقیاس‌پذیری سطح
اطلاعات عمق (تنها برای ویدئوهای برجسته نما)	مقیاس‌پذیری وضوح - عمق یا کیفیت - عمق
اطلاعات جانبی برای بازسازی نماهای مجازی	مقیاس‌پذیری اطلاعات جانبی
نظریه محو‌شدگی	مقیاس‌پذیری بر اساس دریافت‌بیننده با استفاده از نظریه محو‌شدگی

۳-۱-۱-۳ - مثالی از کاربرد مقیاس‌پذیری عمق

محیط مجازی همه‌جانبه در مقیاس وسیع^۱ [94] در یک کلاس آموزش مثالی از کاربردهای ویدئوی چندنمایی سه‌بعدی است. در این محیط، چهره همه شرکت‌کنندگان به صورت سه‌بعدی نمایش داده می‌شود و فاصله شرکت‌کنندگان از یکدیگر به عنوان عمق در نظر گرفته می‌شود. اندازه و کیفیت چهره شخصی که در مرکز توجه است باید بیشتر از سایرین باشد. علاوه بر این، چنین محیطی این امکان را فراهم می‌کند که تعداد زیادی شرکت‌کننده با استفاده از محتوای چندرسانه‌ای با یکدیگر ارتباط متقابل برقرار کنند. بنابراین، هر کاربر هم تولیدکننده و هم گیرنده اطلاعات است. این مسأله نیاز به ارتباط هم‌زمان بین تعداد زیادی کاربر با محتوای گوناگون چندرسانه‌ای مانند ویدئو، صوت، پیام متنی و غیره دارد. با این حال، در هر لحظه از زمان، هر کاربر معمولاً با تعداد محدودی کاربر دیگر ارتباط دارد. ارسال تمام اطلاعات به همه کاربران بدون توجه به اینکه به محتوای داده ارسالی نیاز دارند یا نه، نه تنها کارآمد نیست، بلکه با اعمال ترافیک اضافی روی شبکه روی کیفیت داده‌های دریافتی نیز تأثیر می‌گذارد. کاربران این محیط روی چندین دامنه ارتباطی توزیع شده‌اند. تفاوت در پهنای باند و توان پردازشی گیرنده‌ها، عدم تجانس این محیط را بالا می‌برد. مقیاس‌پذیری و به‌خصوص مقیاس‌پذیری عمق، یک راه‌کار مناسب برای کاهش حجم داده‌ها

¹ Large Scale Immersive Virtual Environment (LSIVE)

در این کاربرد است. در مقیاس‌پذیری عمق برای این کاربرد، لایه پایه برای هر کاربر از نواحی از صحنه تشکیل شده که فاصله خاصی از آن کاربر دارند. لایه افزایشی شامل باقی نواحی صحنه (نواحی با اهمیت کمتر) می‌شود. بنابراین، تنها با دریافت لایه پایه، همه گیرنده‌ها قادر خواهند بود با توجه به زاویه دید خود نواحی مهم صحنه را برای خود بازسازی کنند. گیرنده‌هایی با منابع بیشتر می‌توانند لایه‌های افزایشی را نیز دریافت کرده و کیفیت دریافتی و تجربی خود را بالا ببرند. شکل ۲-۳ (الف) محتوای لایه پایه برای معلم این کلاس آموزشی را در این نوع مقیاس‌پذیری نشان می‌دهد. در لایه پایه تنها یک نفر که در مرکز توجه است، وجود دارد. در شکل ۲-۳ (ب) لایه پایه به‌علاوه لایه‌های افزایشی دیده می‌شود که همه نواحی صحنه را شامل می‌شود.



(الف)



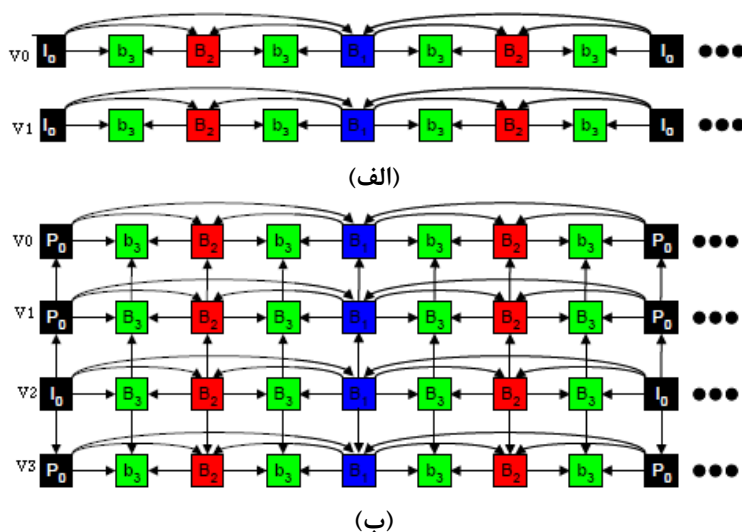
(ب)

شکل ۲-۳ - مقیاس‌پذیری عمق برای کلاس آموزشی (الف) لایه پایه (ب) لایه‌های پایه و افزایشی

۳-۳-۲ - مقیاس پذیری پیچیدگی

ساختار پیش‌بینی پیشنهاد داده شده برای کد کردن ویدئوی چندنمایی این امکان را فراهم می‌کند که هر نما علاوه بر اینکه به صورت زمانی پیش‌بینی می‌شود، از روی سایر نماها و به صورت بین‌نمایی نیز پیش‌بینی شود. این ساختار انطباقی است و بهترین فریم از میان فریم‌های زمانی و بین‌نمایی می‌تواند برای هر فریم بر اساس هزینه نرخ بیت - اعوجاج - پیچیدگی انتخاب شود [8].

ساختار پیش‌بینی پیشنهاد شده می‌تواند به عنوان معیاری برای کنترل پیچیدگی کدگشای ویدئوی چندنمایی در نظر گرفته شود. در مقیاس‌پذیری پیچیدگی، رشته‌بیت‌های مقیاس‌پذیر، ساختارهای پیش‌بینی متفاوت خواهند داشت. به این ترتیب گیرنده قادر خواهد بود رشته‌بیت ویدئویی متناسب با محدودیت‌های پیچیدگی محاسباتی خود انتخاب کند. برای مثال، لایه پایه می‌تواند ساده‌ترین شکل ساختار پیش‌بینی را داشته باشد، مانند حالتی که همه نماها فقط به صورت زمانی کد می‌شوند [8]. لایه‌های افزایشی می‌توانند با ساختار پیش‌بینی پیچیده‌تر که امکان پیش‌بینی بین‌نمایی را نیز فراهم می‌کند، کد شوند که کارایی فشرده‌سازی را بالا می‌برد و در عوض پیچیدگی محاسباتی بالایی نیز دارد. شکل ۳-۳ (الف) ساختار پیش‌بینی مناسب برای لایه پایه در مقیاس‌پذیری پیچیدگی که در آن فقط از پیش‌بینی زمانی استفاده شده است، را نشان می‌دهد. شکل ۳-۳ (ب) ساختار پیش‌بینی مناسب برای لایه‌های افزایشی که پیش‌بینی بین‌نمایی را نیز پشتیبانی می‌کند نشان می‌دهد.



شکل ۳-۳ - ساختار پیش‌بینی مناسب برای (الف) لایه پایه و (ب) لایه‌های افزایشی در مقیاس‌پذیری پیچیدگی

۳-۳-۱- مثالی از کاربرد مقیاس پذیری پیچیدگی

در کاربرد ویدئوی نمای آزاد^۱، بیننده می تواند از بین نماهای موجود، نماهای دلخواه را انتخاب کند. بنابراین، گیرنده به همه نماهای موجود نیازی ندارد. با در نظر گرفتن وابستگی بین نمایی ذاتی موجود در فرآیند پیش بینی نماها با استفاده از ساختار پیش بینی بین نمایی استاندارد کد کردن ویدئوی چندنمایی، به منظور واكد کردن یک نما، همه نماهای استفاده شده در ساختار پیش بینی باید واكد شوند. این کار در همه کاربردهای موجود ممکن و کارآمد نیست. در این حالت، مقیاس پذیری پیچیدگی می تواند به منظور تولید رشته بیتی با ساختار پیش بینی محدود مورد استفاده قرار گیرد. با استفاده از این ساختار پیش بینی محدود، یک نمای مشخص می تواند با پیچیدگی محاسباتی کمتر و البته با کارایی فشرده سازی کمتر واكد شود.

شایان ذکر است که اگرچه مقیاس پذیری پیچیدگی قبلاً در مورد ویدئوی تک نمایی مورد مطالعه قرار گرفته است، اما مفهوم پیچیدگی در ویدئوی چندنمایی سه بعدی متفاوت است زیرا منبع اصلی پیچیدگی در این ویدئوها، پیش بینی بین نمایی است. بنابراین این نوع مقیاس پذیری با مقیاس پذیری پیچیدگی در حالت تک نمایی متفاوت است.

۳-۳-۳- مقیاس پذیری سطح

پروفایل ها و سطوح در یک استاندارد خاص کد کردن ویدئو مانند H.264/AVC، محدودیت هایی را در مورد قابلیت های مورد نیاز برای واكد کردن ویدئو اعمال می کنند. برای هر پروفایل مشخص، سطوح معمولاً منطبق بر سربار پردازشی واكدکننده و قابلیت حافظه هستند. در مقیاس پذیری سطوح، لایه پایه می تواند شامل ویدئوی گذشته با پایین ترین عدد موجود برای سطح در پروفایل انتخاب شده باشد که در این صورت، حداقل نیازهای آن پروفایل پشتیبانی می شود. برای مثال، برای اندازه فریم، لایه پایه می تواند از سطوحی استفاده کند که حداقل تعداد نمونه های روشنایی و ماکروبلوک ها را دارند (۱۴۵۸ نمونه روشنایی و ۳۸۰۱۶۰ ماکروبلوک). سپس، لایه های افزایشی از سطوح دیگر پروفایل با تعداد نمونه های روشنایی و ماکروبلوک های بیشتر استفاده می کنند. برای مثال، لایه افزایشی اول می تواند از سطوحی استفاده کند که ۳۰۰۰ نمونه روشنایی و ۷۶۸۰۰۰ ماکروبلوک را پشتیبانی می کنند و لایه افزایشی دوم می تواند از سطوحی با ۱۹۸۰۰ نمونه روشنایی و

¹ Free View-point Video (FVV)

۵۰۶۸۸۰۰ ماکروبلوک استفاده کند. به‌طور مشابه، برای حداکثر سرعت واکدکردن، حداکثر نرخ ویدئو و غیره می‌توان از قابلیت‌های متفاوت سطوح مختلف یک پروفایل استفاده کرد.

۳-۳-۱- مثالی از کاربرد مقیاس‌پذیری سطح

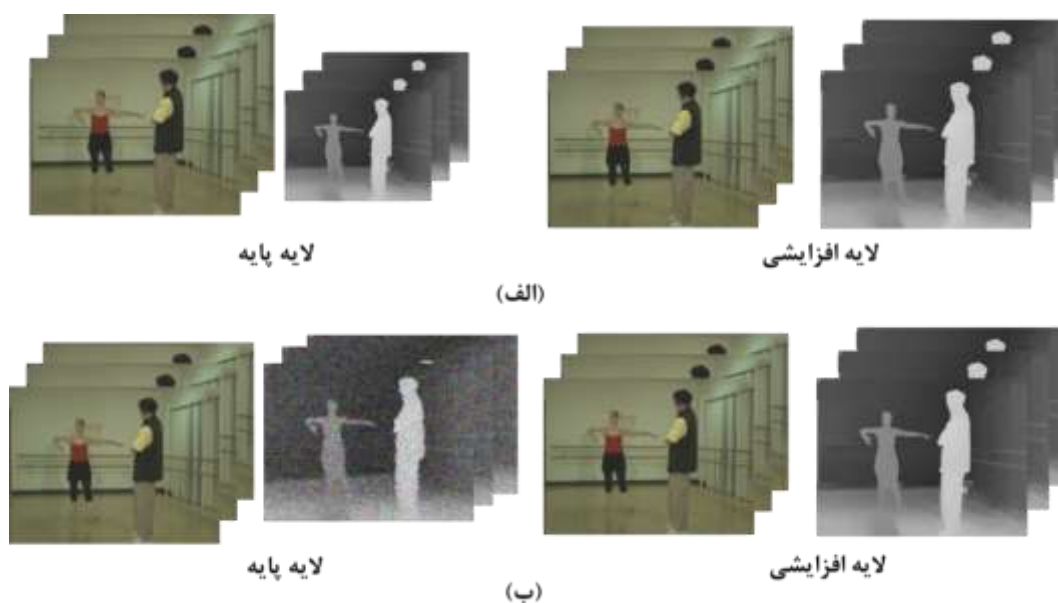
تلویزیون سه‌بعدی موبایل^۱ یکی از کاربردهای اصلی ویدئوی سه‌بعدی است که با استفاده از مزایای فناوری صفحه کوچک و بدون نیاز به عینک مخصوص می‌توان از آن استفاده کرد. در این کاربرد، بینندگان باید قادر باشند که صحنه سه‌بعدی را از زوایای مختلف تماشا کنند و بنابراین، نماهای زیادی باید به‌صورت هم‌زمان واکد شوند. به این ترتیب، این فرآیند به توان پردازشی بالا نیاز دارد که در بسیاری از کاربردهای موبایل قابل پشتیبانی نیست. در این حالت، مقیاس‌پذیری سطح راه حل مناسبی است، چون می‌تواند رشته‌بیت‌های متفاوتی با سطوح مختلف پیچیدگی محاسباتی تولید کند. با دریافت تنها لایه پایه که با سطوح پایین پروفایل کد شده‌اند، همه گیرنده‌ها قادر خواهند بود، همه نماها را با حداقل پیچیدگی محاسباتی تولید کنند. گیرنده‌هایی با منابع بیشتر می‌توانند لایه‌های افزایشی بیشتری دریافت کرده و کیفیت دریافتی خود را بهبود بخشند.

۳-۳-۴- مقیاس‌پذیری عمق-وضوح تصویر/عمق-کیفیت تصویر

ویدئوی برجسته‌نما ساده‌ترین فرم ویدئوی سه‌بعدی است. ویدئوی برجسته‌نما متشکل از اطلاعات رنگ و عمق به دلیل اینکه می‌تواند نیازهای پهنای باند و حافظه مصرفی را کاهش دهد، بسیار مورد توجه قرار گرفته است. تحقیقات اخیر نشان می‌دهد که اطلاعات عمق می‌تواند با کارایی بالا و بدون اینکه روی کیفیت تجربی اثر قابل ملاحظه‌ای بگذارد، فشرده شود [51].

در این نوع مقیاس‌پذیری، لایه پایه می‌تواند از اطلاعات عمق با وضوح تصویر یا کیفیت حداقل تشکیل شود. لایه‌های افزایشی می‌توانند شامل رشته‌بیت‌هایی باشند که در آنها اطلاعات عمق، وضوح تصویر و یا سطح کیفیت بالاتری کد شده‌اند. به این ترتیب می‌توان ویدئوهایی با کیفیت دریافتی یکسان و نرخ بیت‌های متفاوت که مناسب برای گیرنده‌هایی با محدودیت پهنای باند مختلف هستند، تولید کرد. شکل ۳-۴، (الف) و (ب)، به ترتیب، لایه پایه با وضوح تصویر و کیفیت دریافتی برای اطلاعات عمق را نشان می‌دهد. لایه‌های افزایشی، شامل اطلاعات عمق با وضوح تصویر و سطح کیفیت اصلی خواهند بود.

^۱ Mobile 3DTV



شکل ۳-۴ - (الف) لایه‌های پایه و افزایشی در ویدئوی برجسته در مقیاس پذیری عمق - وضوح تصویر (ب) لایه‌های پایه و افزایشی در ویدئوی برجسته در مقیاس پذیری عمق - کیفیت تصویر

۳-۴-۱- مثالی از کاربرد مقیاس پذیری عمق-وضوح تصویر/عمق-کیفیت تصویر
مزایای ویدئوهای برجسته‌نما متشکل از ویدئوهای عمق و رنگ در کاربردهای متفاوتی از جمله پردازش ویدئو در کاربردهای پزشکی مورد توجه قرار گرفته است. در این کاربردها، اطلاعات عمق کیفیت تجربی دریافتی را بهبود می‌بخشد. از آنجایی که ویدئوهای مورد استفاده در کاربردهای پزشکی به خطای فشرده‌سازی بسیار حساس هستند، این نوع مقیاس‌پذیری در ویدئوهای برجسته راه حل مناسبی برای حل مسأله فشرده‌سازی حجم بالای اطلاعات این ویدئوها در این کاربرد خاص است. در لایه پایه، اطلاعات عمق با وضوح تصویر و کیفیت پایین‌تر می‌شوند، در این حالت بدون اینکه کیفیت نهایی تحت تأثیر قرار گیرد، نرخ بیت نهایی کاهش می‌یابد. بنابراین، رشته ویدئویی لایه پایه با نرخ بیت پایین تولید می‌شود، بدون اینکه اثر مخربی روی کیفیت نهایی دریافتی داشته باشد. در عین حال، لایه‌های افزایشی می‌توانند کیفیت نهایی دریافتی را به‌طور مؤثر بهبود بخشند.

۳-۳-۵- مقیاس‌پذیری اطلاعات جانبی

بازتولید نماهای مجازی یک فناوری کلیدی برای عملی ساختن ویدئوی نمای آزاد به عنوان یکی از کاربردهای ویدئوی چندنمایی و سه‌بعدی است. نماهای مجازی با استفاده از نماهای مجاور که دریافت شده‌اند و با استفاده از اطلاعات جانبی بازتولید می‌شوند. بنابراین، کیفیت اطلاعات جانبی می‌تواند به‌طور مؤثر بر روی کارایی نماهای بازتولید شده اثر بگذارد. اما این

اطلاعات اضافی باید همراه با رشته ویدئوی چندنمایی ارسال شوند که منجر به ایجاد ترافیک بالا روی شبکه می‌شود. در این نوع مقیاس‌پذیری، لایه پایه می‌تواند از نماهای ارسالی و اطلاعات جانبی با حداقل کیفیت قابل قبول تشکیل شده باشد. لایه‌های افزایشی شامل اطلاعات جانبی گذشته در سطوح مختلف کیفیت خواهند بود. به این ترتیب، هر گیرنده می‌تواند تنها با دریافت لایه پایه، نماهای مجازی را با کیفیت قابل قبول بازتولید کند. کیفیت نماهای بازتولید شده می‌تواند با دریافت اطلاعات لایه‌های افزایشی بهبود یابد.

۳-۵-۱- مثالی از کاربرد مقیاس‌پذیری اطلاعات جانبی

ویدئوی نمای آزاد یکی از کاربردهای مهم ویدئوی چندنمایی است که به کاربر اجازه می‌دهد نماهای موردنظر خود را به صورت تعاملی انتخاب کرده و نماهای مجازی و جدید نیز علاوه بر نماهای دریافت شده، بازتولید کند. مقیاس‌پذیری اطلاعات جانبی می‌تواند به منظور بازتولید نماهای مجازی که تصویربرداری نشده‌اند، استفاده شود. این نوع مقیاس‌پذیری می‌تواند برای گیرنده‌های متنوع با قابلیت‌های متفاوت پهنای باند و توان پردازشی مورد استفاده قرار گیرد.

۳-۳-۶- مقیاس‌پذیری کیفیت ادراکی با استفاده از نظریه محو شدگی

همان‌طور که پیش‌تر توضیح داده شد، در ویدئوی تک‌نمایی، مقیاس‌پذیری‌های زمانی، مکانی و کیفیت به منظور تولید رشته‌بیت‌های مقیاس‌پذیر مورد استفاده قرار می‌گیرند. اما روشی اصولی به منظور تعمیم دادن این نوع مقیاس‌پذیری‌ها از ویدئوی تک‌نمایی به ویدئوی چندنمایی سه‌بعدی وجود ندارد. در این رساله، یک مفهوم شناخته‌شده در ویدئوی چندنمایی سه‌بعدی، با عنوان نظریه محو شدگی^۱ به منظور تعمیم این نوع مقیاس‌پذیری‌ها مورد استفاده قرار می‌گیرد. نظریه محو شدگی می‌گوید که کیفیت بصری یک ویدئوی برجسته دونمایی از روی کیفیت نمایی که کیفیت بالاتری دارد تعیین می‌شود. به عبارت دیگر، کیفیت نمایی که کیفیت بالاتری دارد، کیفیت بصری بیننده را تحت تأثیر قرار می‌دهد [95]. این مفهوم می‌تواند به ویدئوی چندنمایی دوبعدی و سه‌بعدی نیز تعمیم داده شود. در ویدئوی چندنمایی دوبعدی یا سه‌بعدی، اگر کیفیت مجموعه‌ای از نماها که برای بیننده مهم تر هستند، بیشتر باشد، کیفیت نهایی بصری برای بیننده نیز بالاتر خواهد بود. بنابراین، نماهای غیرمهم می‌توانند به صورت مکانی، زمانی، کیفیت و یا ترکیبی از آنها مقیاس شوند.

¹ Suppression theory

به این صورت، نرخ بیت نهایی کاهش می‌یابد بدون اینکه کیفیت نهایی بصری تحت تأثیر قرار گیرد.

در این مقیاس‌پذیری، لایه پایه شامل زیر مجموعه‌ای از نماهای مهم با سطح قابل قبولی از کیفیت، مقیاس زمانی و مکانی خواهد بود. لایه‌های افزایشی نیز شامل سایر نماها با سطوح متفاوت کیفیت، مقیاس زمانی و مکانی خواهند بود. این روش تعمیم دادن منجر به تولید ویدئوهای چندنمایی با نرخ بیت پایین‌تر و کیفیت بصری تقریباً یکسان خواهد شد.

به‌منظور تعمیم روش ارائه شده به ویدئوی چندنمایی سه‌بعدی نیز می‌تواند مقیاس‌پذیری زمانی، مکانی یا کیفیت را به زیرمجموعه‌ای از نماهای راست یا چپ که از جهت‌های متفاوت تصویربرداری شده‌اند، اعمال کرد.

این نوع مقیاس‌پذیری برای کاربردهای مختلف ویدئوی چندنمایی دوبعدی و سه‌بعدی که معمولاً از میزان حافظه مصرفی بالا رنج می‌برند، مناسب است.

۳-۴- جمع‌بندی

از آنجایی که مقیاس‌پذیری‌های ارائه شده برای ویدئوی تک‌نمایی برای همه کاربردهای ویدئوی چندنمایی و سه‌بعدی کافی نیست، در این بخش از رساله، انواع مناسب مقیاس‌پذیری برای ویدئوهای چندنمایی و سه‌بعدی استخراج شده است. به منظور استخراج این موارد، مجموعه بزرگی از نتایج تحقیقات قبلی در مورد ویدئوی چندنمایی سه‌بعدی و مقیاس‌پذیری در دهه اخیر، با استفاده از نظریه مبنایی و به‌صورت روش‌مند مورد بررسی قرار گرفت. در نهایت به‌منظور نشان دادن اهمیت و لزوم مقیاس‌پذیری‌های جدید ارائه‌شده برای ویدئوی چندنمایی سه‌بعدی ضرورت استفاده از آنها در کاربردهای واقعی ویدئوی چندنمایی سه‌بعدی مورد بررسی قرار گرفته است.

فصل چهارم

روش‌شناسی ارائه‌شده برای اختصاص داده‌های

مناسب به لایه‌های پایه و افزایش درویدنومی چندنمایی بر

اساس کاربرد مورد نظر

در قسمت بعدی چارچوب پیشنهادی این رساله، یک روش‌شناسی برای تخصیص مناسب داده‌ها به لایه‌های مختلف ویدئوی چندنمایی مقیاس‌پذیر ارائه خواهد شد به نحوی که میزان کیفیت نهایی حداکثر شود. روش‌شناسی پیشنهادی در بخش ۴-۱- و زیر بخش‌های آن معرفی می‌شود. سپس در زیر بخش‌های ۴-۲- و ۴-۳- برای دو کاربرد خاص ویدئوی چندنمایی سه‌بعدی اعمال و نتایج حاصل از ارزیابی آن ارائه می‌شود.

۴-۱- روش‌شناسی پیشنهادی

۴-۱-۱- تعاریف و مشاهدات

در مقیاس‌پذیری، حجم داده‌های دریافتی هر گیرنده، بسته به نیازها و قابلیت‌های آن مانند پهنای باند و توان پردازشی، متفاوت خواهد بود. بر این اساس، داده‌های تخصیص داده شده به لایه‌های پایه و افزایشی باید به نحوی باشد که کیفیت تجربی قابل قبول برای هر گیرنده حاصل شود. به عبارت دیگر، نحوه تخصیص داده‌ها باید به گونه‌ای باشد که هر گیرنده بتواند با دریافت حجم مناسبی از داده‌ها بر اساس محدودیت‌های خود، به کیفیت تجربی قابل قبول برسد. بنابراین، در روش‌شناسی ارائه شده در این بخش از رساله، دستیابی به کیفیت تجربی مورد نظر، به عنوان معیاری برای میزان موفقیت کارایی روش تخصیص داده‌ها در نظر گرفته می‌شود.

از طرف دیگر، در کد کردن ویدئوی چندنمایی و به منظور افزایش کارایی از وابستگی‌های بین‌نمایی استفاده شده و نماهای موجود از روی هم پیش‌بینی می‌شوند و داده‌های لایه پایه و افزایشی به یکدیگر وابسته خواهند بود. بنابراین وابستگی‌های بین‌نمایی می‌تواند به عنوان عامل تعیین‌کننده برای تخصیص داده‌ها به لایه‌های مختلف استفاده شود، چون روی کیفیت پیش‌بینی و در نتیجه کارایی فشرده‌سازی اثر خواهد داشت. بر این اساس، روش‌شناسی این رساله پیشنهاد می‌کند که از مفهوم وابستگی‌های بین‌لایه‌ای و درون‌لایه‌ای برای اختصاص مناسب داده‌ها به لایه‌های مختلف استفاده شود، به نحوی که حداکثر کیفیت تجربی در سمت گیرنده‌ها و بر اساس نیازهای آنها حاصل شود. نکته قابل توجه اینکه وابستگی بین‌لایه‌ای و درون‌لایه‌ای بر اساس نوع مقیاس‌پذیری مورد استفاده که در قسمت قبلی چارچوب پیشنهادی استخراج شد، تعریف می‌شوند.

۴-۱-۲ - مراحل روش‌شناسی پیشنهادی

مراحل روش‌شناسی ارائه شده به شرح زیر است: ابتدا حجم داده‌های لایه پایه تعیین می‌شود. از آنجایی که همه گیرنده‌ها باید قابلیت دریافت لایه پایه را داشته باشند، حجم داده‌های آن باید بر اساس کمترین پهنای باند موجود بین گیرنده‌ها تعیین شود. سپس از بین داده‌های موجود باید داده‌های مناسب به لایه‌های پایه و افزایشی اختصاص یابند. همان‌طور که اشاره شد، معیار انتخاب داده‌های مناسب برای لایه‌ها کیفیت تجربی مورد نیاز گیرنده‌ها بر اساس کاربرد موردنظر است. بنابراین باید مجموعه‌ای از محدودیت‌ها و نیازمندی‌های گیرنده‌ها از لحاظ معیار کیفیت تجربی را استخراج کنیم. این محدودیت‌ها می‌تواند شامل پهنای باند، توان پردازشی قابل تحمل، قابلیت پشتیبانی نمای آزاد، محدودیت واگد کردن با تأخیر کم^۱ و غیره باشد.

در نهایت، بر اساس محدودیت‌ها و کیفیت تجربی مورد نیاز در کاربرد موردنظر تعیین می‌کنیم که میزان وابستگی بین داده‌های لایه‌های پایه و افزایشی به یکدیگر کم یا زیاد باشد. برای مثال، وقتی گیرنده‌ها محدودیت پهنای باند دارند، به احتمال زیاد، تنها قادر خواهند بود که داده‌های لایه پایه را دریافت کنند و داده‌های لایه‌های افزایشی به‌طور کامل حذف می‌شوند. بنابراین داده‌های لایه‌های پایه و افزایشی باید وابستگی کمی به هم داشته باشند تا لایه پایه بتواند به تنهایی واگد شود. از طرفی اگر وابستگی بین داده‌های لایه پایه بالا باشد، کیفیت دریافتی داده‌های لایه پایه پس از واگد شدن به تنهایی نیز بالا خواهد بود. به عنوان مثالی دیگر، وقتی که گیرنده‌ها نیاز به قابلیت پشتیبانی نمای آزاد دارند، باید داده‌های لایه پایه و افزایشی وابستگی بیشتری به هم داشته باشند. چون در این حالت در صورت گم شدن لایه‌های افزایشی، داده‌های آنها با کیفیت بالا از روی داده‌های لایه پایه که حتماً به مقصد رسیده‌اند، قابل بازتولید خواهد بود. در حالتی که گیرنده‌ها محدودیت واگد کردن با تأخیر کم دارند نیز باید وابستگی داده‌های لایه پایه و افزایشی کم باشد. به این ترتیب لایه‌های افزایشی که به مقصد رسیده‌اند، می‌توانند مستقل از لایه پایه و با کیفیت بالا و در زمان محدود واگد شوند.

بنابراین، می‌توان گفت که با استفاده از مفهوم وابستگی بین لایه‌ای و درون لایه‌ای برای تخصیص داده‌ها به لایه‌های پایه و افزایشی می‌توان کیفیت تجربی مورد نیاز گیرنده‌ها را بر اساس محدودیت آنها تأمین کرد. استفاده از این مفاهیم ایده اصلی روش‌شناسی ارائه شده در این رساله برای تخصیص مناسب داده‌ها به لایه‌های مختلف است.

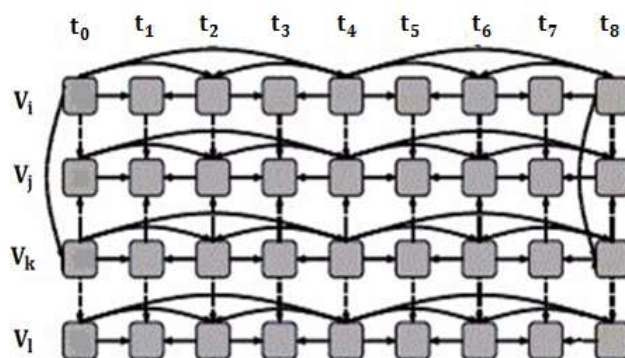
^۱ Low delay decoding

بنابراین باید مفاهیم وابستگی‌های بین‌لایه‌ای و درون‌لایه‌ای تعریف شوند. این مفاهیم بر اساس مقیاس‌پذیری انتخاب شده در قسمت قبلی چارچوب پیشنهادی تعریف می‌شوند. به-عنوان مثال در مقیاس‌پذیری نمایی وابستگی بین‌لایه‌ای به صورت میانگین وابستگی بین نماهای مختلف لایه پایه و افزایشی تعریف می‌شود. وابستگی درون‌لایه‌ای نیز میزان وابستگی بین نماهای مرجع و نمای پیش‌بینی شده بر اساس ساختار پیش‌بینی استفاده شده در هر لایه است. روابط (۱-۴) و (۲-۴) به ترتیب تعاریف وابستگی بین‌لایه‌ای و درون‌لایه‌ای را نشان می‌دهند:

$$\text{average inter_layer disparity} = (\sum_{i=1}^n \sum_{k=1}^m d_{ik}) / (n \times m) \quad (1-4)$$

$$\text{average intra_layer disparity} = (\sum d_{ik}) / (\text{number of predictions}) \quad (2-4)$$

در این روابط، n تعداد نماهای لایه پایه، m تعداد نماهای لایه افزایشی و d_{ik} اختلاف بین-نمایی برای نماهای i از لایه پایه و k از لایه افزایشی و نمای i نمای مرجع برای نمای j است. برای مثال، در حالتی که لایه مورد نظر ۴ نما داشته باشد و ساختار پیش‌بینی به صورت شکل ۱-۴ باشد، مقدار میانگین اختلاف درون‌لایه‌ای به صورت $\frac{d_{02}+d_{01}+d_{12}+d_{23}}{4}$ خواهد بود.



شکل ۱-۴ - مثالی از ساختار پیش‌بینی استفاده شده برای هر لایه

مراحل مختلف روش‌شناسی ارائه شده در شکل ۲-۴ نشان داده شده است.



شکل ۲-۴ - مراحل مختلف روش‌شناسی ارائه شده برای تخصیص داده‌ها به لایه‌های پایه و افزایشی

در ادامه این بخش، روش‌شناسی ارائه شده را به دو کاربرد خاص ویدئوهای چندنمایی و با محدودیت‌های متفاوت اعمال می‌کنیم. به منظور مشخص شدن جزئیات این روش، روش-شناسی ارائه شده، به صورت خاص برای مقیاس‌پذیری نمایی، به عنوان یکی از رایج‌ترین تعاریف مقیاس‌پذیری در ویدئوی چندنمایی، بررسی می‌شود. البته روش مورد نظر می‌تواند به سایر تعاریف ارائه شده برای مقیاس‌پذیری که در بخش قبل توضیح داده شد، نیز تعمیم داده شود.

۲-۴ - اعمال روش‌شناسی پیشنهادی برای کاربردهای موبایل با محدودیت

پهنای باند و توان پردازشی

در کاربردهای موبایل و در حالتی که از مقیاس‌پذیری نمایی استفاده شود، نمایی که برای لایه پایه انتخاب می‌شوند، روی کارایی فشرده‌سازی اثر مهمی دارند، چون در این محیط‌ها، احتمال اینکه لایه‌های افزایشی به‌طور کامل حذف شوند، زیاد است. در این حالت، به دلیل وابستگی ذاتی نماها به یکدیگر در رشته‌بیت ویدئوی چندنمایی (پیش‌بینی شدن نماها از روی یکدیگر به منظور افزایش کارایی فشرده‌سازی)، هیچ چیز در سمت کدگشا قابل واگد کردن نخواهد بود. اکثر روش‌های موجود [19][97] پیشنهاد می‌کنند که نماهای از دست‌رفته لایه افزایشی در سمت گیرنده بازتولید شده و برای واگد شدن نماهای لایه پایه مورد استفاده قرار گیرند. اما بازتولید نماها در سمت گیرنده نیاز به اطلاعات جانبی دارد که برای ارسال آنها نیاز به پهنای باند اضافی داریم. علاوه بر این، الگوریتم‌های بازتولید نماها توان

پردازشی بالایی را به گیرنده تحمیل می‌کنند که در محیط‌ها و دستگاه‌های موبایل قادر به تأمین آن نخواهیم بود.

به‌منظور اعمال روش‌شناسی پیشنهادی برای کاربردهای موبایل و در حالتی که از مقیاس-پذیری نمایی استفاده می‌شود، به‌صورت زیر عمل می‌کنیم. ابتدا، تعداد نماهای لایه پایه را بر اساس حداقل محدودیت گیرنده‌ها از لحاظ پهنای باند تعیین می‌کنم.

سپس محدودیت‌های کاربرد موردنظر برای تعیین کیفیت تجربی مورد نیاز برای کاربر را استخراج می‌کنیم. در کاربردهای موبایل این محدودیت‌ها شامل محدودیت پهنای باند و توان پردازش گیرنده‌ها است. بنابراین، احتمال به مقصد نرسیدن نماهای لایه‌های افزایشی زیاد است و گیرنده‌ها باید به لایه پایه بسنده کنند. در این حالت به‌منظور حداکثر کردن میزان کیفیت تجربی باید نماهای لایه پایه با کیفیت بالایی قابل واکنش باشند.

در گام بعد، باید بر اساس نوع مقیاس‌پذیری مورد استفاده، وابستگی بین‌لایه‌ای و درون-لایه‌ای را تعریف کنیم. همان‌طور که پیش‌تر اشاره شد، در مقیاس‌پذیری نمایی وابستگی بین‌لایه‌ای به‌صورت میانگین وابستگی بین نماهای لایه‌های پایه و افزایشی تعریف می‌شود. وابستگی درون‌لایه‌ای نیز وابستگی بین نماهای مرجع و پیش‌بینی‌شده در ساختار پیش-بینی لایه مورد نظر است.

بر اساس محدودیت‌های اشاره‌شده برای کاربردهای موبایل، روش‌شناسی ارائه شده در این رساله پیشنهاد می‌دهد که نماهای لایه پایه با یکدیگر وابستگی زیاد و با نماهای لایه افزایشی وابستگی کمی داشته باشند. در این حالت می‌توان ساختار پیش‌بینی لایه پایه را محدود کرد، به نحوی که نماهای این لایه تنها از روی یکدیگر و مستقل از نماهای لایه افزایشی کد و واکنش شوند. چون این نماها به گونه‌ای انتخاب شده‌اند که وابستگی کمی به نماهای لایه افزایشی دارند، محدود کردن ساختار پیش‌بینی کارایی فشرده‌سازی و کیفیت نهایی نماهای این لایه را تحت تأثیر قرار نخواهد داد. از سوی دیگر، اگر در شرایط بحرانی و بر اثر محدودیت پهنای باند، هیچ‌یک از لایه‌های افزایشی نیز به گیرنده‌ها نرسند، نماهای لایه پایه می‌توانند به‌طور مستقل واکنش شوند و نیازی به بازتولید نماهای لایه افزایشی نخواهد بود. بنابراین به اطلاعات جانبی مانند اطلاعات عمق یا اطلاعات هندسی نماها که معمولاً برای بازتولید نماها استفاده می‌شود، نیز نیازی نخواهیم داشت.

برای وضوح بیشتر، فرض کنیم که ۴ نما را می‌توانیم به‌عنوان لایه پایه ارسال کنیم. بنابراین، ساختار پیش‌بینی برای لایه پایه همان‌طور که در شکل ۴-۱ نشان داده شده است، باید به ۴ نما محدود شود.

در کد کردن ویدئوی چندنمایی، فریم‌های نمای اول تنها به صورت زمانی و از روی فریم‌های همان نما کد می‌شوند. بنابراین، نمایی که وابستگی بین فریم‌های آن از همه بیشتر باشد (اختلاف بین فریم‌های مختلف آن کم باشد)، باید به عنوان نمای اول (V_0 در شکل ۴-۱) انتخاب شود. بر اساس شکل ۴-۱، به منظور بالابردن وابستگی درون‌لایه‌ای و پیش‌بینی بهتر نماها از روی یکدیگر، V_2 باید حداکثر وابستگی را به V_0 داشته باشد. بنابراین، از بین همه نماهای موجود، نمایی که کمترین اختلاف بین‌نمایی را با نمای V_0 دارد، برای نمای ۲ انتخاب می‌شود. به همین طریق، نمایی با کمترین اختلاف بین‌نمایی با V_0 و V_2 به عنوان نمای V_1 و نمایی با حداقل اختلاف بین‌نمایی با V_2 به عنوان V_3 انتخاب می‌شود. به این ترتیب، لایه پایه می‌تواند به صورت مستقل از لایه افزایشی و با کارایی بالا کد و واکنش شود. از طرف دیگر، برای لایه‌های افزایشی ساختار پیش‌بینی محدود نشده که تمام نماها را شامل می‌شود، استفاده می‌شود تا کارایی فشرده‌سازی آن لایه‌ها بالا رود.

۴-۲-۱- ارزیابی روش‌شناسی ارائه شده برای اختصاص داده‌ها به لایه‌های

پایه و افزایشی برای کاربردهای موبایل

روش‌شناسی ارائه شده برای کاربردهای موبایل با پنج ویدئوی چندنمایی Break-Ballet, Ballroom, Exit, dancer و Dog [98][99][100] با تعداد نماها و وضوح تصویر متفاوت آزمایش شده است. مشخصات ویدئوهای آزمایش شده در جدول ۴-۱ نشان داده شده است. نکته قابل ذکر اینکه در همه ویدئوهای آزمایش شده در این جدول، دوربین‌ها ثابت هستند. نتایج با استفاده از کدکننده JMVC نسخه ۸/۵ استخراج شده است [105].

جدول ۴-۱- مشخصات ویدئوهای آزمایش شده

تعداد فریم	تعداد نما	نرخ فریم (فریم بر ثانیه)	اندازه فریم	ویدئو
۱۰۰	۸	۱۵	۱۰۲۴×۷۶۸	Ballet
۱۰۰	۸	۱۵	۱۰۲۴×۷۶۸	Break-dancer
۲۵۰	۸	۲۵	۶۴۰×۴۸۰	Exit
۲۵۰	۸	۲۵	۶۴۰×۴۸۰	Ballroom
۳۰۰	۸	۱۵	۱۲۸۰×۹۶۰	Dog

فرآیند پیاده‌سازی روش ارزیابی به این صورت است. ابتدا، فرض می‌کنیم که بر اساس حداقل محدودیت‌های گیرنده‌ها، چهار نما قابل دریافت در لایه پایه باشد. همان‌طور که در توضیح روش‌شناسی پیشنهادی برای کاربردهای موبایل اشاره شد، در این حالت باید

وابستگی بین‌نمایی لایه‌های پایه و افزایشی کم و وابستگی نماهای لایه پایه زیاد باشد و نماهای لایه پایه باید به صورت مستقل و تنها از روی یکدیگر پیش‌بینی شوند. بنابراین ساختار پیش‌بینی شکل ۴-۱ برای پیش‌بینی نماهای لایه پایه استفاده می‌شود. برای انتخاب نماهای مناسب برای لایه پایه باید متوسط اختلاف درون‌لایه‌ای و بین‌لایه‌ای را برای هر ویدئو را استخراج می‌کنیم. برای محاسبه این مقادیر باید اختلاف بین نماهای مختلف را به دست آوریم. به این منظور از روش ارائه شده در مقاله [107] استفاده شده است، این روش از اطلاعات عمق و هندسه ویدئوی چندنمایی به منظور کوچک کردن ناحیه جستجو و سرعت بخشیدن به محاسبه اختلاف بین نماها استفاده می‌کند. سپس مقادیر استخراج شده برای محاسبه میانگین اختلاف درون‌لایه‌ای و بین‌لایه‌ای استفاده می‌شوند. برای انتخاب نمای اول در ساختار پیش‌بینی لایه پایه نیز میزان اختلاف بین فریم‌های مختلف نمای اول را محاسبه می‌کنیم. مقادیر اختلاف بین فریم‌های نماهای مختلف و اختلاف‌های بین‌نمایی برای نماهای مختلف در جدول ۴-۲ و جدول ۴-۳ نشان داده شده است.

نحوه انتخاب نماهای مناسب برای لایه پایه بر اساس مقادیر اختلاف درون‌نمایی و بین‌نمایی مربوط به مقیاس‌پذیری نمایی به این صورت است. برای مثال برای ویدئوی Exit، با توجه به مقادیر نشان داده شده در جدول ۴-۳، نمای ۰ به عنوان اولین نمای لایه پایه در ساختار پیش‌بینی شکل ۴-۱ در نظر گرفته می‌شود، چون کمترین مقدار اختلاف درون‌نمایی را از بین ۸ نمای موجود دارد و می‌تواند با کارایی بالا به صورت زمانی کد و واکد شود. همان‌طور که در جدول ۴-۲ دیده می‌شود، از بین نماهای باقی‌مانده از این ویدئو، نمای ۶ کمترین اختلاف بین‌نمایی را با نمای ۰ دارد. بنابراین، این نما باید به عنوان V_2 در ساختار پیش‌بینی شکل ۴-۱ مورد استفاده قرار گیرد. نمای ۱ کمترین اختلاف بین‌نمایی را با نمای ۰ و ۶ دارد، بنابراین به عنوان V_1 انتخاب می‌شود و سرانجام نمای ۳ که کمترین اختلاف را با نمای ۶ دارد، به عنوان V_3 انتخاب می‌شود.

به این ترتیب، در حالتی که تنها نماهای لایه پایه به مقصد برسند، این نماها می‌توانند با کارایی بالایی واکد شده و کیفیت دریافتی قابل قبولی نیز برای کاربر فراهم کنند، بدون اینکه توان پردازشی و پهنای باند اضافی به گیرنده‌ها تحمیل کنند.

جدول ۴-۲ - متوسط اختلاف بین‌نمایی بین نماهای مختلف ویدئوهای آزمایش‌شده

ویدئو		نمای ۰	نمای ۱	نمای ۲	نمای ۳	نمای ۴	نمای ۵	نمای ۶
Exit	نمای ۰							
	نمای ۱	۰/۸						
	نمای ۲	۰/۸۸	۰/۹۰					
	نمای ۳	۰/۸۵	۰/۸۵	۰/۸۷				
	نمای ۴	۰/۸۱	۰/۸۳	۰/۸۶	۰/۹۰			
	نمای ۵	۰/۸۲	۰/۸۵	۰/۸۶	۰/۸۶	۰/۸۹		
	نمای ۶	۰/۷۷	۰/۷۹	۰/۸۱	۰/۸۰	۰/۸۱	۰/۸۵	
	نمای ۷	۰/۷۹	۰/۸۱	۰/۸۲	۰/۸۱	۰/۸۵	۰/۸۶	۰/۸۵
Ballet	نمای ۰							
	نمای ۱	۰/۱۷						
	نمای ۲	۰/۱۵	۰/۱۶					
	نمای ۳	۰/۱۲	۰/۱۳	۰/۲				
	نمای ۴	۰/۰۹	۰/۰۸	۰/۰۹	۰/۱۱			
	نمای ۵	۰/۱۱	۰/۱	۰/۱۴	۰/۱	۰/۲۳		
	نمای ۶	۰/۱۴	۰/۱۵	۰/۱۸	۰/۱۷	۰/۲۳	۰/۲۳	
	نمای ۷	۰/۱۲	۰/۱۱	۰/۱۵	۰/۱۳	۰/۱۷	۰/۱۷	۰/۲
Break-dancer	نمای ۰							
	نمای ۱	۰/۰۶						
	نمای ۲	۰/۰۸	۰/۰۸					
	نمای ۳	۰/۰۸	۰/۰۸	۰/۰۹				
	نمای ۴	۰/۱۱	۰/۱	۰/۰۹	۰/۰۸۲			
	نمای ۵	۰/۰۸	۰/۰۸	۰/۰۹۳	۰/۱۲	۰/۱۲		
	نمای ۶	۰/۰۷	۰/۰۷۳	۰/۰۹۱	۰/۱	۰/۱۲	۰/۱۴	
	نمای ۷	۰/۰۷	۰/۰۷	۰/۰۸۴	۰/۱	۰/۱۱	۰/۱۳	۰/۱
Ballroom	نمای ۰							
	نمای ۱	۰/۰۸۶۶						
	نمای ۲	۰/۰۶۹۷	۰/۰۴۷۲					
	نمای ۳	۰/۰۶۱۶	۰/۰۵۹۱	۰/۰۴۳۵				
	نمای ۴	۰/۰۶۹۰	۰/۰۵۶۰	۰/۰۳۸۱	۰/۰۲۷۹			
	نمای ۵	۰/۰۷۲۵	۰/۰۷۹۸	۰/۰۶۵۵	۰/۰۴۶۰	۰/۰۴۴۵		
	نمای ۶	۰/۱۰۲۰	۰/۰۶۱۱	۰/۰۶۸۱	۰/۰۷۶۱	۰/۰۶۵۹	۰/۰۸۱۰	
	نمای ۷	۰/۰۸۳۰	۰/۰۶۷۹	۰/۰۵۷۱	۰/۰۴۴۱	۰/۰۳۶۲	۰/۰۴۴۰	۰/۰۶۶۱
Dog	نمای ۰							
	نمای ۱	۰/۱۵۳۳						
	نمای ۲	۰/۱۶۶۴	۰/۰۰۱۹					
	نمای ۳	۰/۱۳۶۸	۰/۰۰۱۱	۰/۲۱۱۵				
	نمای ۴	۰/۱۶۱۴	۰/۰۰۲۰	۰/۰۸۷۹	۰/۲۳۰۴			
	نمای ۵	۰/۱۶۶۲	۰/۰۰۱۳	۰/۱۳۵۶	۰/۱۶۲۹	۰/۱۴۵۲		
	نمای ۶	۰/۱۹۰۱	۰/۰۰۱۵	۰/۲۳۱۲	۰/۰۹۰۷	۰/۲۵۹۰	۰/۱۹۵۷	
	نمای ۷	۰/۱۶۰۵	۰/۰۰۱۳	۰/۰۹۲۶	۰/۱۶۲۱	۰/۱۰۹۴	۰/۰۹۴۴	۰/۱۷۹۵

جدول ۴-۳ - متوسط اختلاف درون‌نمایی برای نماهای مختلف از ویدئوهای آزمایش شده

ویدئو	نمای ۷	نمای ۶	نمای ۵	نمای ۴	نمای ۳	نمای ۲	نمای ۱	نمای ۰
Exit	۰/۹۷۸۵	۰/۹۷۹۸	۰/۹۸۰۲	۰/۹۷۹۸	۰/۹۷۹۹	۰/۹۸	۰/۹۸۰۲	۰/۹۷
Ballet	۰/۹۵۴	۰/۹۵۵۴	۰/۹۵۶۱	۰/۹۵۳۸	۰/۹۵۵۴	۰/۹۵۵۳	۰/۹۵۴۶	۰/۹۵۴۸
Break-dancer	۰/۹۳۱۹	۰/۹۳۲۵	۰/۹۳۴۴	۰/۹۳۳۴	۰/۹۳۲۳	۰/۹۳۴۲	۰/۹۳	۰/۹۲
Ballroom	۰/۰۱۴۱	۰/۰۱۵۶	۰/۰۱۴۲	۰/۰۱۴۷	۰/۰۱۵۰	۰/۰۱۴۱	۰/۰۱۴۵	۰/۰۱۵۲
Dog	۰/۰۲۳۹	۰/۰۲۳۶	۰/۰۲۳۷	۰/۰۲۳۸	۰/۰۲۳۷	۰/۰۲۴۳	۰/۰۲۳۶	۰/۰۲۴۳

برای نشان دادن کارایی فشرده‌سازی روش انتخاب نمای ارائه شده، کیفیت نهایی را با حالتی که ۴ نمای مجاور برای لایه پایه انتخاب شوند، مقایسه می‌کنیم. همچنین روش ارائه شده را برای تعداد متفاوتی از نماهای لایه پایه آزمایش کرده و با حالتی که همان تعداد نمای مجاور انتخاب شوند، مقایسه می‌کنیم. PSNR و نرخ بیت استخراج شده از روش ارائه شده و حالتی که نماهای مجاور انتخاب شوند برای ویدئوی Exit در جدول ۴-۴ نشان داده شده است. همان‌طور که می‌بینیم، نماهای انتخاب شده از روش ارائه شده در نرخ بیت پایین‌تر، کیفیت بهتری دارند.

ضمناً، ارزیابی‌ها با ۴ مقدار متفاوت پارامتر کوانتیزاسیون (۱۵، ۲۰، ۲۵ و ۳۰) برای همه ویدئوهای آزمایش شده نیز انجام شده است. برای هر آزمایش، کیفیت با معیار میانگین Bjøntegaard-Delta bitrate [101] اندازه‌گیری شده، BD-PSNR و BD-Bitrate برای هر ویدئو استخراج و میزان بهبود کارایی نسبت به حالتی که نماهای مجاور انتخاب شوند، در جدول ۴-۵ نشان داده شده است. اعداد منفی در این جدول نشان‌دهنده بهبود کارایی کدکردن است.

جدول ۴-۴ - مقایسه کیفیت و نرخ بیت (کیلو بیت بر ثانیه) میانگین نماهای انتخابی روش ارائه شده و نماهای مجاور

برای ویدئوی Exit

QP	دو نما				سه نما				چهار نما			
	نماهای انتخاب شده با روش پیشنهادی		نماهای مجاور		نماهای انتخاب شده با روش پیشنهادی		نماهای مجاور		نماهای انتخاب شده با روش پیشنهادی		نماهای مجاور	
	PSNR	نرخ بیت	PSNR	نرخ بیت	PSNR	نرخ بیت	PSNR	نرخ بیت	PSNR	نرخ بیت	PSNR	نرخ بیت
۳۰	۳۶/۹۶	۶۷۴	۳۶/۶۸	۷۱۹	۳۷/۱۳	۸۶۵	۳۶/۷۸	۹۸۱	۳۷/۲۶	۱۱۳۹	۳۶/۶۳	۱۳۶۵
۲۵	۳۸/۵۸	۱۳۸۱	۳۸/۴۵	۱۴۸۱	۳۸/۷۴	۱۸۰۷	۳۸/۵۴	۲۰۸۴	۳۸/۸۹	۲۳۳۸	۳۸/۴۶	۲۹۲۸
۲۰	۴۰/۲۹	۳۳۱۳	۴۰/۲۳	۴۳۹۷	۴۰/۳۶	۴۴۹۷	۴۰/۳۱	۵۸۴۴	۴۰/۴۷	۵۷۱۱	۴۰/۳۰	۷۷۰۱
۱۵	۴۳/۰۵	۹۱۶۰	۴۳/۰۳	۹۴۵۵	۴۳/۰۴	۱۳۰۳۰	۴۳/۰۵	۱۳۶۸۱	۴۳/۰۸	۱۶۵۹۲	۴۳/۰۴	۱۸۴۰۶

جدول ۴-۵- میزان کارایی روش ارائه‌شده از لحاظ BD-PSNR و BD-Bitrate در مقایسه با حالتی که نماهای مجاور انتخاب شوند. (اعداد منفی بهبود کارایی را نشان می‌دهند).

چهار نما		سه نما		دو نما		ویدئو
BD-Bitrate	BD-PSNR	BD-Bitrate	BD-PSNR	BD-Bitrate	BD-PSNR	
-۰/۲۰۷۰	-۰/۰۱۴۵	-۰/۱۸۷۸	-۰/۱۰۲۹	-۲/۴۰۳۵	-۰/۲۵۵	Ballet
-۰/۸۰۲۰	-۰/۲۳۹۱	-۰/۹۶۳۹	-۰/۲۶۳	۰/۰۳۲۶	-۰/۰۲۶۱	Break-dancer
-۱/۵۳۰۳	-۰/۸۲۱۶	-۱/۰۳۸۸	-۰/۳۸۰۱	-۰/۹۶۰۵	-۰/۴۵۰۶	Exit
-۰/۱۰۱۵	۰/۰۵۵	-۰/۲۵۸	-۰/۵۸	-۰/۵۵۲۲	-۰/۳۰	Ballroom
۰/۷	۱/۲۷	-۰/۱۴۱۸	-۰/۶۶	-۰/۰۲	۰/۲۷	Dog
-۰/۳۹	-۰/۰۵	-۰/۵۲	-۰/۴	-۰/۷۴	-۰/۱۵۲۳	میانگین

همان‌طور که در این جدول دیده می‌شود برای حالت‌های مختلف انتخاب نماها (دو نما، سه نما و چهار نما) کارایی فشرده‌سازی روش پیشنهادی از حالتی که نماهای مجاور انتخاب شوند، بهتر است.

در مواردی که کارایی فشرده‌سازی عدد مثبت را نشان می‌دهد، مانند حالت ۴ نمای انتخابی برای ویدئوی Dog، به این معنی است که نسبت به حالتی که نماهای مجاور انتخاب شوند کارایی فشرده‌سازی نداریم، به این دلیل که در این حالت میانگین اختلاف بین‌نمایی برای نماهایی که برای حالت پیشنهادی انتخاب شده‌اند، بسیار شبیه حالتی است که نماهای مجاور انتخاب شوند. در این ویدئوی آزمایش‌شده، دوربین‌ها بسیار نزدیک به هم قرار دارند، بنابراین اختلاف بین نماهای مختلف چندان زیاد نخواهد بود. قطعاً در حالتی که فاصله بین دوربین‌ها بیشتر و اختلاف بین نماهای مجاور محسوس‌تر باشد، میزان بهبود روش پیشنهادی از لحاظ کارایی فشرده‌سازی نسبت به حالتی که نماهای مجاور انتخاب شوند، نیز محسوس‌تر خواهد بود.

نکته قابل توجه این‌که اگرچه میزان بهبود کارایی فشرده‌سازی روش ارائه‌شده نسبت به حالت مرجع چندان بالا نیست، اما مزیت اصلی این روش این است که محدودیت‌های کاربرد موردنظر را به نحو مناسب تأمین می‌کند. به عبارت دیگر، روش پیشنهادی با در نظر گرفتن ویژگی‌های کاربردهای مختلف ویدئوی چندنمایی می‌تواند به کاربردهای مختلف ویدئوی چندنمایی اعمال شده و محدودیت‌های آنها در کیفیت تجربی مناسب تأمین کند.

۳-۴ - اعمال روش‌شناسی پیشنهادی برای کاربردهای نمای آزاد با قابلیت

کنترل تعاملی محدوده دید توسط کاربر

ویدئوی نمای آزاد به کاربر این اجازه را می‌دهد که صحنه مورد نظر را آزادانه و از جهات مختلف تماشا کند. لازمه این کار ارسال داده‌های نماهای مختلف برای کاربر است که نیاز به پهنای باند بالایی دارد. همان‌طور که پیش‌تر اشاره شد، در مقیاس‌پذیری نمایی، در صورتی که لایه افزایشی به مقصد نرسد، شرایط قابلیت پیمایش صحنه مورد نظر از نماهای مختلف برای کاربر دچار مشکل می‌شود. یک راه حل کارآمد در شرایطی که گیرنده‌ها قابلیت پردازش کافی داشته باشند، استفاده از بازتولید نماها است. در این روش، نماهایی که به دلیل محدودیت پهنای باند یا خطای کانال به مقصد نرسیده‌اند، در سمت گیرنده و با استفاده از داده‌های نماهایی که به مقصد رسیده‌اند، بازتولید می‌شوند.

بر اساس این بحث و محدودیت‌های کاربردهای نمای آزاد، روش‌شناسی ارائه‌شده در این رساله پیشنهاد می‌دهد که نماهای لایه پایه و افزایشی وابستگی زیادی به یکدیگر داشته باشند. در این حالت، اگر نماهای لایه افزایشی به مقصد نرسند، با استفاده از داده‌های لایه پایه می‌توان نماها از دست رفته را با کیفیت بالایی بازتولید کرد. به این ترتیب، نیازهای گیرنده‌ها از جهت پیمایش صحنه تصویربرداری شده، تأمین شده و هزینه اضافی از جهت ارسال اطلاعات جانبی برای بازتولید نماها به شبکه تحمیل نمی‌شود. وابستگی بالای بین نماهای لایه پایه و افزایشی نیز کیفیت دریافتی نماهای بازتولید شده را افزایش می‌دهد.

به‌منظور اعمال این روش به کاربردهای نمای آزاد، در گام نخست تعداد نماهای لایه پایه را بر اساس حداقل محدودیت گیرنده‌ها از لحاظ پهنای باند تعیین می‌کنیم. فرض کنیم که بر اساس این محدودیت‌ها تنها ۴ نما در لایه پایه قابل ارسال باشد. ساختار پیش‌بینی نماهای لایه پایه مانند شکل ۴-۱ محدود می‌شود تا نماهای لایه پایه بدون استفاده از نماهای لایه افزایشی قابل واگرد کردن باشند. بنابر بحث بالا و به منظور برطرف کردن نیاز گیرنده‌ها از لحاظ پیمایش صحنه تصویربرداری شده از جهات مختلف، نماهای لایه پایه و افزایشی به گونه‌ای انتخاب می‌شوند که کمترین میانگین اختلاف بین لایه‌ای را داشته باشند. به این منظور، ترکیبات مختلف برای اختصاص نماها به لایه پایه و افزایشی را در نظر می‌گیریم و از بین آنها حالتی که کمترین میانگین اختلاف بین لایه‌ای برای نماهای لایه‌های پایه و افزایشی را دارند را انتخاب می‌کنیم. میانگین اختلاف بین لایه‌ای برای نماهای لایه پایه و افزایشی از رابطه (۴-۱) به دست می‌آید. همانند بخش قبلی، اختلاف بین‌نمایی از روشی که در [107] ارائه شده است، استخراج شده است.

۴-۳-۱- ارزیابی روش‌شناسی ارائه شده برای اختصاص داده‌ها با لایه‌های

پایه و افزایشی برای کاربردهای ویدئوی نمای آزاد

روش‌شناسی ارائه شده برای کاربردهای ویدئوی نمای آزاد با ویدئوهای چندنمایی Ballet، Break-dancer و Balloons آزمایش شده است. مشخصات ویدئوهای آزمایش شده در جدول ۴-۶ نشان داده شده است. نکته قابل ذکر اینکه در همه ویدئوهای آزمایش شده در این جدول، دوربین‌ها ثابت هستند.

جدول ۴-۶ - مشخصات ویدئوهای آزمایش شده

تعداد فریم	تعداد نما	نرخ فریم (فریم بر ثانیه)	اندازه فریم	ویدئو
۱۰۰	۸	۱۵	۱۰۲۴×۷۶۸	Ballet
۱۰۰	۸	۱۵	۱۰۲۴×۷۶۸	Break-dancer
۲۴۰	۷	۱۵	۱۰۲۴×۷۶۸	Balloons

فرآیند پیاده‌سازی روش ارزیابی به این صورت است. ابتدا، متوسط اختلاف بین نماهای مختلف را برای ویدئوهای آزمایش شده استخراج می‌کنیم. این اعداد که برای استخراج اختلاف بین لایه‌ای استفاده می‌شوند، در جدول ۴-۲ و جدول ۴-۳ نشان داده شده است. فرض می‌کنیم که یک لایه پایه و یک لایه افزایشی هر کدام با ۴ نما داریم. حالت‌های مختلف اختصاص نما برای لایه‌های پایه و افزایشی و میانگین اختلاف بین لایه‌ای دو لایه را در حالت‌های مختلف استخراج کرده و از بین آنها حالتی که میانگین اختلاف بین لایه‌ای حداقل باشد را انتخاب می‌کنیم. نماهای اختصاص داده شده برای لایه پایه و افزایشی و میانگین اختلاف بین لایه‌ای برای هر یک از ویدئوهای آزمایش شده در جدول ۴-۷ نشان داده شده است.

جدول ۴-۷ - نماهای اختصاص داده شده به لایه پایه و افزایشی و میانگین اختلاف بین لایه‌ای برای ویدئوهای

آزمایش شده

ویدئو	نماهای لایه پایه	نماهای لایه افزایشی	میانگین اختلاف بین لایه‌ای
Ballet	۳-۰-۵-۶	۷-۴-۲-۱	۰/۸۳
Break-dancer	۲-۳-۴-۷	۶-۵-۱-۰	۰/۵۲
Balloons	۳-۰-۴-۵	۶-۲-۱	۰/۱۴

حال فرض می‌کنیم که لایه افزایشی به دلیل محدودیت پهنای باند به مقصد نرسیده و با استفاده از داده‌های لایه پایه نماهای لایه افزایشی را بازتولید می‌کنیم. برای بازتولید نماها

از نرم‌افزار مرجع گروه MPEG، با نام View Synthesis Reference Software (VSRS) نسخه ۳ [102] استفاده کرده‌ایم.

به‌منظور نشان دادن کارایی فشرده‌سازی روش ارائه‌شده، کیفیت نماهای بازتولید شده را در روش پیشنهادی با حالتی که نماهای مجاور برای لایه پایه انتخاب شوند، مقایسه می‌کنیم. جدول ۴-۸ BD-bitrate را برای این دو حالت و برای ویدئوهای آزمایش‌شده نشان می‌دهد. از آنجایی که در هر دو حالت فرض شده است که تنها لایه پایه به گیرنده‌ها رسیده است و لایه افزایشی به‌طور کامل حذف شده است، نرخ بیت نهایی تنها نرخ بیت لایه پایه خواهد بود. همان‌طور که مشاهده می‌کنیم، کارایی فشرده‌سازی روش پیشنهادی نسبت به حالتی که نماهای مجاور انتخاب شوند، بهبود یافته است.

جدول ۴-۸- میزان کارایی روش ارائه‌شده از لحاظ BD-Bitrate در مقایسه با حالتی که نماهای مجاور انتخاب شوند. (اعداد منفی بهبود کارایی را نشان می‌دهند.)

BD-bitrate	ویدئو
-۰/۷	Ballet
-۰/۱۴	Break-dancer
-۱/۴	Balloons

همان‌طور که در جدول ۴-۷ نشان می‌دهد، برای ویدئوی Ballet در روش پیشنهادی، نماهای ۰، ۳، ۵ و ۶ برای لایه پایه انتخاب شده و نماهای ۱، ۲، ۴ و ۷ با استفاده از نماهای لایه پایه و به‌عنوان لایه افزایشی بازتولید شده‌اند. در حالتی که نماهای مجاور برای لایه پایه در نظر گرفته شده‌اند، نماهای ۰، ۱، ۲ و ۳ برای لایه پایه انتخاب شده و نماهای ۴، ۵، ۶ و ۷ با استفاده از نماهای لایه پایه و به‌عنوان لایه افزایشی بازتولید شده‌اند. شکل ۴-۳ (الف) و (ب) نمای بازتولید شده ۴ را در دو حالت نشان می‌دهد. از آنجایی که نمای ۴ در دو حالت با نماهای مرجع متفاوت بازتولید شده‌اند، زاویه دید در دو نمای بازتولیدی متفاوت است. اما همان‌طور که در این شکل می‌بینیم، کیفیت نمای بازتولید شده در روش پیشنهادی از حالت مرجع که نماهای مجاور انتخاب شده‌اند، از لحاظ بصری بهتر است. در نمای بازتولید شده توسط نماهای مجاور، اجزای موجود در صحنه نمایش داده شده دوبار تکرار شده‌اند.



(ب)

(الف)

شکل ۳-۴ - مقایسه کیفیت بصری نماهای باز تولید شده در (الف) روش پیشنهادی و (ب) روش مرجع که نماهای مجاور به‌عنوان لایه پایه انتخاب شده‌اند.

به این ترتیب روش‌شناسی ارائه شده تنها با استفاده از مفهوم اختلاف بین‌نمایی و با کم کردن وابستگی نماهای انتخاب شده برای لایه پایه و افزایشی می‌تواند رشته بیت ویدئویی مقیاس‌پذیر متناسب با محدودیت‌های کاربردهای نمای آزاد تولید کند. در این حالت بدون نیاز به ارسال اطلاعات جانبی و تنها با استفاده از داده‌های ارسال شده در لایه پایه، می‌توان نماهای از دست رفته لایه افزایشی را با کیفیت خوبی باز تولید کرد.

۴-۴- جمع‌بندی

در این بخش از رساله، یک روش‌شناسی برای اختصاص مناسب داده‌ها به لایه‌های پایه و افزایشی بر حسب کاربرد خاص ویدئوی چندنمایی مقیاس‌پذیر به‌نحوی که میزان کیفیت نهایی حداکثر شود، ارائه شد. روش‌شناسی ارائه شده از مفاهیم اختلاف بین‌نمایی و درون-نمایی به‌منظور اختصاص مناسب داده‌ها به لایه‌های پایه و افزایشی استفاده می‌کند. این مفاهیم بر حسب نوع مقیاس‌پذیری انتخاب‌شده در قسمت قبل چارچوب پیشنهادی تعریف می‌شوند. نتایج شبیه‌سازی‌ها نشان می‌دهد که روش‌شناسی ارائه شده، نیازهای خاص کاربردهای مورد نظر را تأمین کرده و کیفیت نهایی نیز حداکثر می‌شود.

فصل پنجم

روش‌شناسی ارائه‌شده برای استخراج مدل نرخ‌میت-

اعوجاج برای ویدئوی چندنمایی

در این فصل از رساله یک مدل نرخ بیت - اعوجاج در سطح نما برای ویدئوی چندنمایی سه-بعدی به صورت روش‌مند استخراج می‌شود. در این فصل ابتدا لزوم استخراج مدل نرخ بیت - اعوجاج برای ویدئوی چندنمایی به صورت روش‌مند و نیز نوآوری‌های روش ارائه‌شده در بخش ۵-۱- توضیح داده می‌شود. بخش ۵-۲- گام‌های روش‌شناسی پیشنهادی را معرفی می‌کند. در نهایت مدل نرخ بیت - اعوجاج استخراج‌شده از این روش‌شناسی و نتایج حاصل از ارزیابی آن به ترتیب در بخش‌های ۵-۳- و ۵-۴ ارائه می‌شوند.

۵-۱- لزوم ارائه مدل نرخ بیت - اعوجاج برای ویدئوی چندنمایی سه‌بعدی و نوآوری‌های روش ارائه‌شده

به منظور ارسال کارآمد ویدئوی چندنمایی و سه‌بعدی، الگوریتم‌های مناسب کنترل و اختصاص نرخ بیت برای کاربردهای خاص این نوع ویدئوها مورد نیاز است. برای طراحی این الگوریتم‌ها، خصوصیات ویدئوهای چندنمایی و سه‌بعدی و ویژگی‌های کاربردهای خاص آنها باید مدنظر قرار گیرد.

حجم زیاد افزونگی بین‌نمایی موجود در ویدئوهای چندنمایی سه‌بعدی می‌تواند در طراحی کارآمد الگوریتم‌های کنترل نرخ بیت در این نوع ویدئوها استفاده شود. علاوه بر این، ویژگی‌های خاص کاربردهای مختلف ویدئوهای چندنمایی سه‌بعدی می‌تواند در طراحی الگوریتم‌های کارآمد برای کنترل نرخ بیت مورد توجه قرار گیرد. تغییر آزادانه زاویه دید^۱ یکی از خصوصیات مهم کاربردهای ویدئوها چندنمایی و سه‌بعدی است که اثر قابل ملاحظه‌ای روی کیفیت تجربی دارد. این ویژگی به کاربر این اجازه را می‌دهد که از بین نماهای موجود و به صورت تعاملی، نمای موردنظر خود را انتخاب کند. نکته کلیدی در این مسأله، پوشش دادن تمام زوایای مورد نیاز برای کاربرهای مختلف با بهترین کیفیت ممکن است که به خصوص در کاربردهای موبایل با محدودیت پهنای باند و توان پردازشی چالش برانگیز خواهد بود. این محدودیت‌ها در دستگاه‌های موبایل ممکن است منجر به محدود کردن تعداد نماهای ارسالی برای کاربر شود. بنابراین به منظور ارتقا کیفیت تجربی نهایی، نماهای ارسال‌شده باید در سمت گیرنده بازتولید شوند. حتی اگر پهنای باند نیز کافی باشد، با مشکل محدودیت توان پردازشی و توان مصرفی دستگاه‌های موبایل مواجه هستیم، چون پیچیدگی الگوریتم‌های بازتولید نماها ایجاب می‌کند که به منظور بهبود کیفیت تجربی، توان پردازشی را بالا ببریم. بنابراین، نوعی مصالحه بین کیفیت نهایی، پهنای باند و

^۱ Free view-point switching

توان پردازشی وجود دارد. یک الگوریتم کنترل نرخ بیت مناسب می‌تواند این چالش را با تنظیم کردن کیفیت ویدئو حل کند به نحوی که نیازمندی‌های گیرنده‌ها، ظرفیت شبکه و توان پردازشی موجود به صورت توأم مورد توجه قرار گیرند.

رابطه بین نرخ بیت - اعوجاج در اکثر تکنیک‌های کد کردن ویدئو تحت عنوان مدل نرخ بیت - اعوجاج نشان داده می‌شود. عمده الگوریتم‌های کنترل نرخ بیت ارائه شده برای ویدئوی تک‌نمایی و چندنمایی سه‌بعدی، از مدل نرخ بیت - اعوجاج برای توصیف رابطه بین نرخ بیت و کیفیت ویدئوی گذشته استفاده می‌کنند. از آنجایی که دقت مدل نرخ بیت - اعوجاج بر روی کارایی الگوریتم کنترل نرخ بیت تأثیر به‌سزایی دارد، از جمله بخش‌های کلیدی الگوریتم به حساب می‌آید.

دو روش برای به دست آوردن مدل نرخ بیت - اعوجاج وجود دارد: روش آماری¹ و روش تجربی² [103]. روش‌های آماری فرض می‌کنند که سیگنال وردی از توزیع خاصی، مانند توزیع گاوسی تبعیت می‌کند. بیشتر مدل‌های نرخ بیت - اعوجاج موجود برای ویدئوی چندنمایی مدل‌های آماری هستند که مبنای نظری مشترکی داشته و تفاوت آنها تنها در روشی است که برای ساده کردن مدل یا کاربردی ساختن آن به کار می‌گیرند. اما مدل‌های آماری دقت بالایی ندارند، چون برای همه داده‌های پردازش شده یک مدل را استفاده می‌کنند. یافتن مدل آماری با دقت بالا بسیار دشوار است، چون مدل‌های دقیق تنها برای منابع ساده و تحت شرایط خاص قابل استفاده خواهند بود.

مدل‌های تجربی با در نظر گرفتن ویژگی‌های نرخ بیت - اعوجاج بر حسب انواع مختلف داده‌های ورودی می‌توانند مدل‌های دقیق‌تری ارائه دهند [103]. این مدل‌ها می‌توانند از طریق فرآیند برازش منحنی به صورت پویا به‌روز شده و دقت بالایی را فراهم کنند. اما مدل‌های تجربی ارائه شده برای ویدئوی چندنمایی مشکلاتی نیز دارند. دقت مدل نرخ بیت - اعوجاج تجربی علاوه بر بررسی گسترده داده‌های ورودی به شرایط کد کردن نیز بستگی دارد. اما مدل‌های موجود ویژگی‌های ویدئوی چندنمایی و نیز ساختار و پارامترهای کدکننده را به صراحت در نظر نمی‌گیرند. با در نظر گرفتن این موارد پارامترهای مدل نرخ بیت - اعوجاج می‌تواند به نحو مؤثرتر و کارآمدتر تعیین شود.

در ویدئوی چندنمایی، کیفیت تجربی پارامتر مهمی است که باید در انتخاب و تنظیم پارامترهای کدکننده در نظر گرفته شود. به عبارت دیگر، به منظور برقراری مصالحه مناسب

¹ Statistical method

² Experimental method

بین نرخ بیت و کیفیت دریافتی، پارامترهایی که روی کیفیت تجربی اثر دارند، باید به صورت مؤثر در مدل نرخ بیت - اعوجاج در نظر گرفته شوند.

در این رساله، یک روش‌شناسی مشخص به‌منظور استخراج مدل نرخ بیت - اعوجاج در سطح نما برای ویدئوی چندنمایی سه‌بعدی ارائه می‌شود. معمولاً الگوریتم‌های نرخ بیت - اعوجاج برای ویدئوی چندنمایی در سطوح متفاوت مانند سطح نما، سطح گروهی از تصاویر، سطح فریم و غیره طراحی می‌شوند. به این ترتیب، در هر سطح پارامترها مؤثر برای آن سطح می‌توانند برای تعیین نرخ بیت آن سطح به‌صورت دقیق استفاده شوند. علاوه بر این، این روش سلسله‌مراتبی می‌تواند حافظه موردنیاز و پیچیدگی محاسباتی الگوریتم کنترل نرخ بیت را مدیریت کند. از آنجایی که اختصاص نرخ بیت در سطح نما یکی از مهم‌ترین قسمت‌های الگوریتم اختصاص نرخ بیت در ویدئوی چندنمایی سه‌بعدی است و نیز اینکه ویژگی‌های خاص ویدئوی چندنمایی سه‌بعدی می‌تواند در این سطح به‌طور مؤثر بازتاب داده شود، روش‌شناسی این رساله مدل نرخ بیت - اعوجاج در سطح نما را مورد توجه قرار می‌دهد. البته روش‌شناسی استفاده شده در این رساله می‌تواند به‌سادگی به سطوح دیگر نیز تعمیم داده شده و مدل‌های نرخ بیت - اعوجاج مناسب در سطح گروهی از تصاویر و فریم با در نظر گرفتن پارامترهای مهم آنها تعیین شود.

نوآوری‌های روش ارائه‌شده در این رساله به‌صورت خلاصه شامل موارد زیر است:

- در این رساله یک رویه روش‌مند برای استخراج پارامترهای مدل نرخ بیت - اعوجاج در سطح نما با توجه به ویژگی‌های ویدئوی چندنمایی سه‌بعدی و کاربرد مورد نظر پیشنهاد می‌شود. در این رساله نشان خواهیم داد که انعکاس این ویژگی‌ها در مدل نرخ بیت - اعوجاج منجر به دستیابی به مدلی دقیق‌تر خواهد شد.
- روش‌شناسی پیشنهادی از وابستگی‌های موجود بین فریم‌های ویدئوی چندنمایی، به‌عنوان یکی از خصوصیات مهم ویدئوی چندنمایی سه‌بعدی، برای یافتن پارامترهای مدل نرخ بیت - اعوجاج استفاده می‌کند. اهمیت این موضوع در این است که وابستگی‌های بین نماها و میزان حرکت در فریم‌های مجاور می‌تواند روی فرآیند پیش‌بینی و در نتیجه روی مقدار نهایی نرخ بیت هر نما تأثیر به‌سزایی داشته باشد.
- روش پیشنهادی در این رساله از مفاهیم اختلاف بین‌نمایی و درون‌نمایی به‌منظور توصیف وابستگی‌های آماری در فرآیند کدکردن ویدئوی چندنمایی استفاده می‌کند. بنابراین پیشنهاد می‌شود که نرخ بیت هر نما به‌صورت تابعی از اختلاف بین‌نمایی و درون‌نمایی تعریف شود.

- از آنجایی که روش پیشنهادی برای استخراج مدل نرخ بیت - اعوجاج خصوصیات کاربرد مورد نظر را نیز در نظر می‌گیرد، مدل استخراج شده به راحتی می‌تواند برای محدوده وسیعی از کاربردهای ویدئوی چندنمایی سه‌بعدی تعمیم داده شود.
- اگرچه در این رساله تنها استاندارد H.264/MVC و کاربردهای آن مورد نظر قرار گرفته‌اند، اما روش پیشنهادی برای استخراج مدل نرخ بیت - اعوجاج می‌تواند برای سایر استانداردهای کدکردن ویدئو، مانند استاندارد HEVC و با در نظر گرفتن پارامترهای مرتبط با نرخ بیت و اعوجاج در آن استانداردها مورد استفاده قرار گیرد.

۵-۲- گام‌های روش‌شناسی پیشنهادی

در این بخش، روش‌شناسی پیشنهادی برای استخراج مدل نرخ بیت - اعوجاج برای ویدئوی چندنمایی سه‌بعدی در سطح نما معرفی می‌شود. ساختار پیش‌بینی معمول برای ویدئوی چندنمایی که توسط استاندارد H.264/MVC ارائه شده است، در شکل ۱-۵ نشان داده شده است. به دلیل وجود حجم بالای داده‌ها در ویدئوی چندنمایی سه‌بعدی، فشردگی کارآمد این ویدئوها به منظور ارسال و ذخیره‌سازی آنها لازم است. علاوه بر افزودن زمانی بین فریم‌های متوالی، افزودن مکانی بین فریم‌های نماهای مجاور در ساختار پیش‌بینی نیز می‌تواند به منظور افزایش کارایی فشردگی استخراج شود. برای این منظور از تکنیک‌های جبران‌سازی حرکت و اختلاف بین‌نمایی به منظور استخراج وابستگی‌های زمانی در هر نما و وابستگی‌های بین‌نمایی در نماهای مجاور استفاده می‌شود. بردارهای حرکت^۱ و اختلاف بین‌نمایی^۲ بر اساس معیار نرخ بیت - اعوجاج و به نحوی که نرخ بیت بر اساس محدودیت اعوجاج نهایی حداقل شود، تعیین می‌شوند. بنابراین، نرخ بیت هر نما می‌تواند به طور مؤثر به وسیله وابستگی بین نماهای مجاور در ساختار پیش‌بینی و نیز وابستگی بین فریم‌های متوالی هر نما تحت تأثیر قرار گیرد. بر اساس این بحث، روش‌شناسی ارائه شده در این رساله، پیشنهاد می‌کند که نرخ بیت هر نما باید تابعی از پارامترهای مرتبط با اختلاف‌های درون‌نمایی^۳ و بین‌نمایی^۴ باشد.

بنابراین می‌توان مدل مناسب برای نرخ بیت - اعوجاج ویدئوی چندنمایی در سطح نما را تابعی از اختلاف‌های درون‌نمایی و بین‌نمایی تعریف کرد:

¹ Motion vector

² Disparity vector

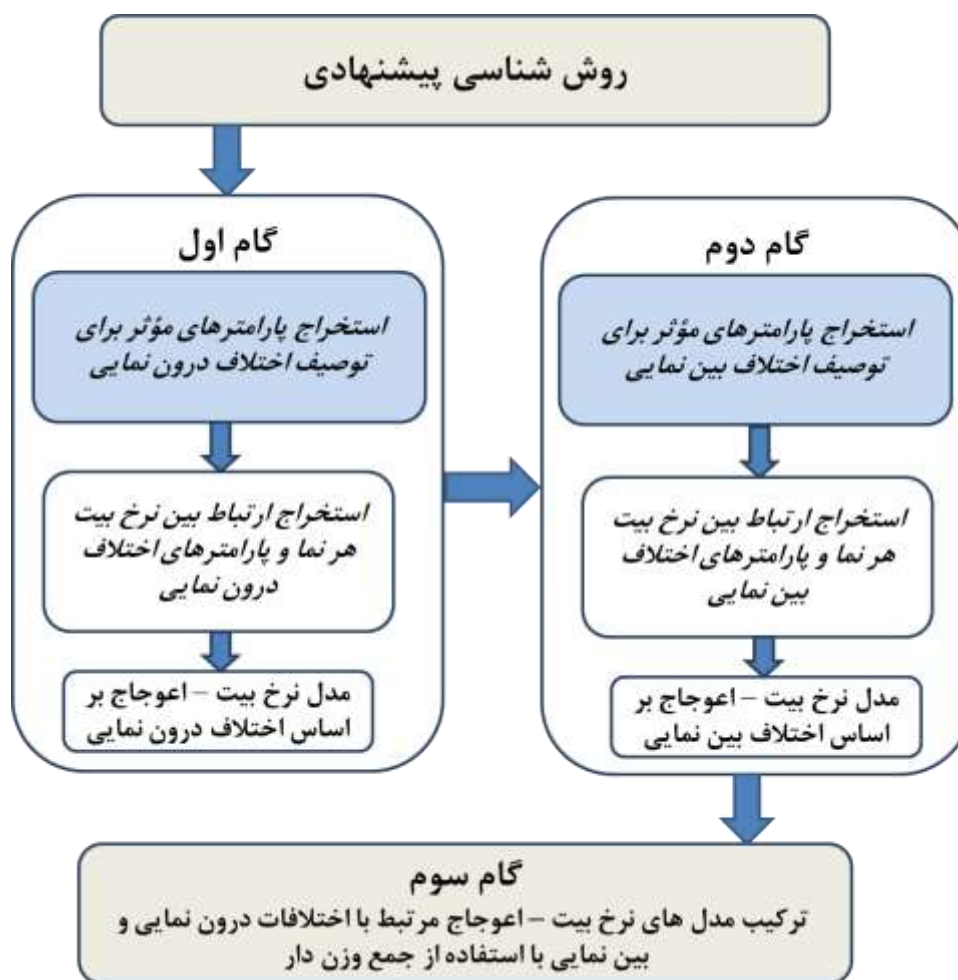
³ Intra-view disparity

⁴ Inter-view disparity

$$R = F(\text{Intra - view disparity, Inter - view disparity}) \quad (1-5)$$

که R نرخ بیت هر نما است و F رابطه بین نرخ بیت و پارامترهای شاخص توصیف‌کننده اختلاف بین‌نمایی و درون‌نمایی را نشان می‌دهد.

بر این اساس روش استخراج مدل نرخ بیت - اعوجاج مناسب در سطح نما باید دو مرحله داشته باشد. در گام‌های اول و دوم، به ترتیب پارامترهای مهم مرتبط با اختلاف‌های درون‌نمایی و بین‌نمایی بر اساس ویژگی‌های خاص ویدئوی چندنمایی سه‌بعدی و کاربرد موردنظر و نیز تابع مناسبی که رابطه بین نرخ بیت نما و این پارامترها را نشان می‌دهد، به صورت تحلیلی استخراج می‌شوند. بنابراین در پایان این دو مرحله، دو مدل نرخ بیت - اعوجاج خواهیم داشت که به ترتیب ارتباط بین نرخ بیت نما با پارامترها اختلاف درون‌نمایی و بین‌نمایی را نشان می‌دهند. در مرحله سوم، این دو مدل با یکدیگر ترکیب شده و مدل نهایی نرخ بیت - اعوجاج در سطح نما برای ویدئوی چندنمایی سه‌بعدی استخراج می‌شود. گام‌های مختلف روش‌شناسی پیشنهادی در شکل ۱-۵ نشان داده شده است. جزئیات مربوط به هر یک از این مراحل نیز در زیرفصل‌های ۱-۲-۵، ۲-۲-۵ و ۳-۲-۵ توضیح داده خواهد شد.



شکل ۵-۱ - ساختار کلی روش شناسی پیشنهادی برای استخراج مدل نرخ بیت - اعوجاج در سطح نما برای ویدئوی چندنمایی سه بعدی

۵-۲-۱- گام اول: استخراج مدل نرخ بیت - اعوجاج بر اساس اختلاف درون نمایی

۵-۲-۱-۱- استخراج پارامترهای مؤثر برای توصیف اختلاف درون نمایی

همان طور که پیش تر اشاره شد، تکنیک های جبران سازی حرکت به منظور استخراج وابستگی زمانی بین فریم های یک نما استفاده می شوند. بنابراین، فرآیند پیش بینی و در نتیجه نرخ بیت هر نما به طور مؤثر تحت تأثیر وابستگی بین فریم های متوالی هر نما خواهد بود. این وابستگی می تواند به پارامترهای مختلفی از جمله طول گروه تصاویر^۱، تعداد فریم های مرجع^۲، پیچیدگی محتوای ویدئو^۱ و غیره وابسته باشد. به منظور یافتن پارامترهای مؤثر

^۱ Group Of Picture (GOP)

^۲ Reference picture

برای توصیف اختلاف درون‌نمایی، باید پارامترهای مهمی که روی پیش‌بینی درون‌نمایی اثر می‌گذارند را استخراج کنیم.

۵-۲-۱-۱-۱- استخراج ارتباط بین نرخ بیت هر نما و پارامترهای شاخص اختلاف درون-

نمایی

در این مرحله باید ارتباط بین نرخ بیت نهایی هر نما و پارامترهای مرتبط با اختلاف درون-نمایی استخراج شود. روش‌شناسی ارائه‌شده در این رساله پیشنهاد می‌دهد که از یک رویه تحلیلی با استفاده از برازش منحنی برای این منظور استفاده شود. با استفاده از این رویه، یک مدل نرخ بیت - اعوجاج که ارتباط بین نرخ بیت هر نما و پارامترهای شاخص اختلاف درون‌نمایی را نشان می‌دهد، به دست می‌آید.

۵-۲-۲- گام دوم: استخراج مدل نرخ بیت - اعوجاج بر اساس اختلاف بین-

نمایی

۵-۲-۲-۱- استخراج پارامترهای مؤثر برای توصیف اختلاف بین‌نمایی

همان‌طور که پیش‌تر اشاره شده، علاوه بر افزونگی زمانی بین فریم‌های متوالی، افزونگی مکانی بین فریم‌های نماهای مجاور در ساختار پیش‌بینی ویدئوی چندنمایی نیز می‌تواند روی فرآیند پیش‌بینی و نرخ بیت نهایی هر نما تأثیر بگذارد. بنابراین همانند مرحله قبل باید اختلاف بین‌نمایی نیز توصیف شده و پارامترهای مؤثر متناظر استخراج شوند.

۵-۲-۲-۱- استخراج ارتباط بین نرخ بیت هر نما و پارامترهای شاخص اختلاف بین‌نمایی

همانند مرحله قبل، در این مرحله باید ارتباط بین نرخ بیت نهایی هر نما و پارامترهای مرتبط با اختلاف بین‌نمایی به صورت تحلیلی و با استفاده از برازش منحنی استخراج شود. به این ترتیب، در پایان این مرحله نیز یک مدل نرخ بیت - اعوجاج که ارتباط بین نرخ بیت نهایی هر نما و پارامترهای شاخص اختلاف‌های بین‌نمایی را نشان می‌دهد، استخراج می‌شود.

¹ Video content complexity

۵-۲-۳- مرحله سوم: مدل نرخ بیت - اعوجاج نهایی در سطح نما برای

ویدئوی چندنمایی سه بعدی

در مرحله نهایی باید دو مدل نرخ بیت - اعوجاج استخراج شده در دو مرحله قبل با یکدیگر ترکیب شده و مدل نهایی نرخ بیت - اعوجاج در سطح نما برای ویدئوی چندنمایی سه بعدی به دست آید.

همان طور که در ساختار پیش بینی شکل ۱-۵ نشان داده شده است، نماهای مختلف ویدئوی چندنمایی از پیش بینی های درون نمایی و بین نمایی به منظور افزایش کارایی فشرده سازی استفاده می کنند. نسبت تعداد پیش بینی های درون نمایی و بین نمایی برای نماهای مختلف متفاوت است. نتایج شبیه سازی ها برای ویدئوهای مختلف نشان می دهد که برای نماهایی که تنها یک نمای مرجع برای پیش بینی بین نمایی دارند، مانند V_2 در شکل ۱-۵، تعداد پیش بینی های بین نمایی به مراتب کمتر از تعداد پیش بینی های درون نمایی است. به طور مشابه، نسبت بین تعداد پیش بینی های بین نمایی به تعداد پیش بینی های درون نمایی برای نماهایی که دو نمای مرجع دارند، مانند V_1 در شکل ۱-۵، به مراتب بزرگتر است.

هرچه تعداد پیش بینی های بین نمایی بیشتر باشد، اهمیت اختلاف های بین نمایی در تعیین نرخ بیت نهایی بیشتر خواهد بود و هر چه تعداد پیش بینی های درون نمایی بیشتر باشد، اهمیت اختلاف های درون نمایی در تعیین نرخ بیت نهایی بیشتر خواهد بود. علاوه بر این، بردارهای حرکت و اختلاف بین نمایی که از پیش بینی های درون نمایی و بین نمایی استخراج می شوند، کاملاً از یکدیگر مستقل هستند. بر اساس این بحث، در این رساله پیشنهاد شده است که مدل نرخ بیت - اعوجاج در سطح نما برای ویدئوی چندنمایی به صورت حاصل جمع وزن دار نرخ بیت حاصل از اختلاف های درون نمایی و نرخ بیت حاصل از اختلاف های بین نمایی در نظر گرفته شود. روش شناسی ارائه شده در این رساله پیشنهاد می دهد که از نسبت بین تعداد پیش بینی های درون نمایی و بین نمایی در هر نما به منظور استخراج مقادیر وزن ها در جمع وزن دار استفاده شود.

۵-۳- استخراج مدل نرخ بیت - اعوجاج در سطح نما برای ویدئوی چندنمایی

سه بعدی بر اساس روش شناسی ارائه شده

در این بخش، با استفاده از سه مرحله روش شناسی پیشنهادی یک مدل نرخ بیت - اعوجاج در سطح نما برای ویدئوی چندنمایی سه بعدی به صورت تجربی استخراج می شود. جزئیات

روش اعمال سه مرحله این روش‌شناسی در زیر بخش‌های ۵-۳-۱، ۵-۳-۲ و ۵-۳-۳- توضیح داده خواهد شد.

۵-۳-۱- مرحله اول: استخراج مدل نرخ بیت - اعوجاج بر اساس اختلاف درون‌نمایی

۵-۳-۱-۱- استخراج پارامترهای مؤثر برای توصیف اختلاف درون‌نمایی

همان‌طور که در مرحله اول روش‌شناسی پیشنهادی توضیح داده شد، به‌منظور پیدا کردن پارامترهای مؤثر برای توصیف اختلاف درون‌نمایی، پارامترهایی که بیشترین اثر روی نرخ بیت نماها را دارند، باید استخراج شود.

در [104] یک روش‌شناسی مشخص برای انتخاب مناسب پارامترهای مؤثر در نرخ بیت نماهای یک ویدئوی چندنمایی ارائه شده است. در روش‌شناسی ارائه‌شده در این مقاله، ابتدا همه پارامترهای موجود در الگوریتم‌های کنترل نرخ بیت پیشنهاد شده قبلی برای ویدئوی تک‌نمایی و چندنمایی از روی منابع موجود شامل مقالات کنفرانس‌ها و مجلات و نیز پارامترهای قابل تنظیم در کدکننده‌های استاندارد، جمع‌آوری و دسته‌بندی شده و سپس میزان تأثیر هر یک از این پارامترها بر روی نرخ بیت هر نما از ویدئوی چندنمایی بررسی شده و در نهایت، پارامترهایی که بیشترین تأثیر را بر نرخ بیت هر نما داشته‌اند، استخراج شده است. پارامترهای تأثیرگذار در نرخ بیت هر نما بر اساس نتایج ارائه شده در [104] در جدول ۵-۱ نشان داده شده است.

در روش ارائه‌شده در [104] برای بررسی اثر هر یک از پارامترهای جدول ۵-۱ روی میزان نرخ بیت هر نما، اثر تغییر هر پارامتر روی نرخ بیت نهایی هر نما در کیفیت یکسان بررسی شده است. بدین منظور، با استفاده از کدکننده استاندارد H.264/MVC نسخه ۸/۵ [105]، تعدادی از نماهای مربوط به ویدئوهای چندنمایی مختلف با مقادیر مختلفی از هر پارامتر، کد شده‌اند، به نحوی که اثر پارامترهای دیگر ثابت در نظر گرفته شود. نتایج ارائه شده در [104] نشان می‌دهد که اثر پارامتر پیچیدگی محتوای ویدئو روی نرخ بیت هر نما از سایر پارامترها بیشتر است.

بنابراین بر اساس نتایج این مقاله، در این رساله به منظور توصیف اثر اختلاف درون‌نمایی از مفهوم پیچیدگی محتوای ویدئو استفاده خواهیم کرد.

۵-۳-۱-۱- استخراج ارتباط بین نرخ بیت هر نما و پارامترهای شاخص اختلاف درون-

نمایی

بر اساس گام‌های روش‌شناسی پیشنهادی در این رساله، در این مرحله باید ارتباط بین نرخ بیت هر نما و مفهوم پیچیدگی محتوای ویدئو را استخراج کنیم. به این منظور، ابتدا باید مفهوم پیچیدگی محتوای ویدئو را با پارامترهای مناسب توصیف کنیم. روش‌های متفاوتی برای پارامتری کردن پیچیدگی محتوای ویدئو در تحقیقات قبلی ارائه شده است. در این رساله، برای سادگی از پارامترهای ارائه شده در مقاله [106] برای پارامتری کردن پیچیدگی محتوای ویدئو استفاده می‌کنیم. استفاده از پارامترهای ارائه شده در این مقاله مزایایی دارد. از جمله اینکه، این پارامترها می‌توانند از روی متغیرهای مرتبط با کدکننده محاسبه شوند.

جدول ۵-۱ - پارامترهای تأثیرگذار بر نرخ بیت هر نما در ویدئوی چندنمایی [104]

پارامترهای تأثیرگذار در نرخ بیت نماها در ویدئوی چندنمایی
• تعداد بیت‌های سرآیند ^۱
• بردارهای حرکتی
• مود کد کردن در استانداردهای فشرده‌سازی ویدئو ^۲
• بیت‌های مربوط به داده‌های خطا ^۳
• طول گروه تصاویر ^۴
• محدوده جستجو ^۵
• تعداد فریم‌های مرجع
• تعداد نماها
• نوع ویدئو (بافت ^۶ یا عمق)
• ناحیه مورد توجه
• پیچیدگی محتوای ویدئو
• ترجیحات کاربر ^۷
• نوع فریم (I,P,B)
• نوع نما (I,P,B)
• پهنای باند کانال و نرخ انتقال

¹ Slice header bits

² Coding mode

³ Bits for error data

⁴ Length of GOP

⁵ Search range

⁶ Texture

⁷ User preference

بنابراین محاسبه آنها نسبت به حالتی که پیچیدگی محتوا از روی مقادیر پیکسل‌های فریم‌های فشرده نشده محاسبه شود، هزینه محاسباتی کمتری دارد. اگرچه کاهش پیچیدگی محتوا دقت محاسبات را نیز کاهش می‌دهد، اما نتایج شبیه‌سازی‌های ما نشان می‌دهد که محاسبه پارامترها به این روش دقت قابل قبولی دارد.

نکته قابل توجه این که استفاده از این پارامترها، تغییری در کلیت روش ارائه‌شده ایجاد نخواهد کرد، هر پارامتر دیگری که توصیف‌کننده پیچیدگی محتوای ویدئو باشد، نیز می‌تواند برای این منظور مورد استفاده قرار گیرد.

پارامترهای استفاده شده برای توصیف پیچیدگی محتوای ویدئو در این رساله بر اساس [106]، پیچیدگی صحنه^۱ و سطح حرکت^۲ است که معادله‌های (۲-۵) و (۳-۵) برای محاسبه این دو پارامتر ارائه شده‌اند [106]:

$$\text{Scene Complexity}(C) = \frac{\text{Bitrate}_I}{2 \times 10^6 \times 0.89^{QP_I}} \quad (2-5)$$

$$\text{Level of Motion}(M) = \frac{\text{Bitrate}_P + \text{Bitrate}_B}{2 \times 10^6 \times 0.89^{(QP_P + QP_B)}} \quad (3-5)$$

در این معادلات Bitrate_I ، Bitrate_P و Bitrate_B به ترتیب نرخ بیت فریم‌های گذشته I ، P و B و QP_I ، QP_P و QP_B نیز به ترتیب میانگین پارامترهای کوانتیزاسیون فریم‌های I ، P و B هستند. مقادیر ثابت استفاده شده در این معادلات بر این اساس انتخاب شده‌اند که در استاندارد H.264، ۵۲ مقدار برای اندازه گام کوانتیزاسیون وجود دارد که با مقادیر پارامتر کوانتیزاسیون (QP) نشان داده می‌شوند. این مقادیر به گونه‌ای مرتب شده‌اند که افزایش ۶تایی در مقدار پارامتر کوانتیزاسیون به معنی دو برابر کردن مقدار گام کوانتیزاسیون است. بنابراین افزایش یک واحدی در مقدار پارامتر کوانتیزاسیون مطابق با کاهش نرخ بیت به اندازه حدوداً $1 - \frac{1}{2^6} = 0.89$ خواهد بود.

همان‌طور که پیش‌تر اشاره شد، روش ارائه شده در این رساله مستقل از انتخاب پارامتر مناسب برای مدل کردن پیچیدگی محتوای ویدئو است و پارامترهای انتخاب شده، کلیت روش ارائه‌شده برای استخراج مدل نرخ بیت - اعوجاج را تحت تأثیر قرار نخواهد داد. حال می‌توانیم ارتباط بین نرخ بیت نهایی هر نما و مفهوم پیچیدگی محتوای ویدئو را با استفاده از پارامترهای C و M استخراج کنیم. جزئیات این فرآیند به شرح زیر است.

¹ Scene complexity

² Level of motion

از لحاظ تئوری، تابع پیچیدگی محتوای ویدئو با حاصل ضرب پارامتر کوانتیزاسیون و نرخ بیت به صورت زیر تعریف می‌شود [108]:

$$\text{Coding Complexity} = QP \times R \quad (4-5)$$

با استفاده از این معادله، می‌توان به صورت تجربی و با استفاده از برازش منحنی^۱، پیچیدگی محتوا را به صورت تابعی از پارامترهای پیچیدگی در نظر گرفته شده در این رساله به صورت زیر به دست آورد:

$$QP \times R = F(C, M) \quad (5-5)$$

به عبارت دیگر، پیچیدگی کد کردن می‌تواند به صورت تابعی از پارامترهای شاخص پیچیدگی محتوای ویدئو، که به طور مستقیم روی پیچیدگی کد کردن اثر می‌گذارد، تعریف شود. F در معادله (۵-۵) این تابع را نشان می‌دهد که می‌تواند به صورت تحلیلی و با استفاده از برازش منحنی استخراج شود.

همان‌طور که در اکثر مقالات مربوط به مدل سازی نرخ بیت - اعوجاج عنوان شده است [108]، به منظور استخراج این مدل به صورت تحلیلی به روش زیر عمل می‌کنیم. تعداد ۱۰۰ فریم از تعداد زیادی نما از ویدئوهای چندنمایی مختلف و با پیچیدگی و وضوح تصویر^۲ متفاوت انتخاب و با پارامتر کوانتیزاسیون ثابت و با کد کننده H.264/MVC کد شده‌اند. جدول ۲-۵ مشخصات ویدئوهای آزمایش شده را نشان می‌دهد. این نماها با پارامترهای کوانتیزاسیون ثابت و با استفاده از کد کننده استاندارد H.264/MVC نسخه ۸/۵ [105] کد می‌شوند. نکته قابل ذکر اینکه در همه ویدئوهای آزمایش شده در این جدول به جز ویدئوی Race، دوربین‌ها ثابت هستند.

سپس مقادیر پارامتر کوانتیزاسیون و نرخ بیت نماهای کد شده در فرآیند برازش منحنی استفاده می‌شود تا ارتباط بین $QP \times R$ و پارامترهای C و M همان‌طور که در (۵-۵) بیان شده، تعیین شود. هدف فرآیند برازش منحنی پیدا کردن پارامترهای یک مدل ریاضی است که قادر به توصیف داده‌های موجود باشد به نحوی که اختلاف بین مدل استخراج شده و داده‌ها حداقل شود.

¹ Curve fitting

² Resolution

جدول ۲-۵ - مشخصات ویدئوهای کد شده برای استخراج پارامترهای مدل نرخ بیت - اعوجاج

تعداد فریم	تعداد نما	نرخ فریم (فریم بر ثانیه)	اندازه فریم	ویدئو
۱۰۰	۸	۱۵	۱۰۲۴×۷۶۸	Ballet
۱۰۰	۸	۱۵	۱۰۲۴×۷۶۸	Break-dancer
۵۰۰	۷	۲۵	۱۰۲۴×۷۶۸	Balloons
۴۰۰	۷	۲۵	۱۰۲۴×۷۶۸	Kendo
۱۰۰۰	۵	۱۵	۶۴۰×۴۸۰	Crowd
۱۰۰۰	۵	۱۵	۶۴۰×۴۸۰	Flamenco
۶۲۵	۷	۱۵	۶۴۰×۴۸۰	Object
۵۳۰	۸	۱۵	۶۴۰×۴۸۰	Race
۵۰۰	۸	۱۵	۱۲۸۰×۹۶۰	Tower

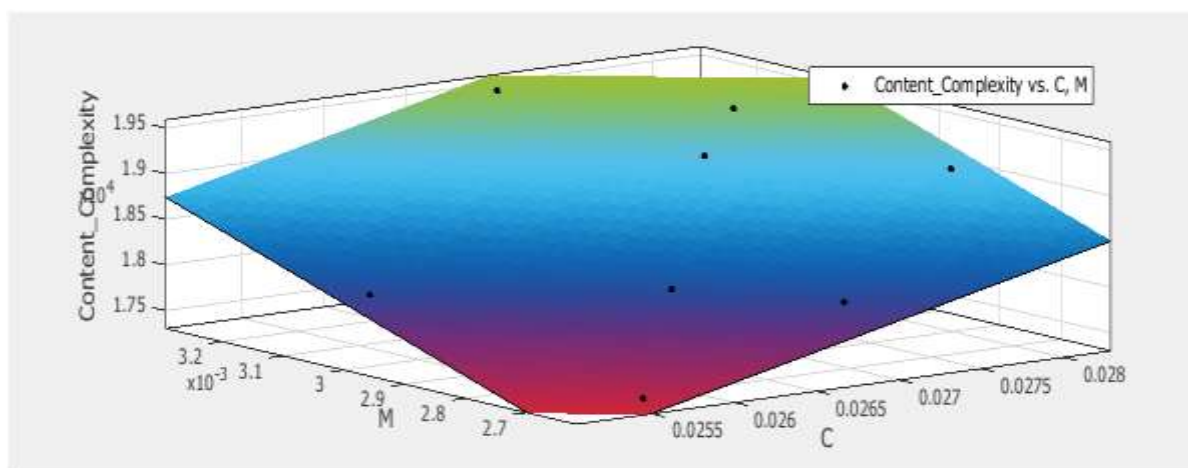
شکل ۲-۵ میزان پیچیدگی ویدئوی کد شده $(QP \times R)$ را به صورت تابعی از پارامترهای پیچیدگی محتوا، C و M ، برای ویدئوی Ballet و با پارامتر کوانتیزاسیون ۲۰ نشان می‌دهد. این شکل نشان می‌دهد که معادله چندجمله‌ای درجه اول برای توصیف داده‌های موجود مناسب است.

دقت فرآیند برازش منحنی با استفاده از R-square و خطای جذر میانگین مربعات (RMSE)^۱ اندازه‌گیری می‌شود. این مقادیر آماری نشان می‌دهند که مدل به‌دست آمده توسط فرآیند برازش منحنی تا چه اندازه با داده‌های اصلی تطبیق دارد. R-square یک پارامتر آماری است که میزان تطبیق نتیجه برازش منحنی با میزان تغییرات داده اصلی نشان می‌دهد. مقدار ۱ برای این پارامتر نشان‌دهنده این است که خط رگرسیون دقیقاً بر داده‌ها منطبق است. این مقادیر از روابط زیر استخراج می‌شوند:

$$RMSE = \sqrt{\frac{\sum_{i=0}^{n-1} (y_i - f(x_i))^2}{DOF}} \quad (۶-۵)$$

$$R - Square = 1 - \frac{\sum_{i=0}^{n-1} (y_i - f(x_i))^2}{\sum_{i=0}^{n-1} (y_i - \bar{y})^2} \quad (۷-۵)$$

^۱ Root Mean Square Error



شکل ۵-۲- پیچیدگی محتوای واقعی به صورت تابعی از پارامترهای ارائه شده برای اندازه‌گیری پیچیدگی ویدئو (C, M)

در این معادلات، y_i مقدار پیسکل i داده اصلی، $f(x_i)$ مقدار پیسکل i در منحنی استخراج شده، DOF درجه آزادی، n تعداد کل پیسکل‌ها و \bar{y} مقدار میانگین داده‌های اصلی است. میزان خطای جذر میانگین مربعات (RMSE)، در شکل ۵-۲، 0.0009 و مقدار R-square برابر ۱ است. در این رساله برای فرآیند برازش منحنی و استخراج مقادیر آماری از ابزار برازش منحنی نرم‌افزار MATLAB استفاده شده است.

نتیجه شکل ۵-۲ برای یک پارامتر کوانتیزاسیون ثابت است و برای نقاط آزمایش شده نشان می‌دهد که رابطه نرخ بیت با پارامترهای در نظر گرفته شده برای پیچیدگی، رابطه‌ای خطی است. به عبارت دیگر، برای پارامتر کوانتیزاسیون ثابت، نتایج برازش منحنی می‌تواند به صورت زیر بیان شود:

$$R = \alpha \times C + \beta \times M + \gamma \quad (۸-۵)$$

که R نرخ بیت هر نما و α ، β و γ ضرایب ثابتی هستند که از برازش منحنی به دست آمده‌اند. بنابراین، در حالت کلی که پارامتر کوانتیزاسیون ثابت نباشد، خواهیم داشت:

$$QP \times R = \alpha (QP) \times C + \beta(QP) \times M + \gamma(QP) \quad (۹-۵)$$

بنابراین

$$R = \frac{\alpha(QP)}{QP} \times C + \frac{\beta(QP)}{QP} \times M + \frac{\gamma(QP)}{QP} = a(QP) \times C + b(QP) \times M + c(QP) \quad (۱۰-۵)$$

توابع $a(QP)$ و $b(QP)$ و $c(QP)$ جایگزین α ، β و γ در حالت کلی هستند. البته به منظور تطابق با مدل‌های قبلی نرخ بیت - اعوجاج، در مدل ارائه شده در این رساله نیز عکس

پارامتر کوانتیزاسیون را در نظر می گیریم. بنابراین رابطه نرخ بیت هر نما با پارامترهای پیچیدگی محتوای ویدئو را می توان به صورت زیر نشان داد:

$$R(QP^{-1}) = a(QP^{-1}) \times C + b(QP^{-1}) \times M + c(QP^{-1}) \quad (11-5)$$

به منظور استخراج ضرایب $a(QP^{-1})$ ، $b(QP^{-1})$ و $c(QP^{-1})$ رابطه (11-5) باید فرآیند برازش منحنی برای مقادیر مختلف پارامتر کوانتیزاسیون تکرار شود. برای هر مقدار پارامتر کوانتیزاسیون، یک مقدار برای $a(QP^{-1})$ ، $b(QP^{-1})$ و $c(QP^{-1})$ از فرآیند برازش منحنی استخراج می شود. این مقادیر می توانند به صورت زیر برای استخراج $a(QP^{-1})$ ، $b(QP^{-1})$ و $c(QP^{-1})$ مورد استفاده قرار گیرند.

تعداد 100 فریم از نماهای مختلف از ویدئوهای چندنمایی آزمایش شده از جدول 5-2 با پارامترهای کوانتیزاسیون 15، 20، 25 و 30 کد شدند. سپس برای هر نمای گذشته مقادیر C و M از روی روابط (2-5) و (3-5) به دست آمد. جدول 5-3 مقادیر استخراج شده برای این پارامترها و نیز حاصل ضرب نرخ بیت و پارامتر کوانتیزاسیون برای ویدئوی Ballet و به ازای پارامتر کوانتیزاسیون 15 را نشان می دهد. به طریق مشابه، مقدار این پارامترها برای نماهای مختلف گذشته برای سایر ویدئوهای چندنمایی جدول 5-2 و نیز به ازای مقادیر مختلف پارامتر کوانتیزاسیون استخراج شد.

سپس با استفاده از روش برازش منحنی، نمودارهای مقدار حاصل ضرب نرخ بیت در پارامتر کوانتیزاسیون بر حسب مقادیر C و M رسم شده و ضرایب a، b و c رابطه (11-5) استخراج شد. مقادیر استخراج شده برای این ضرایب و به ازای مقادیر متفاوت پارامتر کوانتیزاسیون و نیز پارامترهای میانگین مربعات خطا و R-square نشان دهنده دقت فرآیند برازش منحنی در جدول 5-4 نشان داده شده است.

جدول 5-3 - مقادیر پارامترهای C و M و حاصل ضرب نرخ بیت در پارامتر کوانتیزاسیون برای نماهای مختلف

ویدئوی Ballet و به ازای پارامتر کوانتیزاسیون 15

M	C	Bitrate x QP	نمای گذشته
0.0068	0.0423	49355/78	نمای 0
0.0062	0.0425	47466/12	نمای 1
0.0067	0.0432	49682/14	نمای 2
0.0062	0.0411	46749/83	نمای 3
0.0060	0.0416	46262/21	نمای 4
0.0059	0.0425	46476/81	نمای 5
0.0065	0.0436	49264/43	نمای 6
0.0065	0.0422	48246/75	نمای 7

جدول ۴-۵ - مقادیر ضرایب استخراج شده a، b و c رابطه (۵-۱۱) و پارامترهای میانگین مربعات خطا و R-square مربوط به فرآیند برازش منحنی

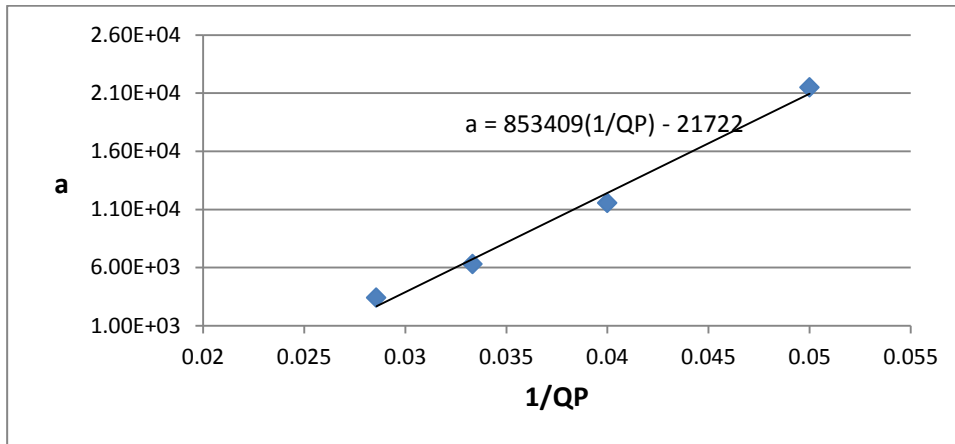
QP = 15				
a	b	c	RMSE	R-square
۴۰۸۹۳/۳۳	۲۳۰۲۶۶/۷	۰/۰۹۹۶	۰/۰۵۷۵۵	۱
QP = 20				
a	b	c	RMSE	R-square
۲۱۵۰۰	۱۲۱۱۰۰	۰/۱۰۱۱۵	۰/۰۸۲۲۸	۱
QP = 25				
a	b	c	RMSE	R-square
۱۱۵۸۴	۶۵۰۴۰	-۰/۹۴۲۸	۱۰۴	۰/۹۹۹۷
QP = 30				
a	b	c	RMSE	R-square
۶۲۸۶/۶۶۷	۳۵۴۳۳/۳۳	۰/۱۵۳۳۶۷	۹/۴۹۶	۱

سپس نمودار مقادیر این ضرایب بر حسب عکس پارامتر کوانتیزاسیون و معادلات متناظر با آنها استخراج شد. شکل ۳-۵ نمودارهای مربوط را نشان می‌دهد. معادلات استخراج شده برای ضرایب a، b و c نیز به ترتیب در (۵-۱۲)، (۵-۱۳) و (۵-۱۴) نشان داده شده‌اند. به این ترتیب همه مقادیر این ضرایب به‌ازای پارامترهای کوانتیزاسیون مختلف را می‌توان از این معادلات استخراج کرد. در نهایت نیز با استفاده از این ضرایب و با توجه به رابطه (۵-۱۱) می‌توان نرخ بیت هر نما را بر اساس پارامترهای میزان پیچیدگی محتوای آنها به‌دست آورد.

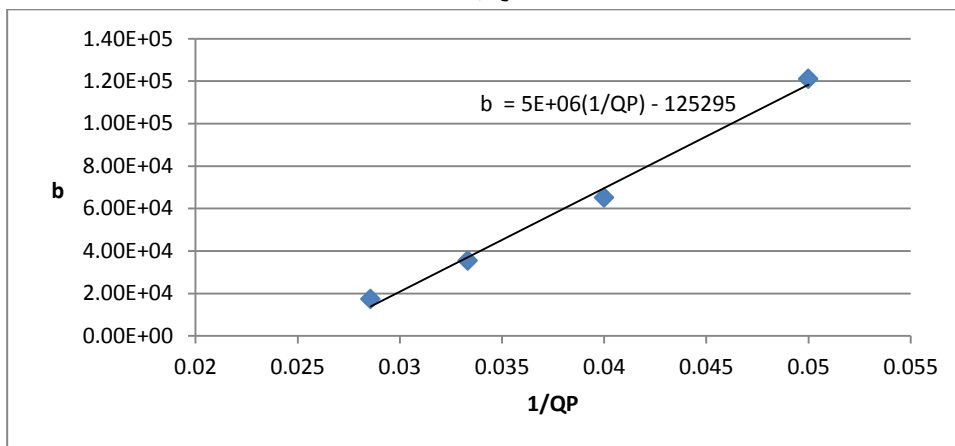
$$a(QP^{-1}) = 853409(QP^{-1}) - 21722 \quad (۵-۱۲)$$

$$b(QP^{-1}) = 5E + 6(QP^{-1}) - 125295 \quad (۵-۱۳)$$

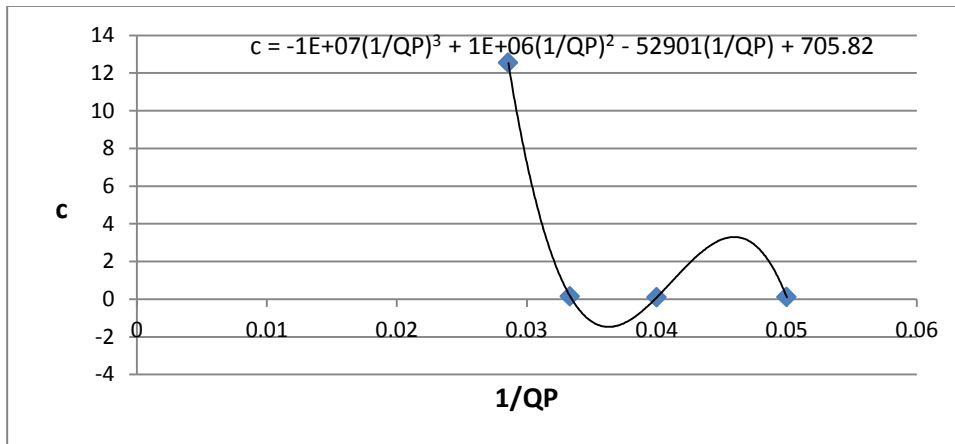
$$c(QP^{-1}) = -1E + 07(QP^{-1})^3 + 1E + 6(QP^{-1})^2 - 52901(QP^{-1}) + 705.82 \quad (۵-۱۴)$$



ضریب a



ضریب b



ضریب c

شکل ۳-۵ - ضرایب a، b و c مربوط به رابطه (۵-۱۱) بر حسب مقدار عکس پارامتر کوانتیزاسیون

۵-۳-۲- مرحله دوم: استخراج مدل نرخ بیت - اعوجاج بر اساس اختلاف

بین‌نمایی

۵-۳-۲-۱- استخراج پارامترهای مؤثر برای توصیف اختلاف بین‌نمایی

همان‌طور که در شکل ۵-۱ نشان داده شده است، علاوه بر وابستگی‌های درون‌نمایی، وابستگی‌های بین‌نمایی نیز در کارایی فشرده‌سازی ویدئوی چندنمایی تأثیرگذار است. در ساختار پیش‌بینی نشان داده شده در این شکل، فریم‌های نمای V_0 برای کد شدن تنها از پیش‌بینی‌های درون‌نمایی یا زمانی^۱ استفاده می‌کنند. فریم‌های نمای V_2 علاوه بر پیش‌بینی‌های درون‌نمایی از پیش‌بینی‌های بین‌نمایی نیز به منظور افزایش کارایی فشرده‌سازی استفاده می‌کنند. به‌طور مشابه، فریم‌های نمای V_1 نیز علاوه بر پیش‌بینی‌های زمانی از نماهای V_0 و V_2 نیز به منظور پیش‌بینی بین‌نمایی استفاده می‌کنند.

به منظور تحلیل توزیع نرخ بیت ویدئوی چندنمایی، این ساختار پیش‌بینی را برای کد کردن ۴ نما از چندین ویدئوی چندنمایی استفاده کردیم. برای کد کردن این ویدئوها دو حالت مختلف با میانگین اختلاف بین‌نمایی متفاوت در نظر گرفته شده است. در حالت اول، نماها میانگین اختلاف بین‌نمایی پایین و در حالت دوم، نماها میانگین اختلاف بین‌نمایی بالایی دارند.

به‌منظور پیدا کردن نماهایی با کمترین میانگین اختلاف بین‌نمایی، به این ترتیب عمل می‌کنیم. اول نمای V_0 به عنوان نمای اصلی در شکل ۵-۱ انتخاب می‌شود. بر اساس ساختار پیش‌بینی نشان داده شده در این شکل، V_2 باید از روی V_0 و به صورت بین‌نمایی پیش‌بینی شود. بنابراین، از بین تمام نماهای باقی‌مانده، نمایی با حداقل اختلاف بین‌نمایی با V_0 به عنوان V_2 انتخاب می‌شود. به‌طور مشابه، نمایی با حداقل اختلاف بین‌نمایی با V_0 و V_2 به عنوان V_1 و نمایی با حداقل اختلاف بین‌نمایی با V_2 به عنوان V_3 انتخاب می‌شود. به‌صورت مشابه، نماهای مربوط به حالت دوم با بیشترین میانگین اختلاف بین‌نمایی نیز انتخاب می‌شوند. برای مثال، برای ویدئوی Ballet و از بین ۸ نمای موجود، نماهای ۰، ۱، ۲ و ۴ برای حالت کمترین میانگین اختلاف بین‌نمایی و نماهای ۰، ۲، ۱ و ۵ برای حالت بیشترین میانگین اختلاف بین‌نمایی انتخاب می‌شوند. سپس ۱۰۰ فریم از هریک از نماهای این ویدئوها با استفاده از کدکننده H.264/MVC [105] کد شدند. نتایج این آزمایش‌ها برای ویدئوی Ballet در جدول ۵-۵ نشان داده شده است. این نتایج نشان می‌دهد که هر نما با میانگین اختلاف بین‌نمایی پایین‌تر می‌تواند با نرخ بیت پایین‌تری در کیفیت یکسان کد شود.

¹ Temporal prediction

جدول ۵-۵- توزیع نرخ بیت در سطح نما برای ویدئوی Ballet در دو حالت مختلف از لحاظ میانگین اختلاف بین‌نمایی

حالت دوم		حالت اول		نما
میانگین اختلاف بین‌نمایی بالا		میانگین اختلاف بین‌نمایی پایین		
QP = ۱۵				
نرخ بیت (کیلو بیت بر ثانیه)	PSNR	نرخ بیت (کیلو بیت بر ثانیه)	PSNR	
۲۵۲۲	۴۳/۶۲	۲۵۲۲	۴۳/۶۲	نمای ۰
۲۴۴۵	۴۳/۵۳	۲۲۸۵	۴۳/۶۴	نمای ۱
۲۳۹۲	۴۳/۶۵	۲۳۶۳	۴۳/۶۴	نمای ۲
۲۵۱۴	۴۳/۵۴	۲۳۷۱	۴۳/۷۱	نمای ۳
QP = ۲۰				
نرخ بیت (کیلو بیت بر ثانیه)	PSNR	نرخ بیت (کیلو بیت بر ثانیه)	PSNR	
۳۷۹	۴۱/۶۷	۳۷۹	۴۱/۶۷	نمای ۰
۳۲۳	۴۱/۵۷	۲۹۲	۴۱/۶۳	نمای ۱
۳۵۶	۴۱/۸۵	۳۳۶	۴۱/۸۵	نمای ۲
۳۶۱	۴۱/۵۳	۳۴۱	۴۱/۵۸	نمای ۳
QP = ۲۵				
نرخ بیت (کیلو بیت بر ثانیه)	PSNR	نرخ بیت (کیلو بیت بر ثانیه)	PSNR	
۳۷۲	۴۱/۷۵	۳۷۲	۴۱/۷۵	نمای ۰
۳۰۸	۴۱/۶۴	۲۸۰	۴۱/۷۲	نمای ۱
۳۰۹	۴۱/۷۲	۳۲۶	۴۱/۹۴	نمای ۲
۳۵۲	۴۱/۶۵	۳۱۲	۴۱/۸۹	نمای ۳
QP = ۳۰				
نرخ بیت (کیلو بیت بر ثانیه)	PSNR	نرخ بیت (کیلو بیت بر ثانیه)	PSNR	
۲۰۸	۴۰/۲۰	۲۰۸	۴۰/۲۰	نمای ۰
۱۶۷	۳۹/۹۳	۱۵۳	۴۰/۰۲	نمای ۱
۱۶۵	۳۹/۹۶	۱۸۲	۴۰/۳۵	نمای ۲
۱۹۳	۴۰	۱۶۶	۴۰/۲۱	نمای ۳

از آنجایی که اندازه‌گیری بصری کیفیت زمان‌بر و پیچیده است، به منظور تحلیل سریع داده-ها، مانند بسیاری از تحقیقات قبلی، از معیار کمی کیفیت نسبت سیگنال به نویز (PSNR) استفاده شده است. اگرچه PSNR معیار خیلی دقیقی برای ارزیابی کیفیت تجربی در ویدئوی دوبعدی و سه‌بعدی نیست، اما می‌تواند برای تخمین کیفیت به‌خصوص هنگامی که

محتوای ویدئو و نوع اعوجاج ثابت می‌مانند و تنها سطوح اعوجاج تغییر می‌کنند استفاده شود [109].

علاوه بر نتایج جدول ۵-۵، آزمایشات مشابهی نیز برای ویدئوهای Kendo, Break-dancer و Balloons با پارامترهای کوانتیزاسیون مختلف (در محدوده ۱۵ تا ۳۰) انجام شده است که نتایج جدول ۵-۵ را تأیید می‌کند.

این نتایج نشان می‌دهد که برای کارایی بهتر فشرده‌سازی ویدئوی چند نمایی، هر نما باید از روی نماهایی که کمترین اختلاف بین نمایی را با آن نما دارند، پیشگویی شود. در این رساله، این مسأله با کنترل کردن نرخ بیت هر نما با استفاده از مفهوم میانگین اختلاف بین نمایی حل شده است. به عبارت دیگر، در مدل نرخ بیت - اعوجاج پیشنهادی در این رساله، رابطه‌ای مستقیم بین نرخ بیت هر نما و پارامتر میانگین اختلاف بین نمایی در نظر گرفته شده است.

علاوه بر این پارامتر، توان پردازشی گیرنده نیز پارامتر دیگری است که باید در مدل نرخ بیت - اعوجاج ویدئوی چند نمایی در نظر گرفته شود. در حقیقت در کاربردهای ویدئوی چند نمایی، نوعی مصالحه بین کیفیت، پهنای باند و توان پردازشی وجود دارد. این مسأله را با در نظر گرفتن مورد واقعی زیر توضیح می‌دهیم:

(۱) برای یک کدگشا با محدودیت توان پردازشی و پهنای باند کافی، مانند اکثر تبلت‌ها یا گوشی‌های هوشمند که از ارتباطات با سرعت بالا (3G/4G) به منظور بهبود کیفیت تجربی استفاده می‌کنند، بهترین راه حل ارسال تمام نماها و خودداری از باز تولید نماها با الگوریتم‌های پیچیده است. در این حالت، واحد کنترل کننده نرخ بیت باید پهنای باند موجود را بین همه نماها به نحو مناسب تقسیم کند.

(۲) برای یک کدگشا با توان پردازشی کافی و محدودیت پهنای باند، مانند یک لپ‌تاپ یا یک تبلت با شبکه ارتباطی ضعیف یا همین دستگاه‌ها در شرایطی که با باتری کار می‌کنند، کم کردن توان مصرفی ارسال داده‌ها (که بیشترین قسمت توان مصرفی را به خود اختصاص می‌دهد) لازم است. بنابراین لازم است که نرخ بیت با ارسال نکردن تعدادی از نماها به مقدار قابل توجهی کاهش یابد. در این حالت، کنترل کننده نرخ بیت باید پهنای باند موجود را تنها به نماهای با اهمیت بیشتر اختصاص دهد و سایر نماها را حذف کند. به این ترتیب، نماهای حذف شده در سمت گیرنده می‌توانند از روی نماهای ارسال شده که کیفیت خوبی دارند، به نحو کارآمدی باز تولید شوند.

موارد اشاره شده در بالا اهمیت سه پارامتر میانگین اختلاف بین نمایی، توان پردازشی و کیفیت تجربی را در مدل نرخ بیت - اعوجاج ویدئوی چندنمایی سه بعدی نشان می‌دهد.

در این رساله، توان پردازشی به توان مصرفی مورد نیاز برای هر گیرنده اشاره دارد که از روی پیشینه کاربرد مورد نظر یا از روی پروفایل کدگشا در سمت کدکننده قابل تشخیص است. برای مثال، پروفایل‌های مختلفی برای تبلت‌ها، نوت بوک‌ها^۱ و گوشی‌های هوشمند^۲ با محدودیت‌های متفاوت برای توان پردازشی و مصرفی وجود دارد. در این رساله، یک معیار ساده برای توان مصرفی برای کاربردهای ویدئوی چندنمایی سه‌بعدی ارائه شده است. این معیار تعداد نماهای قابل بازتولید در سمت گیرنده بر اساس محدودیت‌های توان پردازشی آن گیرنده برحسب کاربرد یا پروفایل کدگشا است.

علاوه بر این، مدل پیشنهادی از مفهوم اختلاف بین نمایی به منظور توزیع کارآمد پهنای باند محدود بین نماهای مختلف و بر اساس محدودیت‌های کاربرد موردنظر استفاده می‌کند. همانند قسمت‌های پیشین، برای محاسبه اختلاف بین نمایی از روشی که در [107] ارائه شده است، استفاده می‌شود. در نهایت مدل نرخ بیت - اعوجاج پیشنهادی با استفاده از پارامترهای مصرف توان، میانگین اختلاف بین نمایی و پارامتر کوانتیزاسیون استخراج می‌شود. با استفاده از این مؤلفه‌ها، مدل پیشنهادی می‌تواند در الگوریتم‌های مختلف کنترل نرخ بیت استفاده شده و برای محدوده گسترده‌ای از کاربردهای ویدئوی چندنمایی سه‌بعدی نیز تنظیم‌پذیر خواهد بود.

۵-۳-۱-۱-۱- استخراج ارتباط بین نرخ بیت هر نما و پارامترهای شاخص اختلاف درون-

نمایی

بر اساس مراحل توضیح داده شده در روش‌شناسی پیشنهاد شده، در این مرحله باید ارتباط بین نرخ بیت هر نما و پارامترهای شاخص اختلاف بین نمایی استخراج شود. همانند مرحله قبل، در این مرحله نیز از روش تحلیلی استفاده خواهد شد.

بر اساس توضیحات بخش قبل، می‌توان گفت که نرخ بیت نهایی برای هر نمای گذشته ویدئوی چندنمایی و همین‌طور کیفیت نهایی به میانگین اختلاف بین نمایی و میزان توان پردازشی بستگی دارد. به منظور سادگی و بدون تغییر دادن کلیت روش ارائه شده، از بین پارامترهای کدکننده که روی نرخ بیت و کیفیت ویدئوی گذشته اثر می‌گذارند، مانند روش‌های قبلی ارائه شده در پیشینه تحقیق [108]، تنها مقیاس کوانتیزاسیون^۳ را در نظر می‌گیریم. بنابراین، ارتباط بین نرخ بیت (R)، مقیاس کوانتیزاسیون (QS)، میانگین اختلاف بین نمایی (avg_inter_VD) و توان پردازشی (p) می‌تواند به صورت زیر بیان شود:

¹ Notebook

² Smart phone

³ Quantization Scale(QS)

$$R = F(QS, \text{avg_inter_VD}, p) \quad (15-5)$$

در این رابطه، F نشان دهنده رابطه بین نرخ بیت هر نما و پارامترهای مذکور است. میانگین اختلاف بین‌نمایی برای نمای z نیز از رابطه زیر به دست می‌آید:

$$\text{avg_inter_VD}_j = (\sum_{i=1}^V d_{ji}) / V \quad (16-5)$$

که V تعداد کل نماها در ساختار پیش‌بینی و d_{ji} اختلاف بین‌نمایی برای نماهای i و z است. همان‌طور که توضیح داده شد، اختلاف بین نماها از روشی که در [107] ارائه شده است، به دست می‌آید.

بر اساس توضیحات بخش قبل و نتایج جدول ۵-۵، رابطه مستقیمی بین نرخ بیت نهایی هر نما و پارامتر میانگین اختلاف بین‌نمایی وجود دارد. علاوه بر این، همان‌طور که پیش‌تر توضیح داده شد، اگر توان پردازشی کافی برای بازتولید نماهای حذف شده داشته باشیم، نرخ بیت موجود می‌تواند تنها به نماهای با اهمیت اختصاص داده شود تا با کیفیت بهتری در سمت گیرنده واکنش شوند. نماهای کم اهمیت می‌توانند به‌طور کامل حذف شده و در سمت گیرنده با کمک نماهایی موجود و با کیفیت بالا بازتولید شوند. بر اساس این توضیحات، بین نرخ بیت نهایی نماهای ارسال شده و معیار توان مصرفی نیز رابطه مستقیمی وجود دارد. به‌طور خلاصه و برای سادگی بیشتر، پارامتر جدیدی با نام X به صورت حاصل ضرب میانگین اختلاف بین‌نمایی و توان پردازشی تعریف می‌کنیم:

$$X = \text{avg_inter_VD} \times p \quad (17-5)$$

بنابراین می‌توان معادله (۱۵-۵) را به صورت تابعی از پارامتر X و مقیاس کوانتیزاسیون (QS) نوشت:

$$R = F(QS, X) \quad (18-5)$$

در همه استانداردهای کدکردن ویدئو، اعوجاج تابع افزایشی یکنواختی از QS است. در استاندارد H.264/AVC نیز ارتباط بین QS و پارامتر کوانتیزاسیون (QP) به صورت زیر است:

$$QS = 2^{QP/6} \quad (19-5)$$

بنابراین ارتباط بین نرخ بیت و QP می‌تواند به عنوان مدل نرخ بیت - اعوجاج در نظر گرفته شده و به صورت زیر بیان شود:

$$R(D) = F(QP, X) \quad (۲۰-۵)$$

روش استخراج مدل نهایی نرخ بیت - اعوجاج شبیه روشی است که در مقالات قبلی [108] بیان شده است. ابتدا تعداد زیادی نما از ویدئوهای چندنمایی شامل Break-dancer, Ballet و Kendo و Balloons [110][98] را انتخاب می‌کنیم. مشخصات این ویدئوها در جدول ۲-۵ ارائه شده است. سپس حالت‌های مختلفی برای میانگین اختلاف بین‌نمایی و میزان توان پردازشی در نظر گرفته می‌شود. به این ترتیب که نماهای انتخاب‌شده با مقادیر متفاوت QP (در محدوده ۱۵ تا ۳۰) و ترکیبات مختلف میانگین اختلاف بین‌نمایی (در محدوده ۰/۱ تا ۰/۵) و میزان توان پردازشی در محدوده (۲ تا ۶) کد شدند. کدکننده H.264/MVC نسخه ۸/۵ [105] و ساختار پیش‌بینی شکل ۱-۵ برای کدکردن ویدئوهای آزمایش‌شده مورد استفاده قرار گرفته است.

نتایج شبیه‌سازی‌ها برای ویدئوی Ballet برای مقادیر مختلف پارامتر کوانتیزاسیون (QP) و محدوده متفاوت میانگین اختلاف بین‌نمایی و توان پردازشی (p) در جدول ۵-۶ نشان داده شده است.

همان‌طور که قبلاً توضیح دادیم، با افزایش توان پردازشی، تعداد بیشتری نما می‌تواند در سمت گیرنده بازتولید شود. بنابراین، پهنای باند موجود می‌تواند بین تعداد کمتری نما توزیع شود و نرخ بیت هر نما افزایش یابد. اعداد نشان‌داده شده در سطرهای جدول ۵-۶ این فرضیه را ثابت می‌کند. همچنین، با افزایش میانگین اختلاف بین‌نمایی، نماها با کارآمدی کمتر از روی یکدیگر پیش‌بینی خواهند شد و بنابراین نرخ بیت آنها افزایش می‌یابد. ستون‌های نرخ بیت در جدول ۵-۶ نیز این مسأله را نشان می‌دهند. شبیه‌سازی‌های مشابه برای ویدئوهای دیگر و ترکیبات مختلف از پارامترهای ذکر شده انجام شده که نتایج ارائه شده در این جدول را تأیید می‌کند.

جدول ۵-۶- مقادیر انتخابی برای پارامتر کوانتیزاسیون، میانگین اختلاف بین نمایی و توان پردازشی و نرخ بیت (کیلوبیت بر ثانیه) و کیفیت نهایی برای ویدئوی Ballet

حالت اول - میانگین اختلاف بین نمایی کم															
نما	QP = 15, p = 2			QP = 14, p = 3			QP = 13, p = 4			QP = 12, p = 5			QP = 11, p = 6		
	PSNR	نرخ بیت	میانگین اختلاف بین-نمایی	PSNR	نرخ بیت	میانگین اختلاف بین-نمایی	PSNR	نرخ بیت	میانگین اختلاف بین-نمایی	PSNR	نرخ بیت	میانگین اختلاف بین-نمایی	PSNR	نرخ بیت	میانگین اختلاف بین-نمایی
۰	۴۳/۶۲	۲۵۲۲	۰/۱۲۵	۴۳/۹۲	۳۲۹۰	۰/۱۳۲	۴۴/۳۵	۴۴۴۹	۰/۱۲۶	۴۴/۸۷	۵۹۱۳	۰/۱۳	۴۵/۵۴	۷۸۹۷	۰/۰۹
۱	۴۳/۶۴	۲۲۸۵	۰/۱۲۸	۴۳/۹۳	۳۰۱۴	۰/۱۳۵	۴۴/۳۴	۴۱۴۸	۰/۱۲۶	۴۴/۸۵	۵۶۱۲	۰/۱۲۵	۴۵/۵۳	۷۷۲۴	۰/۰۹
۲	۴۳/۶۴	۲۳۶۳	۰/۱۲	۴۳/۹۴	۳۱۱۰	۰/۰۹۲	۴۴/۳۵	۴۲۶۰	۰/۰۹۳	۴۴/۸۶	۵۷۲۳	۰/۰۸۵			
۳	۴۳/۵۳	۲۴۳۶	۰/۱۴۸	۴۳/۸۴	۳۲۱۸	۰/۱۵	۴۴/۴۱	۴۲۲۵	۰/۱۲						
۴	۴۳/۷۱	۲۳۷۱	۰/۱۴	۴۴/۰۱	۳۱۰۲	۰/۱۴									
۵	۴۳/۶۱	۲۳۵۸	۰/۱۳۶												
حالت دوم - میانگین اختلاف بین نمایی زیاد															
نما	QP = 15, p = 2			QP = 14, p = 3			QP = 13, p = 4			QP = 12, p = 5			QP = 11, p = 6		
	PSNR	نرخ بیت	میانگین اختلاف بین-نمایی	PSNR	نرخ بیت	میانگین اختلاف بین-نمایی	PSNR	نرخ بیت	میانگین اختلاف بین-نمایی	PSNR	نرخ بیت	میانگین اختلاف بین-نمایی	PSNR	نرخ بیت	میانگین اختلاف بین-نمایی
۰	۴۳/۶۲	۲۵۲۲	۰/۱۴۴	۴۳/۹۲	۳۲۹۰	۰/۱۴۲	۴۴/۳۵	۴۴۴۹	۰/۱۵۳	۴۴/۸۷	۵۹۱۳	۰/۱۶	۴۵/۵۴	۷۸۹۷	۰/۱۷
۱	۴۳/۵۳	۲۴۴۵	۰/۱۳۲	۴۳/۸۴	۳۲۱۲	۰/۱۵۷	۴۴/۲۷	۴۴۲۳	۰/۱۶۳	۴۴/۸۱	۵۹۵۳	۰/۱۵۵	۴۵/۵۴	۷۷۶۸	۰/۱۷
۲	۴۳/۶۵	۲۳۹۲	۰/۱۶۲	۴۳/۹۴	۳۱۱۱	۰/۱۴۵	۴۴/۳۶	۴۲۶۲	۰/۱۷	۴۴/۸۷	۵۷۳۰	۰/۱۶۵			
۳	۴۳/۵۸	۲۲۸۹	۰/۱۸۶	۴۳/۸۷	۳۲۲۹	۰/۱۴۵	۴۴/۲۸	۴۴۷۵	۰/۱۵۶						
۴	۴۳/۵۴	۲۵۱۴	۰/۱۴۲	۴۳/۸۵	۳۲۹۴	۰/۱۷۵									
۵	۴۳/۶۴	۲۳۶۰	۰/۱۴۴												

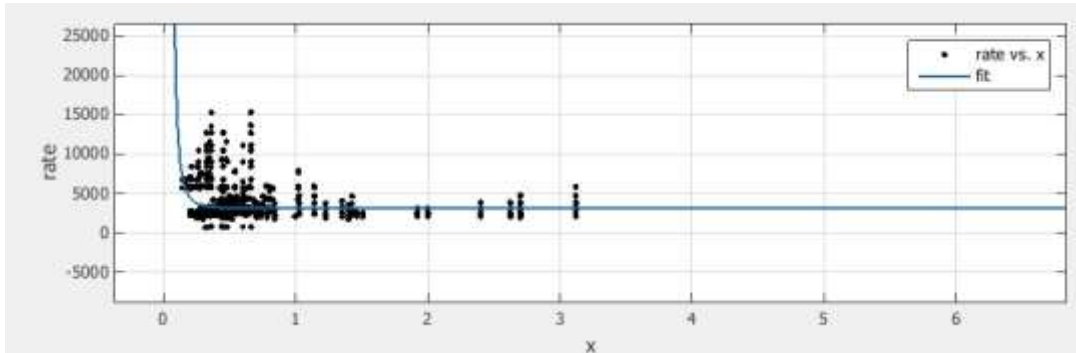
بعد از مشخص شدن همه متغیرها، ارتباط بین نرخ بیت نهایی هر نما و پارامترهای X و QP توسط برازش منحنی تعیین می‌شود. نتایج شکل ۵-۴ نشان می‌دهد که برای نقاط آزمایش- شده، ارتباط بین نرخ بیت با پارامترهای مذکور به صورت توانی است که در معادله زیر نیز نشان داده شده است:

$$R(QP) = d(QP^{-1})X^{e(QP^{-1})} + f(QP^{-1}) \quad (۲۱-۵)$$

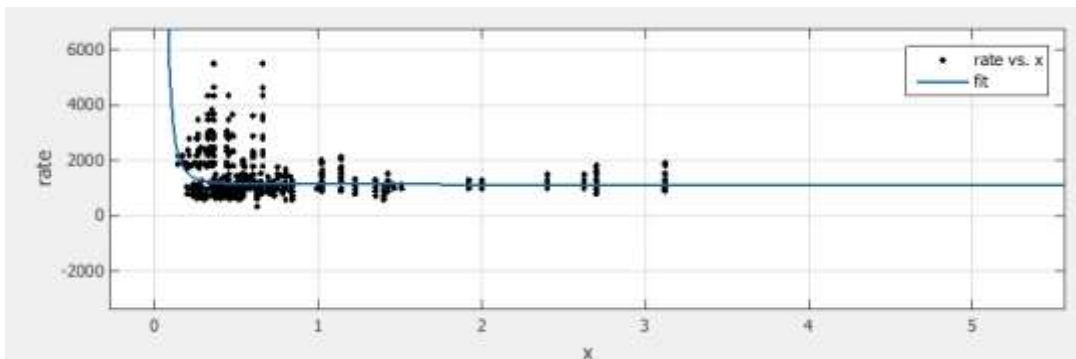
مقدار R-Square برای برازش منحنی و برای همه مقادیر آزمایش شده پارامتر کوانتیزاسیون برابر ۰/۷ است که نشان می‌دهد به‌طور میانگین ۷۰٪ تغییرات داده‌ها در منحنی ارائه شده توسط ابزار برازش منحنی نشان داده شده است.

برای هر مقدار پارامتر کوانتیزاسیون، مقداری برای ضرایب $d(QP)$ ، $e(QP)$ و $f(QP)$ پیدا می‌شود که ضرایب منحنی حاصل از برازش منحنی هستند. روابط مناسب برای یافتن این مقادیر نیز می‌توانند با یک برازش منحنی ساده بین مقادیر مختلف پارامتر کوانتیزاسیون و

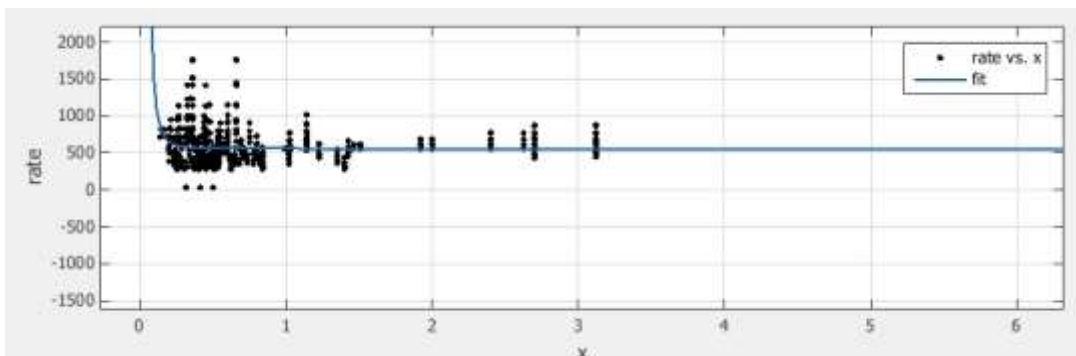
این ضرایب به دست آید. البته به منظور هماهنگی با مدل‌های نرخ بیت قبلی ارائه شده، برازش منحنی را بین مقادیر این ضرایب و عکس پارامتر کوانتیزاسیون انجام دادیم. نتایج در شکل ۵-۵ نشان داده شده است.



(الف)



(ب)



(ج)

شکل ۵-۴ - نرخ بیت واقعی ویدئوهای کدشده به صورت تابعی از پارامتر X برای (الف) $QP = 15$ (ب) $QP = 20$ (ج) $QP = 25$

مقادیر استخراج‌شده برای این ضرایب و به‌ازای مقادیر متفاوت پارامتر کوانتیزاسیون و نیز پارامترهای میانگین مربعات خطا و R-square نشان‌دهنده دقت فرآیند برازش منحنی در جدول ۵-۷ نشان داده شده است.

سپس نمودار مقادیر این ضرایب بر حسب عکس پارامتر کوانتیزاسیون استخراج شد. شکل ۵-۵ نمودارهای مربوط را نشان می‌دهد. معادلات استخراج شده برای این ضرایب نیز در معادلات (۲۲-۵)، (۲۳-۵) و (۲۴-۵) نشان داده شده‌اند. به این ترتیب برای مقادیر دیگر پارامتر کوانتیزاسیون نیز می‌توان مقادیر این ضرایب را از روی این معادلات به دست آورد. حال با استفاده از این ضرایب و با توجه به رابطه (۲۱-۵) می‌توان نرخ بیت هر نما را بر اساس پارامترهای میانگین اختلاف بین‌نمایی و توان پردازشی گیرنده به دست آورد.

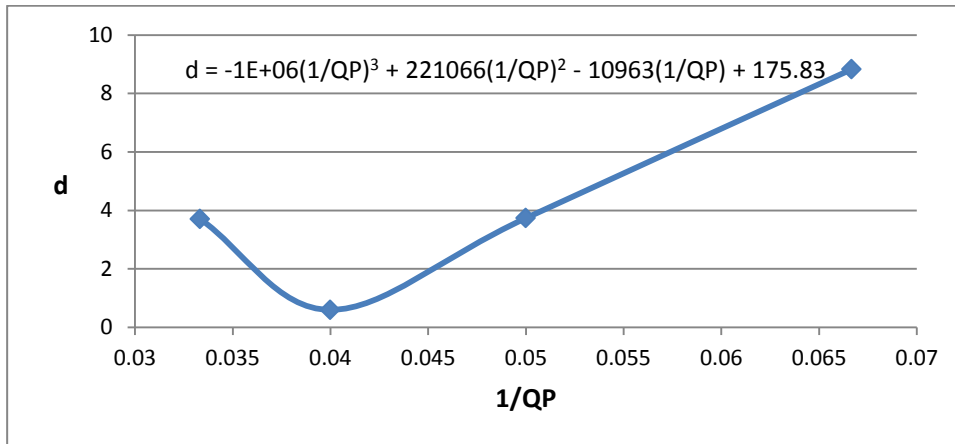
جدول ۷-۵ - مقادیر ضرایب استخراج شده d ، e و f رابطه (۲۱-۵) و پارامترهای میانگین مربعات خطا و R-square مربوط به فرآیند برازش منحنی

QP = 15				
d	e	f	RMSE	R-square
۸/۸۲۷	-۳/۱۱۱	۳۱۲۹	۴۰۷/۷	۰/۹۷۶
QP = 20				
d	e	f	RMSE	R-square
۳/۷۳۶	-۲/۹۰۶	۱۱۳۸	۱۲۵/۵	۰/۹۷۷۸
QP = 25				
d	e	f	RMSE	R-square
۰/۵۹۸۱	-۳/۱۲۶	۵۵۹/۵	۴۰/۰۹	۰/۹۷۵
QP = 30				
d	e	f	RMSE	R-square
۳/۷۰۵	۰/۸۱۲۲	۳۱۰/۴	۱۰۹/۷	۰/۹۷۵

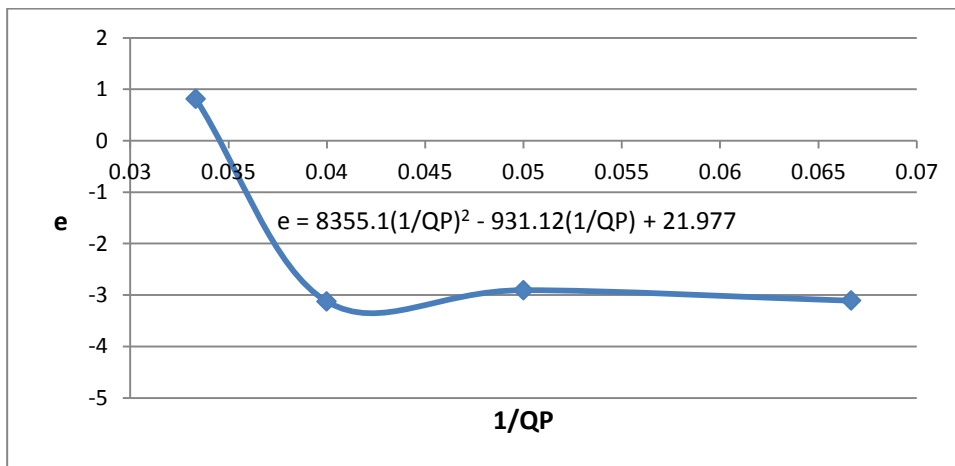
$$d(QP^{-1}) = -1E + 6(QP^{-1})^3 + 221066(QP^{-1})^2 - 10963(QP^{-1}) + 175.83 \quad (۲۲-۵)$$

$$e(QP^{-1}) = 8355(QP^{-1})^2 - 931.12(QP^{-1}) + 21.977 \quad (۲۳-۵)$$

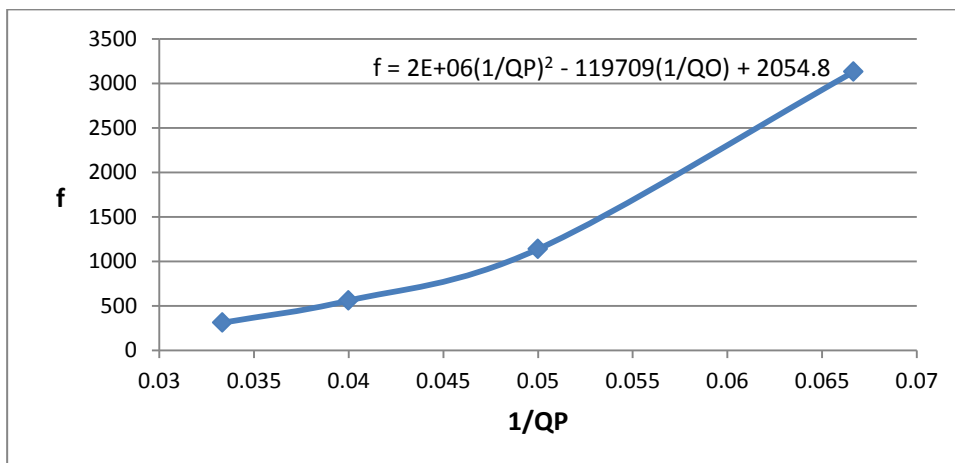
$$f(QP^{-1}) = 2E + 6(QP^{-1})^2 - 119709(QP^{-1}) + 2054.8 \quad (۲۴-۵)$$



(الف)



(ب)



(ج)

شکل ۵-۵ - ضرایب d , e و f رابطه (۵-۲۱) به صورت تابعی از عکس پارامتر کوانتیزاسیون

۵-۳-۳- مرحله سوم: مدل نهایی نرخ بیت - اعوجاج در سطح نما برای

ویدئوی چندنمایی

همان‌طور که قبلاً اشاره شد، در مرحله نهایی روش‌شناسی پیشنهادی، دو مدل نرخ بیت - اعوجاج استخراج‌شده در مراحل قبلی باید با یکدیگر ترکیب شوند تا مدل نهایی نرخ بیت - اعوجاج برای ویدئوی چندنمایی در سطح نما با استفاده از جمع وزن‌دار حاصل شود. مقادیر مناسب برای وزن‌ها نیز باید بر اساس نسبت تعداد پیش‌بینی‌های درون‌نمایی و بین‌نمایی تعیین شود.

برای استخراج مقادیر مناسب برای وزن‌ها، ۱۰۰ فریم از ویدئوهای آزمایش‌شده در جدول ۲-۵ و با ساختار پیش‌بینی شکل ۱-۵ کد شدند و تعداد پیش‌بینی‌های درون‌نمایی و بین‌نمایی برای فریم‌های مختلف نماهای متفاوت این ویدئوها استخراج شد. نتایج این آزمایش نشان می‌دهد که برای نماهایی که یک نمای مرجع برای پیش‌بینی بین‌نمایی دارند، به‌طور متوسط ۹۶ درصد از پیش‌بینی‌ها درون‌نمایی و تنها ۴ درصد پیش‌بینی‌ها بین‌نمایی است. برای نماهایی که دو نمای مرجع دارند، ۷۰ درصد پیش‌بینی‌ها درون‌نمایی و ۳۰ درصد آنها بین‌نمایی است. به این ترتیب رابطه نهایی مدل پیشنهادی نرخ بیت - اعوجاج برای ویدئوی چندنمایی به صورت زیر خواهد بود:

$$R(QP) = \omega_{\text{intra_pred}} [a(QP^{-1}) \times C + b(QP^{-1}) \times M + c(QP^{-1})] + \omega_{\text{inter_pred}} [d(QP^{-1})X^{e(QP^{-1})} + f(QP^{-1})] \quad (۲۵-۵)$$

در این معادله، X حاصل ضرب توان پردازشی گیرنده و میانگین اختلاف بین‌نمایی است که از معادله (۵-۱۷) به دست می‌آید. C پیچیدگی صحنه و M سطح حرکت، پارامترهای مدل-کننده پیچیدگی محتوا هستند که به ترتیب از معادلات (۵-۲) و (۵-۳) به دست می‌آیند. $\omega_{\text{intra_pred}}$ و $\omega_{\text{inter_pred}}$ به ترتیب وزن‌های پیش‌بینی بین‌نمایی و درون‌نمایی هستند. برای نماهایی که یک نمای مرجع دارند، مقادیر وزن‌ها به ترتیب ۰/۰۴ و ۰/۹۶ و برای نماهایی که دو نمای مرجع دارند، مقادیر وزن‌ها به ترتیب ۰/۳ و ۰/۷ خواهد بود. a, b, c, d, e, f ضرایب مدل هستند که به ترتیب از (۵-۱۲)، (۵-۱۳)، (۵-۱۴)، (۵-۲۲)، (۵-۲۳) و (۵-۲۴) تعیین می‌شوند و نحوه استخراج آنها در بخش‌های ۵-۱-۱-۲-۲-۵ و ۵-۱-۱-۲-۲-۵-۱-۱-۲-۲-۵ توضیح داده شده است. میزان حساسیت خطای پیش‌بینی به مقادیر وزن‌ها در جدول ۵-۸ نشان داده شده است. مقادیر این جدول نشان می‌دهد که برای مثال با تغییر مقادیر وزن‌ها به اندازه ۰/۱، مقدار خطای پیش‌بینی به‌طور میانگین ۱/۳۹ برابر می‌شود.

جدول ۵-۸- میزان حساسیت نرخ بیت پیش‌بینی‌شده به تغییرات وزن‌های مدل نرخ بیت - اعوجاج

۰/۴	۰/۳	۰/۲	۰/۱	میزان تغییر وزن‌های مدل نرخ بیت - اعوجاج
۲/۲۲	۱/۹۴	۱/۶۵	۱/۳۶	مقدار متوسط تغییرات نرخ بیت پیش‌بینی‌شده

همان‌طور که در این جدول دیده می‌شود، میزان حساسیت نرخ بیت به وزن‌ها به نسبت بالا است. استدلال ما در مورد وزن‌ها در مدل نهایی نرخ بیت اعوجاج این بود که برای نماهایی که تنها یک نمای مرجع دارند، حدود ۹۰ درصد پیش‌بینی‌ها درون‌نمایی و تنها ۱۰ درصد بین‌نمایی است. بر این اساس وزن مربوط به پیش‌بینی درون‌نمایی برای این نوع نماها ۰/۹ و وزن مربوط به پیش‌بینی بین‌نمایی ۰/۱ در نظر گرفته شد. در مورد نماهایی که دو نمای مرجع دارند بر اساس همین استدلال وزن مربوط به پیش‌بینی بین‌نمایی ۰/۳ و وزن مربوط به پیش‌بینی درون‌نمایی ۰/۷ در نظر گرفته شد. بر این اساس تغییر وزن‌ها به معنی تغییر نسبت پیش‌بینی‌های بین‌نمایی و درون‌نمایی است. بنابراین طبیعی است که نرخ بیت تخمین زده شده و خطای پیش‌بینی به مقادیر وزن‌ها حساسیت زیادی داشته باشد. از آنجایی که پیش‌بینی درون‌نمایی و بین‌نمایی بر اساس بهینه‌سازی نرخ بیت - اعوجاج انتخاب می‌شوند، نسبت پیش‌بینی‌ها نشان می‌دهد که از لحاظ کارایی فشرده‌سازی پیش‌بینی‌های درون‌نمایی مهم‌تر هستند. در نتیجه اگر ما با تغییر مقادیر وزن‌ها در مدل پیشنهادی وزن بیشتری به پیش‌بینی بین‌نمایی نسبت به درون‌نمایی بدهیم طبیعی است که نرخ بیت تخمینی به نرخ بیت واقعی نزدیک نباشد و خطای بالایی داشته باشد. بنابراین میزان بالای حساسیت نرخ بیت به وزن‌ها طبیعی است.

۵-۴- ارزیابی مدل پیشنهادی

به منظور ارزیابی مدل پیشنهادی، تعداد زیادی نما از ویدئوهای چندنمایی و سه-بعدی با پیچیدگی و وضوح تصویر متفاوت انتخاب شد. این ویدئوها با ویدئوهایی که برای استخراج مدل پیشنهادی استفاده شده‌اند، متفاوت است. مشخصات این ویدئوها در جدول ۵-۹ نشان داده شده است. نکته قابل ذکر اینکه در همه ویدئوهای آزمایش‌شده در این جدول، دوربین‌ها ثابت هستند.

جدول ۵-۹ - مشخصات ویدئوهای کد شده برای ارزیابی مدل نرخ بیت - اعوجاج پیشنهادی

تعداد فریم	تعداد نما	نرخ فریم (فریم بر ثانیه)	اندازه فریم	ویدئو
۲۵۰	۷	۱۵	۶۴۰×۴۸۰	Ballroom
۲۵۰	۷	۱۵	۶۴۰×۴۸۰	Exit
۵۰۰	۷	۱۵	۱۲۸۰×۹۶۰	Pantomim
۱۰۰	۵	۱۵	۱۰۲۴×۷۶۸	Book Arrival

سپس این ویدئوهای با کدکننده استاندارد H.264/MVC نسخه ۸/۵ [105] کد شدند و نرخ بیت واقعی حاصل از کدشدن استخراج شد. سپس با استفاده از مدل پیشنهادی ارائه شده در معادله (۵-۲۵)، نرخ بیت تخمین زده شده توسط این مدل نیز به دست آمد. در نهایت، مقادیر واقعی و تخمین زده شده نرخ بیت با هم مقایسه شدند. نماهای انتخاب شده با مقادیر متفاوت پارامتر کوانتیزاسیون (در محدوده ۱۵ تا ۳۰) و توان پردازشی (در محدوده ۲ تا ۶) کد شدند. جدول ۵-۱۰ میانگین خطای تخمین را برای ویدئوهای آزمایش شده نشان می دهد. خطای تخمین از رابطه زیر به دست می آید:

$$A. E. E = \text{mean} \left(\frac{100 \times \text{abs}(\text{Real Bits} - \text{Estimated Bits})}{\text{Real bits}} \right) \quad (۵-۲۶)$$

نتایج جدول ۵-۱۰ نشان می دهد که میانگین خطای تخمین برای نماهای مختلف ویدئوهای آزمایش شده با مقادیر متفاوت پارامتر کوانتیزاسیون ۱۲٪ است. در این جدول مقدار P برابر ۲ در نظر گرفته شده است، به این معنی که کدکننده قادر است دو نمای مجازی را باز تولید کند. بنابراین تعداد نماهایی که ارسال می شوند ۵ نما است. تعداد بالای نماهای ارسالی باعث شده که در مقادیر بالای پارامتر کوانتیزاسیون (QP = 30) و نرخ بیت های پایین، مقدار خطای تخمین مدل پیشنهادی زیاد باشد.

جدول ۵-۱۱ مقدار میانگین خطای تخمین را برای مدل استخراج شده از روش پیشنهادی در مقایسه با مدل نرخ بیت - اعوجاج خطی [111] برای ۴ نما از ویدئوهای آزمایش شده نشان می دهد. همان طور که در این جدول می بینیم، مدل پیشنهادی ما می تواند نرخ بیت واقعی هر نما را با دقت بالاتری تخمین بزند.

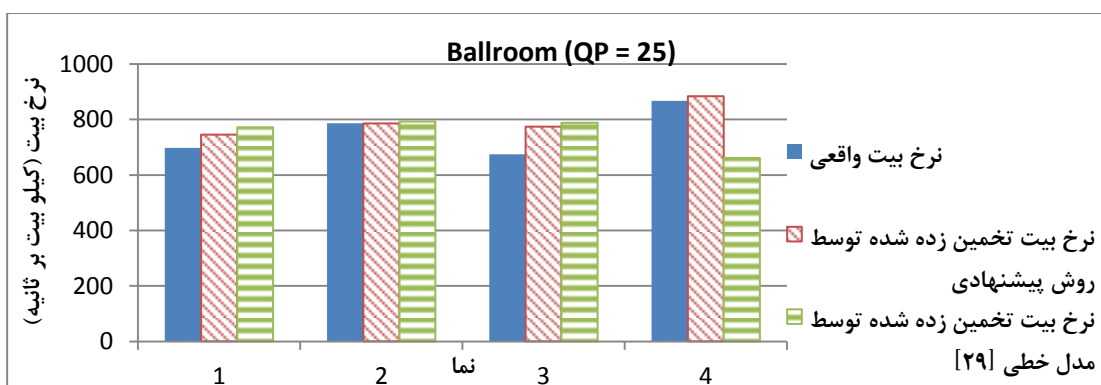
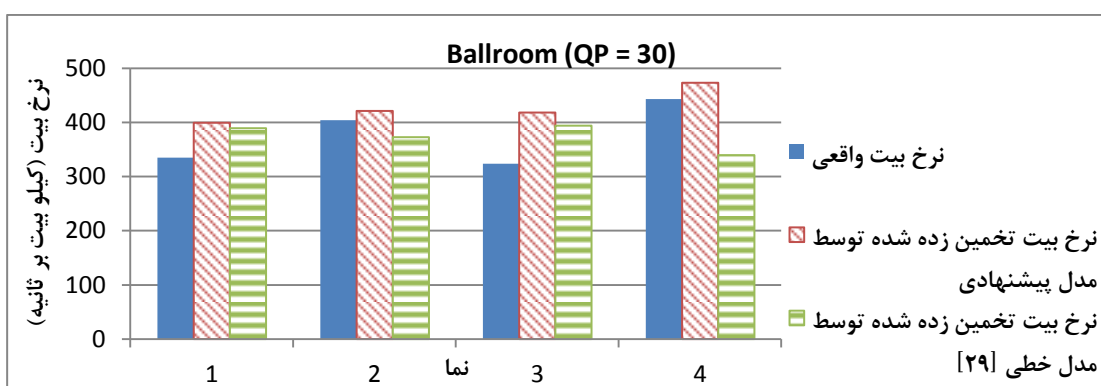
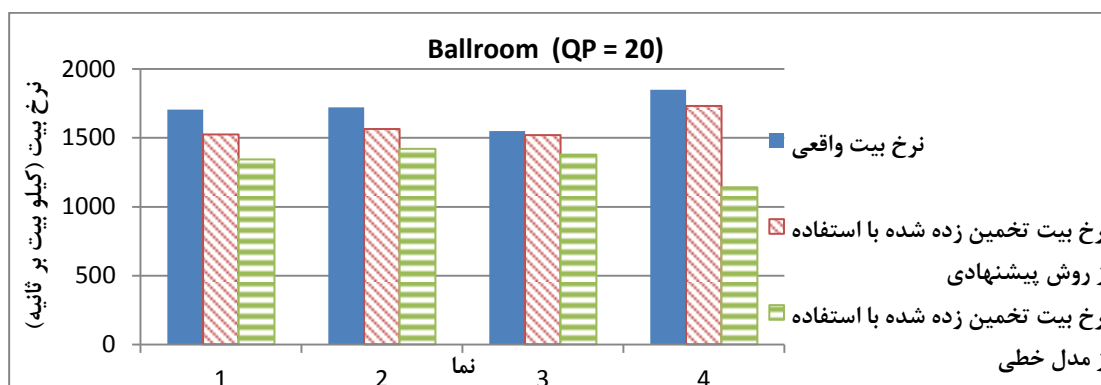
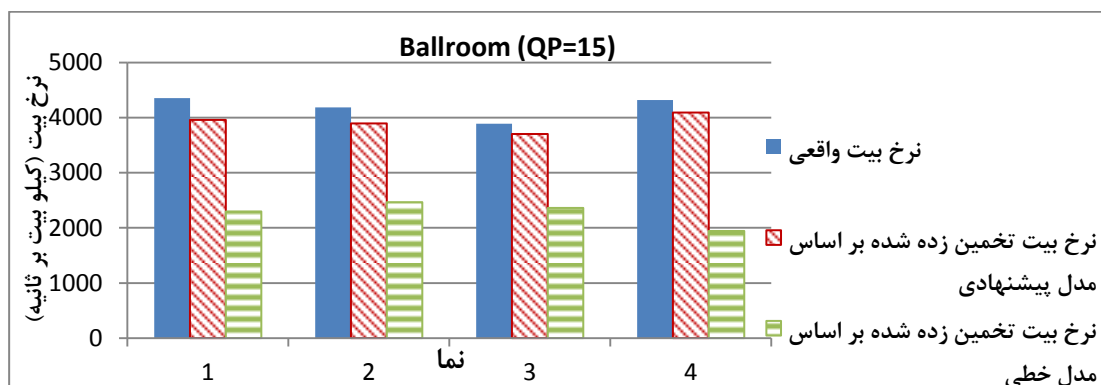
علاوه بر این، برای وضوح بیشتر، مقادیر نرخ بیت واقعی و تخمین زده شده توسط مدل پیشنهادی نرخ بیت - اعوجاج و مدل خطی برای مقادیر مختلف پارامتر کوانتیزاسیون برای ویدئوی Ballroom در شکل ۵-۶ نشان داده شده است.

جدول ۵-۱۰ - میانگین خطای تخمین مدل نرخ بیت - اعوجاج پیشنهادی برای نماهای مختلف ویدئوهای آزمایش -
شده و برای مقادیر مختلف پارامتر کوانتیزاسیون

Ballroom					
پارامتر کوانتیزاسیون	نمای ۱	نمای ۲	نمای ۳	نمای ۴	میانگین خطای تخمین
	خطای تخمین	خطای تخمین	خطای تخمین	خطای تخمین	
۱۵	%۹/۱	%۶/۹	%۴/۷	%۵/۲	%۶/۵
۲۰	%۱۰/۶	%۹/۱	%۲	%۶/۴	%۷
۲۵	%۶/۷	%۰/۲	%۱۴/۷	%۱/۹	%۵/۹
۳۰	%۱۹/۱	%۴/۳	%۲۹/۱	%۶/۸	%۱۴/۸
Exit					
پارامتر کوانتیزاسیون	نمای ۱	نمای ۲	نمای ۳	نمای ۴	میانگین خطای تخمین
	خطای تخمین	خطای تخمین	خطای تخمین	خطای تخمین	
۱۵	%۱۱/۴	%۱۳/۵	%۱۲/۲	%۱۱/۵	%۱۲/۱
۲۰	%۱۳	%۲۱/۲	%۱۵/۲	%۱۵/۴	%۱۶/۲
۲۵	%۲۵/۶	%۷	۹	%۵/۳	%۱۱/۷
۳۰	%۶۱	%۵/۸	%۳/۵	%۲۳/۱	%۲۳/۴
Pantomim					
پارامتر کوانتیزاسیون	نمای ۱	نمای ۲	نمای ۳	نمای ۴	میانگین خطای تخمین
	خطای تخمین	خطای تخمین	خطای تخمین	خطای تخمین	
۱۵	%۱۵/۷	%۷/۲	%۱۰/۷	%۶/۹	%۱۰/۱
۲۰	%۱۴/۳	%۳/۹	%۳/۳	%۴	%۶/۱
۲۵	%۱۹/۴	%۵/۵	%۱۶/۳	%۵/۸	%۱۱/۸
۳۰	%۲۱/۵	%۱۰/۴	%۳۸	%۱۰/۷	%۲۰/۱

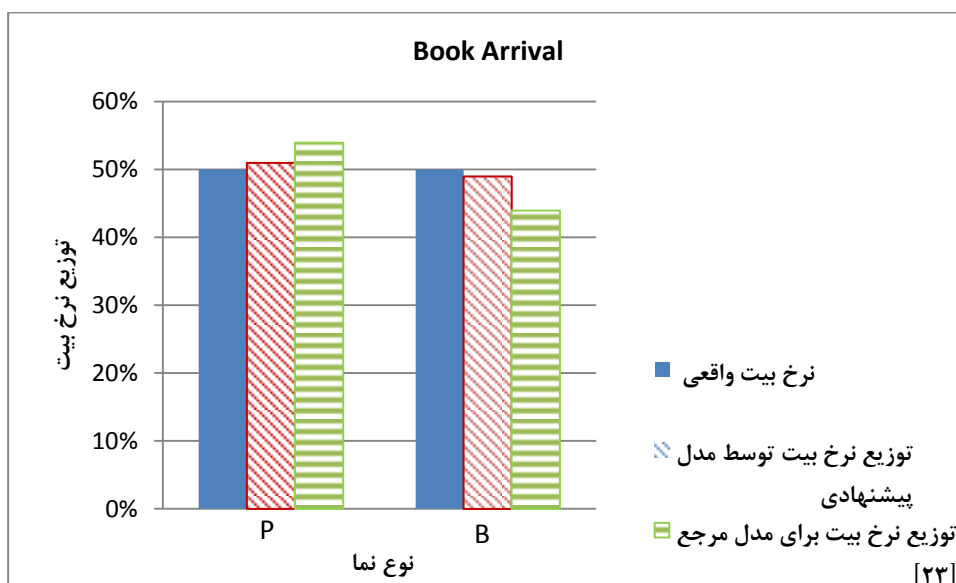
جدول ۵-۱۱ - مقایسه میانگین خطای تخمین برای مدل نرخ بیت - اعوجاج پیشنهادی و مدل نرخ بیت - اعوجاج خطی [111] برای ۴ نما از ویدئوهای آزمایش شده

میانگین خطای تخمین	پارامتر کوانتیزاسیون	ویدئوی آزمایش شده
مدل خطی	روش پیشنهادی	
%۴۲	%۶	۱۵
%۲۲	%۷	۲۰
%۱۲	%۵	۲۵
%۱۷	%۱۴	۳۰
%۲۷	%۱۲	۱۵
%۲۸	%۱۶	۲۰
%۵۰	%۱۳	۲۵
%۶۰	%۲۳	۳۰
%۶۲	%۱۱	۱۵
%۴۵	%۶	۲۰
%۳۰	%۱۳	۲۵
%۳۲	%۲۳	۳۰



شکل ۵-۶ - مقایسه مقادیر نرخ بیت واقعی و تخمین زده شده برای نماهای مختلف ویدئوی Ballroom با پارامترهای کوانتیزاسیون متفاوت

علاوه بر این، کارایی مدل نرخ بیت - اعوجاج پیشنهادی با یک مدل تجربی دیگر در سطح نما که بر مبنای توزیع حالات پیش‌بینی مورد استفاده در انواع مختلف نماها است [25]، نیز مقایسه شده است. شکل ۵-۷ کارایی مدل پیشنهادی ارائه‌شده در این رساله را در مقایسه با این مدل برای ویدئوی آزمایش‌شده Book Arrival و برای محدوده وسیعی از پارامترهای کوانتیزاسیون (در محدوده ۱۵ تا ۳۸) نشان می‌دهد. اعداد این شکل درصد توزیع نرخ بیت برای انواع مختلف نما نسبت به نرخ بیت کل را نشان می‌دهد. همان‌طور که در این شکل دیده می‌شود مدل پیشنهادی در این رساله توزیع نرخ بیت را هم برای نماهای نوع P و هم برای نماهای نوع B با دقت بالاتری پیش‌بینی می‌کند.



شکل ۵-۷ - مقایسه توزیع نرخ بیت برای انواع مختلف نما نسبت به نرخ بیت کل برای مدل پیشنهادی و مدل ارائه‌شده در [25]

یکی از کاربردهای مهم مدل نرخ بیت - اعوجاج پیشنهادشده در این رساله می‌تواند در کد کردن یکی از فرمت‌های خاص ویدئوی سه‌بعدی با عنوان ویدئو به‌علاوه عمق^۱ باشد. در این فرمت ویدئو اطلاعات بافت و عمق در یک رشته‌بیت واحد قرار می‌گیرند [112]. فرمت ویدئو به‌علاوه عمق از این جهت که قابلیت بازتولید نماهای مجازی را فراهم می‌کند، در کاربردهای نمای آزاد ویدئوی سه‌بعدی بسیار مورد توجه است.

در این فرمت ویدئویی، فشرده‌سازی بر اساس همان الگوریتم‌های کد کردن ویدئوی چندنمایی انجام می‌شود که وابستگی‌های بین فریم‌های متوالی از نظر زمانی و بین‌نمایی را

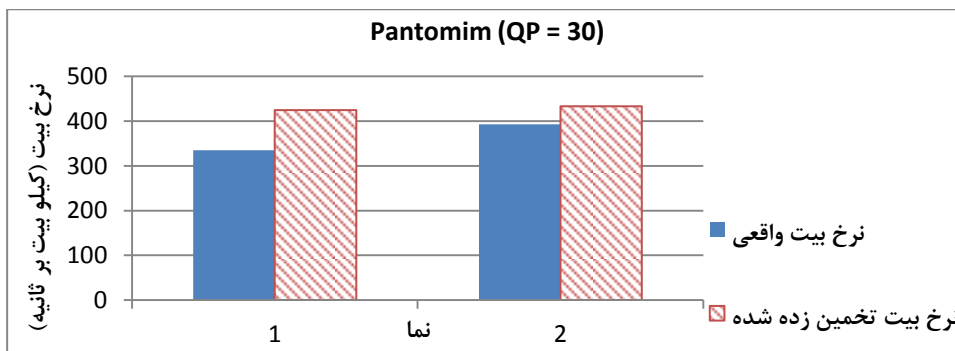
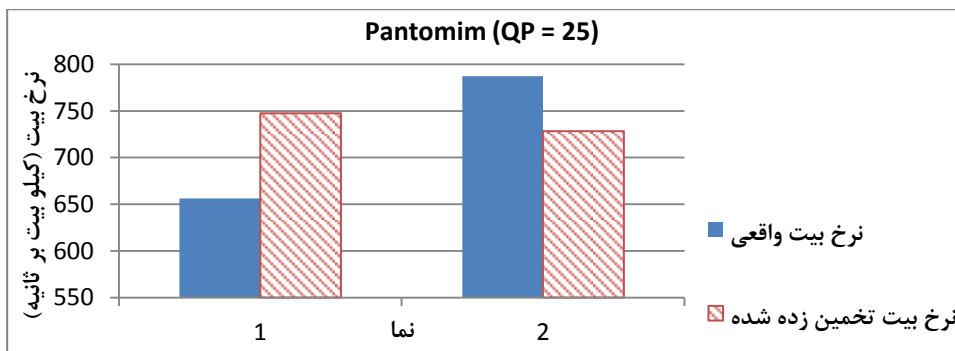
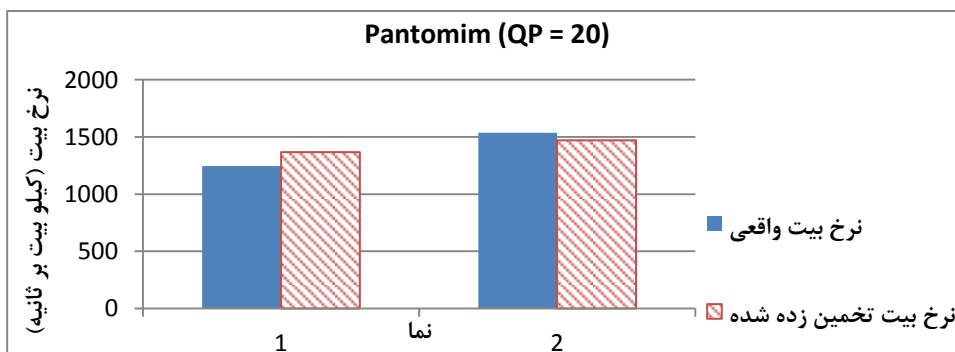
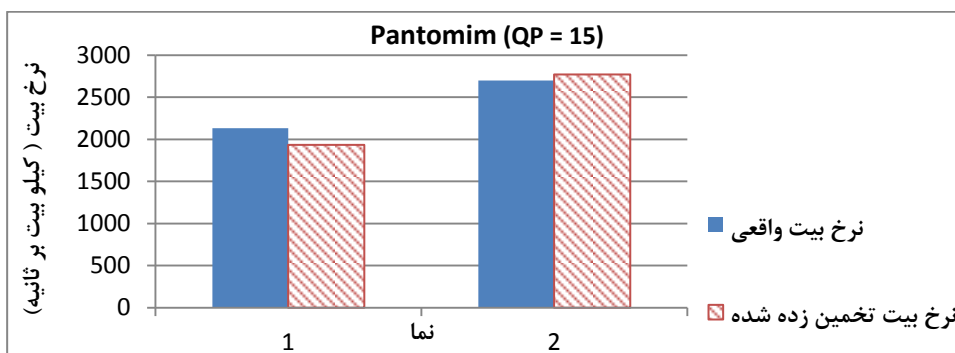
^۱ Multi-view video plus depth

برای پیش‌بینی بهتر اطلاعات عمق و بافت استخراج می‌کند [113]. به منظور ارزیابی کارایی مدل نرخ بیت - اعوجاج برای ویدئوی سه‌بعدی، این فرمت ویدئویی به‌عنوان یکی از کاربردهای مهم ویدئوی چندنمایی سه‌بعدی مورد بررسی قرار گرفته است. برای این منظور آزمایشی ترتیب دادیم که در آن نماهای عمق از روی نماهای عمق مجاور به‌عنوان مرجع پیش‌بینی شوند. ابتدا، پارامترهای مدل مانند میانگین اختلاف بین‌نمایی، پیچیدگی محتوای ویدئو و توان پردازشی برای ویدئوی عمق استخراج شد. سپس نرخ بیت تخمینی با استفاده از معادله (۵-۲۵) محاسبه شد و در نهایت این مقدار با نرخ بیت واقعی حاصل از کدکردن ویدئو مقایسه و خطای تخمین با استفاده از معادله (۵-۲۶) استخراج شد. نتایج این آزمایش برای ویدئوی Pantomim و برای مقادیر متفاوت پارامتر کوانتیزاسیون در جدول ۵-۱۲ نشان داده شده است. علاوه بر این، خطای تخمین مدل نرخ بیت - اعوجاج پیشنهادی برای نماهای عمق ویدئوی Pantomim در شکل ۵-۸ نشان داده شده است. میانگین خطای تخمین مدل پیشنهادی برای نماهای متفاوت این ویدئو برابر ۱۰٪ است.

جدول ۵-۱۲ - میانگین خطای تخمین برای نماهای مختلف عمق برای ویدئوی Pantomim

پارامتر کوانتیزاسیون	نمای ۱	نمای ۲	میانگین خطای تخمین
	خطای تخمین	خطای تخمین	
۱۵	٪۹	٪۲	٪۵/۵
۲۰	٪۹	٪۴	٪۶/۵
۲۵	٪۱۳	٪۷	٪۱۰
۳۰	٪۲۶	٪۱۰	٪۱۸

در مجموع نتایج شبیه‌سازی‌ها نشان می‌دهد که مدل نرخ بیت - اعوجاج استخراج شده از روش‌شناسی پیشنهادی می‌تواند نرخ بیت هر نما را با خطای تخمین پایین ۱۲٪ برای ویدئوهای بافت و ۱۰٪ برای ویدئوهای عمق تخمین بزند.



شکل ۵-۸ - مقایسه مقادیر نرخ بیت واقعی و تخمین زده شده برای نماهای عمق ویدئوی Pantomim با پارامترهای کوانتیزاسیون متفاوت

۵-۵- جمع بندی

در این بخش از رساله، یک روش‌شناسی به منظور استخراج مدل نرخ بیت - اعوجاج مناسب در سطح نما برای ویدئوی چندنمایی سه‌بعدی پیشنهاد شده است که ویژگی‌های خاص کدکردن ویدئوی چندنمایی مانند وابستگی‌های درون‌نمایی و بین‌نمایی و اثر آنها روی کارایی کدکردن را در نظر می‌گیرد. به این منظور ابتدا پارامترهای مؤثر برای توصیف این مفاهیم استخراج شده و سپس رابطه نرخ بیت هر نما با این پارامترها به دست آمده است. نتایج شبیه‌سازی‌ها نشان می‌دهد که مدل نرخ بیت - اعوجاج استخراج شده از روش‌شناسی پیشنهادی ما می‌تواند نرخ بیت هر نما از ویدئوی چندنمایی را به طور متوسط با دقت ۱۱٪ تخمین بزند.

فصل شش

روش‌شناسی ارائه‌شده برای استخراج معیار ارزیابی کمی

کیفیت درویدنومی چندنمائی مقیاس پذیر

در این بخش از رساله، روش‌شناسی پیشنهادی برای استخراج معیار اندازه‌گیری کمی کیفیت برای ویدئوی مقیاس‌پذیر چندنمایی معرفی می‌شود. در ابتدا، چند تعریف و نیز مشاهداتی که برای استخراج این روش‌شناسی مورد استفاده قرار گرفته است، در بخش ۶-۱-۱ معرفی می‌شود. روش‌شناسی پیشنهادی در بخش ۶-۱-۲ و زیربخش‌های آن توضیح داده می‌شود.

۶-۱- روش‌شناسی پیشنهادی

۶-۱-۱- تعاریف و مشاهدات

همان‌طور که قبلاً اشاره شد، رشته‌بیت ویدئویی مقیاس‌پذیر از نمونه‌های کوچک‌تر ویدئوی اصلی با وضوح تصویر، نرخ فریم، کیفیت یا تعداد نماهای متفاوت تشکیل شده است. بنابراین کیفیت نهایی هر لایه از روی کیفیت این واحدهای کوچک‌تر تعیین می‌شود. برای مثال، می‌توان گفت که کیفیت نهایی دریافتی یک لایه در مقیاس‌پذیری نمایی از مجموع کیفیت نماهای موجود در آن لایه به دست می‌آید. به منظور تعمیم‌دادن روش ارزیابی کیفیت ارائه‌شده به سایر مقیاس‌پذیری‌ها، در این رساله، نمونه‌های کوچک‌تر از ویدئوی اصلی در هر لایه را "واحد مرتبه پایین"^۱ می‌نامیم. این واحدهای مرتبه پایین بر اساس نوع خاص مقیاس‌پذیری تعریف می‌شوند. در ادامه این تعریف جدید را با ذکر مثال توضیح می‌دهیم.

مقیاس‌پذیری نمایی یکی از پرکاربردترین انواع مقیاس‌پذیری در ویدئوی چندنمایی و سه-بعدی است. در این حالت، رشته ویدئوی مقیاس‌پذیر یک لایه پایه دارد که از حداقل تعداد نماهایی تشکیل شده که گیرنده‌ها می‌توانند دریافت کنند. هر گیرنده بر حسب قابلیت‌های پهنای باند و توان پردازشی خود می‌تواند یک یا چند لایه افزایشی نیز دریافت کرده و محدوده بیشتری از صحنه تصویربرداری شده را پوشش دهد. در این نوع مقیاس‌پذیری، هر لایه خود از یک ویدئوی چندنمایی / سه‌بعدی تشکیل شده است. در نتیجه، برای مقیاس‌پذیری نمایی، واحدهای مرتبه پایین، نماهای هر لایه هستند.

در مقیاس‌پذیری عمق، هر لایه از قسمت‌های مختلف نماهای مختلف با یک سطح خاص عمق تشکیل شده است. این قسمت‌ها واحدهای مرتبه پایین را در این نوع خاص مقیاس‌پذیری تشکیل می‌دهند. شایان ذکر است که در این نوع مقیاس‌پذیری، ویدئوی اصلی به صورت ویدئوی چندنمایی به اضافه عمق در نظر گرفته می‌شود.

^۱ Lower Order Unit (LOU)

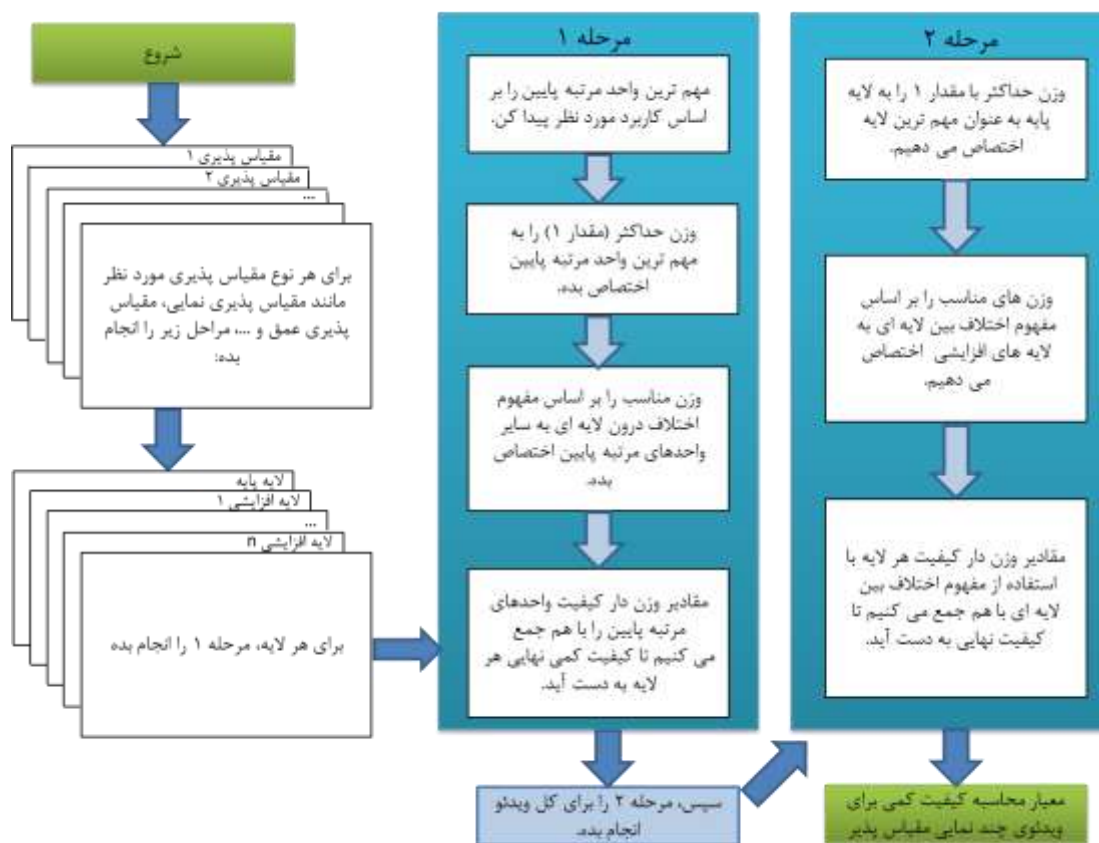
در بعضی کاربردها، واحدهای مرتبه پایین مختلف اثر یکسانی روی کیفیت نهایی ویدئوی چندنمایی مقیاس‌پذیر ندارند و ممکن است بعضی از آنها برای کاربران اهمیت بیشتری داشته باشند. برای مثال، در بعضی کاربردها، نمای وسط ممکن است برای کاربران مهم‌تر باشد، بنابراین اثر کیفیت آن در ارزیابی کیفیت نهایی از اثر سایر نماها بیشتر خواهد بود. به‌عنوان مثالی دیگر، در ویدئو کنفرانس چندنمایی، ارتباط چهره به چهره نقش مهمی در القای تجربه ارتباط واقعی دارد [114]. بنابراین، کیفیت نمای مرتبط به دوربینی که از چهره گوینده تصویربرداری می‌کند، نسبت به سایر نماها اهمیت بیشتری دارد. به‌طور مشابه، از آنجایی که یک ویدئوی چندنمایی مقیاس‌پذیر از لایه‌های پایه و افزایشی تشکیل شده است، کیفیت نهایی آن از روی کیفیت لایه‌های آن تعیین می‌شود. از آنجایی که لایه‌های افزایشی تنها در صورت وجود لایه پایه می‌توانند واکنش دهند، کیفیت لایه پایه نسبت به کیفیت لایه‌های افزایشی اهمیت بیشتری در کیفیت نهایی خواهد داشت. علاوه بر این، اثر لایه‌های افزایشی مختلف در کیفیت نهایی نیز در کاربردهای مختلف یکسان نخواهد بود.

۶-۱-۲ - مراحل روش‌شناسی پیشنهادی

بر اساس مشاهدات بالا، روش استخراج معیار مناسب برای ارزیابی کیفیت ویدئوی چندنمایی مقیاس‌پذیر باید دو مرحله داشته باشد. در گام نخست، کیفیت هر لایه باید به صورت مستقل تعیین شود. در این مرحله و با در نظر گرفتن اثر متفاوت واحدهای مرتبه پایین هر لایه در کیفیت نهایی آن لایه، ابتدا باید مهم‌ترین واحد مرتبه پایین در هر لایه، بر اساس کاربرد مورد نظر تعیین شود. وزن کیفیت مهم‌ترین واحد مرتبه پایین در ارزیابی کیفیت نهایی حداکثر وزن ممکن خواهد بود. وزن سایر واحدهای مرتبه پایین بر اساس میزان اهمیت آنها نسبت به مهم‌ترین واحد مرتبه پایین در نظر گرفته می‌شود. در این رساله، مفهوم "اختلاف درون لایه‌ای" برای مشخص کردن ارتباط بین واحدهای مرتبه پایین و تعیین وزن مناسب برای آنها معرفی شده است. اختلاف درون لایه‌ای به صورت اختلاف بین مهم‌ترین واحد مرتبه پایین و سایر واحدها و بر اساس انواع مختلف مقیاس‌پذیری تعریف می‌شود. به عبارت دیگر، برای هر نوع مقیاس‌پذیری، خصوصیات و ویژگی‌های آن به منظور تعیین تعریف مناسب برای اختلاف درون لایه‌ای برای آن مقیاس‌پذیری خاص مورد استفاده قرار می‌گیرد.

در گام دوم، مقادیر کیفیت کمی لایه‌های مختلف با یکدیگر ترکیب شده و کیفیت نهایی کل ویدئوی مقیاس‌پذیر تعیین می‌شود. این کار با اختصاص دادن حداکثر وزن ممکن به کیفیت

لایه پایه به عنوان مهم ترین لایه موجود انجام می شود. سپس وزن مناسب برای لایه های افزایشی بر اساس میزان اهمیت هریک از آنها نسبت به لایه پایه و میزان اهمیت آنها برای هر گیرنده تعیین می شود. اهمیت هر یک از لایه های افزایشی نیز بر اساس نوع مقیاس پذیر مورد نظر تعیین می شود. در روش شناسی پیشنهاد شده، مفهوم "اختلاف بین لایه ای" به منظور تعیین اهمیت هر یک از لایه های افزایشی در کیفیت نهایی و تعیین وزن مناسب برای کیفیت آنها معرفی شده است. این مفهوم برای ترکیب کردن کیفیت لایه های مختلف به منظور به دست آوردن یک معیار واحد برای کل ویدئوی چندنمایی مقیاس پذیر مورد استفاده قرار می گیرد. این مفهوم به صورت میانگین اختلاف بین واحدهای مرتبه پایین لایه پایه و لایه های افزایشی تعریف می شود. همانند مراحل قبلی، ویژگی های خاص هر نوع مقیاس پذیری به منظور تعریف مفهوم اختلاف بین لایه ای در آن نوع خاص مقیاس پذیری استفاده می شود. روش شناسی ارائه شده در شکل ۱-۶ نشان داده شده است. دو مرحله پیشنهادی برای این روش شناسی به ترتیب در زیرفصل های ۱-۶-۱ و ۱-۶-۲-۱-۲ توضیح داده خواهند شد.



شکل ۱-۶. مراحل مختلف روش شناسی ارائه شده برای استخراج معیار ارزیابی کمی کیفیت برای ویدئوی چندنمایی مقیاس پذیر

۶-۱-۲-۱- ارزیابی کیفیت برای هر لایه در ویدئوی چندنمایی مقیاس‌پذیر

همان‌طور که در بخش ۶-۱-۱-۱ توضیح داده شد، هر لایه از ویدئوی چندنمایی مقیاس‌پذیر از چندین واحد مرتبه پایین تشکیل شده است. یکی از ساده‌ترین روش‌های ارزیابی کیفیت هر لایه جمع کردن معیارهای کیفیت این واحدهای مرتبه پایین با یکدیگر است. برای مثال، در مقیاس‌پذیری نمایی، می‌توان معیار کیفیت نماهای مختلف هر لایه را با یکدیگر جمع کرد.

اگرچه این روش پیچیدگی محاسباتی کمی دارد و تعداد واحدهای مرتبه پایین را نیز در میزان کیفیت نهایی نظر می‌گیرد، اما در بسیاری از کاربردها گمراه‌کننده خواهد بود. برای مثال، فرض کنید یک ویدئو با دو نمای مختلف داریم که میانگین مربعات خطای آنها به ترتیب ۱۰۰ و ۴۰ است. حال ویدئوی دیگری را با دو نما و میانگین مربعات خطای ۷۰ و ۷۰ در نظر بگیرید. با جمع کردن اعوجاج دو نمای این دو ویدئو، میانگین مربعات خطای نهایی هر دو ویدئو یکسان خواهد بود، که قطعاً گمراه‌کننده است، زیرا در حالت اول، تقریباً تمام جزئیات نمای اول (نمایی که میانگین مربعات خطای آن ۱۰۰ است) از بین می‌رود ولی این موضوع در کیفیت نهایی بازتاب داده نمی‌شود. این نقص به‌خصوص در کاربردهایی مانند نمای آزاد بسیار مهم خواهد بود. بنابراین، جمع ساده معیار کیفیت واحدهای مرتبه پایین، اثر کیفیت هر واحد را در کیفیت نهایی به‌خوبی بازتاب نمی‌دهد.

در این حالت، استفاده از جمع وزن‌دار می‌تواند راه‌حل بهتری باشد. مقدار وزن هر واحد مرتبه پایین نیز می‌تواند بر حسب میزان اهمیت آن واحد در کیفیت نهایی دریافتی توسط کاربر یا بر اساس کاربرد مورد نظر تعیین شود. کارایی این روش به میزان قابل توجهی به مقادیر انتخاب‌شده برای وزن‌های واحدهای مرتبه پایین و قابلیت تغییر دادن آنها در موقعیت‌های متفاوت بستگی خواهد داشت. ویژگی‌های متفاوتی می‌تواند برای انتخاب مقادیر مناسب برای وزن‌ها در نظر گرفته شود، ولی بدون شک استفاده از ویژگی‌های خاص ویدئوی چندنمایی مقیاس‌پذیر برای انتخاب آنها به تخمین مناسب‌تر میزان کیفیت نهایی ویدئوی چندنمایی مقیاس‌پذیر منجر خواهد شد.

در روش‌شناسی پیشنهادی در این رساله، از جمع وزن‌دار کیفیت واحدهای مرتبه پایین مختلف به عنوان روشی برای ارزیابی کیفیت نهایی هر لایه استفاده می‌شود. سپس پیشنهاد شده است که از مفهوم اختلاف درون‌لایه‌ای بین واحدهای مرتبه پایین به عنوان یک خصوصیات ذاتی ویدئوی چندنمایی مقیاس‌پذیر برای تعیین مقادیر مناسب برای وزن‌ها استفاده شود. همان‌طور که پیش‌تر اشاره شد، اختلاف درون‌لایه‌ای باید بر اساس ویژگی‌های

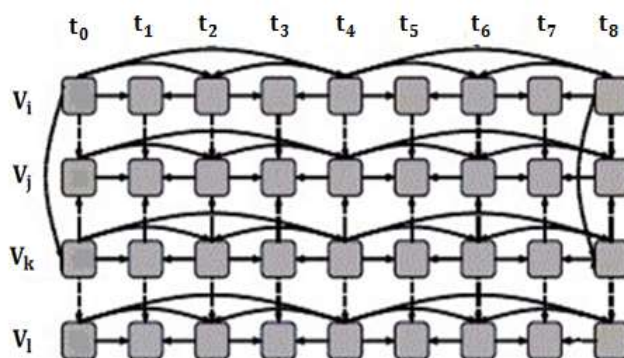
خاص هر نوع مقیاس پذیری تعریف شود. همان طور که در بخش ۶-۱-۱- رساله توضیح داده شد، در این رساله این مفهوم برای دو نوع مقیاس پذیری نمایی و عمق تعریف شده است. به منظور تعمیم دادن این مفهوم به سایر مقیاس پذیری‌ها، کافی است که ویژگی اصلی که برای استخراج لایه‌ها مورد استفاده قرار گرفته را در نظر بگیریم. با استفاده از این ویژگی می‌توان مفهوم اختلاف درون لایه‌ای را نیز تعریف کرد. برای مثال، در مقیاس پذیری کیفیت، فریم‌هایی با کیفیت متفاوت برای لایه‌های مختلف تولید می‌شوند. این موضوع دلالت بر این دارد که مفهوم "کیفیت" به صورت نسبت سیگنال به نویز به عنوان ویژگی اصلی برای تولید لایه‌های مختلف در نظر گرفته می‌شود. بنابراین، اختلاف بین سطوح کیفیت از نظر نسبت سیگنال به نویز برای فریم‌های مختلف می‌تواند به عنوان مفهوم "اختلاف" و برای محاسبه اختلاف درون نمایی مورد استفاده قرار گیرد. به عنوان مثالی دیگر، در مقیاس پذیری مکانی، وضوح تصویر فریم‌های تولید شده برای لایه‌های مختلف با هم متفاوت است. بنابراین، مفهوم "وضوح تصویر" به عنوان یک ویژگی اصلی برای استخراج لایه‌های مختلف در نظر گرفته می‌شود. اختلاف بین میزان وضوح تصاویر فریم‌های لایه‌های مختلف به منظور تعریف مفهوم "اختلاف" و برای محاسبه اختلاف درون نمایی می‌تواند مورد استفاده قرار گیرد. به عبارت دیگر، تنها چیزی که برای تعمیم دادن مفهوم اختلاف درون نمایی برای مقیاس‌پذیری‌های مختلف نیاز داریم، ویژگی اصلی آن مقیاس‌پذیری خاص است که برای استخراج لایه‌های مختلف در آن مقیاس‌پذیری مورد استفاده قرار گرفته است. از آنجایی که پیدا کردن این ویژگی در مقیاس‌پذیری‌های مختلف بدیهی و ساده است، به راحتی می‌توان مفهوم "اختلاف درون نمایی" را برای انواع مختلف مقیاس‌پذیری استخراج کرد. برای وضوح بیشتر، جزئیات روش پیشنهادی برای دو نوع مقیاس‌پذیری خاص: مقیاس-پذیری نمایی و عمق در زیرفصل‌های بعد توضیح داده خواهد شد.

۶-۱-۲-۱-۱- ارزیابی کیفیت برای هر لایه در مقیاس‌پذیری نمایی

در کد کردن ویدئوی چندنمایی، به منظور فشردن سازی کارآمد، هر نما باید از روی نماهایی که بیشترین شباهت را با آن دارند، کد شود. بنابراین، میزان اعوجاج هر نما با میزان اعوجاج نماهای مرجع آن وابستگی بالایی خواهد داشت. بنابراین، وزن اعوجاج هر نما در ارزیابی کیفیت نهایی، باید از وزن نماهای مرجع آن کمتر باشد. با استفاده از این روش، اثر اعوجاج نمای مرجع در میزان اعوجاج نهایی بیش از حد برآورد نمی‌شود. در کد کردن ویدئوی چندنمایی، پیش‌بینی بین نمایی به منظور استخراج وابستگی بین نماها مورد استفاده قرار می‌گیرد [115]. بر این اساس، اختلاف بین فریم‌های نماهای مختلف، معیار مناسبی برای

انتخاب مرجع مناسب برای پیش‌بینی بین‌نمایی است، زیرا میزان شباهت بین نماهای مختلف را نشان داده و می‌تواند به‌عنوان معیاری برای نشان دادن کارایی فرآیند پیش‌بینی در نظر گرفته شود. کارایی فرآیند پیش‌بینی اثر مستقیمی روی کیفیت نهایی دارد. در نتیجه، اختلاف بین فریم‌های نماهای مختلف می‌تواند تخمین خوبی برای میزان کیفیت دریافتی نهایی باشد. به این دلیل، روش‌شناسی ارائه شده در این رساله از مفهوم اختلاف بین فریم‌های نماهای مختلف تحت عنوان اختلاف درون‌نمایی برای هر نوع خاص مقیاس-پذیری استفاده می‌کند. سپس این مفهوم برای تعیین مقادیر وزن‌ها برای کیفیت نماهای مختلف مورد استفاده قرار می‌گیرد.

برای استخراج مقادیر وزن‌ها با استفاده از اختلاف درون‌نمایی به‌صورت زیر عمل می‌کنیم. فرض کنید که چهار نما در لایه پایه وجود داشته باشد. بنابراین، ساختار پیش‌بینی برای این لایه به‌صورت شکل ۲-۶ خواهد بود.



شکل ۲-۶ - ساختار پیش‌بینی انتخابی برای کدکردن ویدئوی چندنمایی

فرض کنید که V_i مهم‌ترین واحد مرتبه پایین بوده و بیشترین مقدار وزن را دارد. بر اساس شکل ۲-۶، V_k باید از روی V_i پیش‌بینی شود. اگر اختلاف بین این دو نما کم باشد، V_k می‌تواند با کارایی بهتری از روی V_i پیش‌بینی شود. بنابراین، مقدار وزن آن در محاسبه کیفیت نهایی باید کمتر از وزن V_i باشد. به این ترتیب، مقدار نرخ بیت تخصیص داده شده به V_k کاهش می‌یابد، بدون اینکه روی کیفیت نهایی اثر بدی بگذارد، زیرا می‌تواند به خوبی از روی V_i که شباهت زیادی با آن دارد، پیش‌بینی شود. بر اساس این استدلال، مقدار وزن نمای V_k کمتر از مقدار وزن V_i خواهد بود و با استفاده از مقدار اختلاف بین این دو نما محاسبه می‌شود. به‌طور مشابه، مقدار مناسب وزن برای V_j و V_l نیز انتخاب می‌شود. معادله ۱-۶ برای محاسبه مقادیر وزن نماها مورد استفاده قرار می‌گیرد:

$$w_k = \max_{i \in R} (d_{ik} \times w_i) \quad (1-6)$$

در این معادله، R مجموعه نماهای مرجع برای نمای k و d_{ik} اختلاف بین V_i و V_k و w_i مقدار وزن نمای مرجع V_i در ارزیابی کیفیت نهایی این لایه است. همان طور که معادله (۶-۱) نشان می‌دهد، w_k مقدار وزن نمای V_k با استفاده از w_i مقدار وزن نمای مرجع آن V_i و بر اساس اختلاف بین نمایی آنها محاسبه می‌شود. علاوه بر این، اگر نمایی بیش از یک نمای مرجع داشته باشد، مانند V_j در شکل ۶-۲، مقدار وزن به ازای هر نمای مرجع محاسبه شده و بزرگترین مقدار به عنوان وزن نهایی انتخاب می‌شود. تابع ماکزیمم به این منظور استفاده شده است که اثر کیفیت نمای فعلی در ارزیابی کیفیت نهایی به واسطه وجود نماهای مرجع با کارایی کمتر، کمتر از حد تخمین زده نشود. نتایج شبیه‌سازی نشان می‌دهد که با استفاده از این روش، تخمین دقیق‌تری برای میزان کیفیت نهایی خواهیم داشت.

۶-۱-۲-۱-۲- ارزیابی کیفیت برای هر لایه در مقیاس پذیری عمق

روش‌شناسی پیشنهادی می‌تواند برای استخراج معیار ارزیابی کمی کیفیت برای مقیاس-پذیری عمق نیز اعمال شود. ابتدا باید مهم‌ترین واحد مرتبه پایین را بر اساس کاربرد مورد نظر و ترجیحات کاربران انتخاب کنیم. برای مثال، اگر کاربر ترجیح دهد که قسمت‌های نزدیک‌تر صحنه تصویربرداری شده را با وضوح بیشتری ببیند، واحدهای مرتبه پایینی که شامل سطوح عمق این قسمت‌های صحنه هستند، به عنوان مهم‌ترین واحدهای مرتبه پایین در نظر گرفته می‌شوند. سپس حداکثر وزن ممکن به کیفیت این واحدها اختصاص داده می‌شود. واحدهای مرتبه پایین با سطوح عمق دورتر از مکان‌هایی که مورد توجه کاربر است، تأثیر زیادی در بالا بردن کیفیت تجربی کاربر در این کاربرد ندارد. در نتیجه، روش‌شناسی پیشنهادی ارائه شده در این مقاله از مفهوم فاصله بین سطوح عمق مربوط به واحدهای مرتبه پایین به عنوان معیار اختلاف درون‌نمایی در این مقیاس‌پذیری خاص استفاده کرده است. سپس این معیار برای تعیین مقادیر وزن‌ها استفاده می‌شود. بر اساس این استدلال، وزن مربوط به واحد مرتبه پایین k به صورت زیر محاسبه می‌شود:

$$w_k = \frac{1}{1+d_{ik}} \times w_i \quad (۶-۲)$$

که d_{ik} اختلاف بین سطوح عمق برای واحدهای مرتبه پایین k و i است، که واحد مرتبه پایین i مهم‌ترین واحد مرتبه پایین و w_i وزن مربوط به این واحد است. از آنجایی که مخرج کسر معادله (۶-۲) همیشه بزرگ‌تر از صورت آن است، مقدار وزن همه واحدهای مرتبه پایین کمتر از مقدار وزن مهم‌ترین واحد مرتبه پایین خواهد بود. علاوه بر این، این معادله به واحدهای مرتبه پایین با اختلاف کمتر با مهم‌ترین واحد مرتبه پایین، وزن‌هایی با مقادیر

بیشتر اختصاص می‌دهد. به این ترتیب، تأثیر کیفیت این واحدها در ارزیابی کیفیت نهایی بیشتر انعکاس داده می‌شود.

به‌طور مشابه، همان‌طور که در مرحله اول شکل ۶-۱ نشان داده شده است، روش جمع‌کردن وزن‌دار می‌تواند برای استخراج کیفیت نهایی هر لایه برای سایر مقیاس‌پذیری‌ها نیز استفاده شود. وزن‌ها با استفاده از ویژگی‌های خاص هر نوع مقیاس‌پذیری و مفهوم اختلاف درون‌نمایی متناظر با آن مقیاس‌پذیری تعیین می‌شود. برای استخراج وزن مناسب برای واحدهای مرتبه پایین هر نوع مقیاس‌پذیری، ابتدا باید ویژگی اصلی آن مقیاس‌پذیری که واحدهای مرتبه پایین را به هم ارتباط می‌دهد و چگونگی این ارتباط مشخص شود. برای مثال، در مقیاس‌پذیری نمایی، واحدهای مرتبه پایین از طریق ساختار پیش‌بینی با هم ارتباط برقرار می‌کنند. در مقیاس‌پذیری عمق، این ارتباط از طریق فاصله بین سطوح عمق تعریف می‌شود. در مرحله بعد، باید تعیین شود که این ارتباط چگونه روی کیفیت نهایی اثر می‌گذارد. برای مثال، بر اساس ساختار پیش‌بینی، از آنجایی که نمای مرجع هر نما اثر بیشتری روی کیفیت نهایی دارد، باید وزن بیشتری نیز در محاسبه کیفیت نهایی داشته باشد. این مسأله در معادله (۶-۱) برای مقیاس‌پذیری نمایی بیان شده است. در مقیاس‌پذیری عمق، از آنجایی که ارتباط بین واحدهای مرتبه پایین بر اساس فاصله سطوح عمق آنها تعریف می‌شود، واحدهای مرتبه پایین با فاصله بیشتر از مهم‌ترین واحد مرتبه پایین، برای بینندگان اهمیت کمتری دارند و کیفیت آنها باید وزن کمتری در ارزیابی کیفیت نهایی داشته باشد. بنابراین تابع محاسبه وزن برای این نوع مقیاس‌پذیری به‌صورت معادله (۶-۲) خواهد بود.

۶-۱-۲-۲- ارزیابی کیفیت نهایی برای ویدئوی چندنمایی مقیاس‌پذیر

در این بخش از رساله، توضیح می‌دهیم که چگونه کیفیت ارزیابی شده برای لایه‌های مختلف را با هم ترکیب کرده و کیفیت نهایی ویدئوی چندنمایی مقیاس‌پذیر را استخراج کنیم. می‌توان روش پیشنهاد شده در بخش ۶-۱-۲-۱ را تعمیم داد و برای ارزیابی کیفیت نهایی ویدئوی چندنمایی مقیاس‌پذیر، وزن‌های مختلفی به کیفیت لایه‌های متفاوت ویدئوی مقیاس‌پذیر تخصیص داد و کیفیت نهایی را از جمع وزن‌دار کیفیت لایه‌های مختلف استخراج کرد. روش‌شناسی ارائه شده در این رساله پیشنهاد می‌کند که تعداد نماهایی که به گیرنده رسیده‌اند و اختلاف بین لایه‌ای به‌منظور تخمین کیفیت نهایی ویدئوی چندنمایی مقیاس‌پذیر استفاده شوند. مفهوم اختلاف بین لایه‌ای برای هر نوع مقیاس‌پذیری تعریف می‌شود. در این رساله، این مفهوم را برای دو نوع مقیاس‌پذیری نمایی و عمق تعریف می‌کنیم،

ولی این مفهوم می تواند به سایر مقیاس پذیری ها نیز به صورت مشابه آنچه در ۱-۲-۱-۶- توضیح داده شد، تعمیم داده شود.

در ویدئوی چندنمایی مقیاس پذیر، کیفیت لایه های افزایشی بر اساس میزان اختلاف بین- لایه ای آنها با لایه پایه، روی کیفیت نهایی تأثیرگذار خواهد بود. وقتی که میانگین اختلاف بین واحدهای مرتبه پایین لایه های پایه و افزایشی کم باشد، کیفیت دریافتی بیننده با دریافت کردن لایه های افزایشی با کیفیت بالا بهبود می یابد. بنابراین، در این حالت باید وزن بیشتری به کیفیت لایه های افزایشی اختصاص یابد. بر این اساس، روش شناسی ارائه شده پیشنهاد می دهد که از مفهوم اختلاف بین لایه ای برای یافتن وزن مناسب برای کیفیت لایه های پایه و افزایشی استفاده شود.

علاوه بر این، مفهوم اختلاف بین لایه ای می تواند کیفیت نهایی دریافتی را با تغییر دادن کیفیت نماهای بازتولید شده نیز تغییر دهد. هنگامی که لایه های افزایشی به واسطه محدودیت های شبکه یا گیرنده حذف می شوند، واحدهای مرتبه پایین از دست رفته باید در سمت گیرنده بازتولید شوند. بازتولید این واحدهای از طریق اطلاعاتی که به گیرنده رسیده مانند اطلاعات لایه پایه انجام می شود. بنابراین، اختلاف بین لایه ای کمتر بین لایه های پایه و افزایشی می تواند به بازتولید نماهایی با کیفیت بهتر منجر شود.

توضیحات بالا نشان می دهد که کیفیت دریافتی نهایی علاوه بر کیفیت هر لایه، به تعداد واحدهای مرتبه پایین دریافت شده برای لایه افزایشی و نیز به اختلاف بین لایه ای لایه های پایه و افزایشی بستگی خواهد داشت. به این ترتیب، معادله نهایی برای محاسبه کیفیت نهایی در ویدئوی چندنمایی مقیاس پذیر، پنج پارامتر مختلف دارد، Q_e ، Q_b ، w_b ، w_e و میانگین اختلاف بین لایه ای. این پارامترها و روش محاسبه کیفیت نهایی در ادامه توضیح داده می شود.

w_e و w_b به ترتیب وزن های مربوط به کیفیت لایه های پایه و افزایشی هستند. از آنجایی که لایه پایه مهم ترین لایه است، به w_b حداکثر مقدار ممکن یعنی ۱ را اختصاص می دهیم. سپس وزن لایه های افزایشی به صورت زیر محاسبه می شود:

$$w_e = \frac{1}{1 + \text{avg_inter_LD}} w_b \quad (3-6)$$

که avg_inter_LD میانگین اختلاف بین لایه ای لایه های پایه و افزایشی است. همان طور که این معادله نشان می دهد، w_e همیشه از w_b کوچکتر خواهد بود و ارتباط بین آنها از طریق میانگین اختلاف بین لایه ای تعیین می شود.

Q_e و Q_b به ترتیب کیفیت نهایی لایه‌های پایه و افزایشی هستند که بر طبق معادلات (۴-۶) و (۵-۶) تعریف می‌شوند:

$$Q_b = \sum_{i=1}^L w_i Q_i \quad (۴-۶)$$

$$Q_e = \sum_{j=i}^E \sum_{k=1}^{K_E} w_{jk} Q_{jk} \quad (۵-۶)$$

که L تعداد کل واحدهای مرتبه پایین لایه پایه، E تعداد لایه‌های افزایشی، Q_i کیفیت هر یک از واحدهای مرتبه پایین لایه پایه، w_i و w_{jk} به ترتیب مقادیر وزن‌های مربوط به واحدهای مرتبه پایین لایه‌های پایه و افزایشی و K_E تعداد واحدهای مرتبه پایین هر یک از لایه‌های افزایشی است. معادلات (۴-۶) و (۵-۶) از روش‌های پیشنهادی برای ارزیابی کیفیت هر لایه در بخش ۶-۱-۲-۱- به دست آمده است که پیشنهاد می‌دهد کیفیت هر لایه به صورت جمع وزن‌دار کیفیت واحدهای مرتبه پایین آن در نظر گرفته شود. بنابراین برای مثال، مقادیر وزن w_i و w_{jk} برای مقیاس‌پذیری نمایی و عمق به ترتیب از روابط (۶-۱) و (۶-۲) به دست می‌آیند.

avg_inte_LD میانگین اختلاف بین لایه‌های بین واحدهای مرتبه پایین لایه‌های پایه و افزایشی است و از معادله زیر به دست می‌آید:

$$avg_inter_LD = (\sum_{i=1}^P \sum_{k=1}^{K_E} d_{ik}) / (P \times K_E) \quad (۶-۶)$$

که P تعداد کل واحدهای مرتبه پایین لایه پایه است. به این ترتیب، معادله نهایی برای محاسبه کیفیت نهایی ویدئوی چندنمایی مقیاس‌پذیر به صورت زیر خواهد بود:

$$Q_T = w_b Q_b + w_e Q_e - \alpha (avg_inter_LD)(LoU_T - LoU_{rec-EL}) \quad (۷-۶)$$

که LoU_T تعداد کل واحدهای مرتبه پایین و LoU_{rec-EL} تعداد واحدهای مرتبه پایین دریافت‌شده برای لایه‌های افزایشی است. پارامتر α به صورت تجربی محاسبه شده و برای مقیاس‌کردن محدوده مناسب برای کیفیت استفاده می‌شود. همان‌طور که پیش‌تر توضیح داده شد، واحدهای مرتبه پایین بر اساس انواع مختلف مقیاس‌پذیری تعریف می‌شوند. این معادله با در نظر گرفتن موارد زیر استخراج شده است. هنگامی که میانگین اختلاف بین لایه‌های ناچیز باشد، تعداد واحدهای مرتبه پایین دریافتی لایه‌های افزایشی در تعیین کیفیت نهایی اهمیت کمتری خواهد داشت زیرا واحدهایی از لایه افزایشی که حذف شده‌اند، می-

توانند با استفاده از داده‌های لایه پایه و با کیفیت بالا بازتولید شوند. در این حالت، قسمت دوم معادله (۶-۷) برابر صفر خواهد بود و کیفیت نهایی جمع وزن دار کیفیت لایه‌های مختلف خواهد بود. اما وقتی که میانگین اختلاف بین لایه‌های زیاد باشد و تعدادی از واحدهای مرتبه پایین لایه افزایشی حذف شوند، کیفیت دریافتی واحدهای مرتبه پایین حذف شده بعد از بازتولید، پایین خواهد بود. بنابراین، جمع وزن دار کیفیت لایه‌های مختلف بیش از کیفیت واقعی دریافتی می‌شود. در این حالت قسمت دوم معادله (۶-۷) این برآورد بیش از اندازه را جبران خواهد کرد. بنابراین، کیفیت نهایی به میزان کیفیت دریافتی توسط بیننده نزدیک تر خواهد بود.

همان‌طور که در مرحله ۲ شکل ۶-۱ نشان داده شده است، این روش می‌تواند برای انواع مختلف مقیاس پذیری و برای محاسبه کیفیت نهایی ویدئوی چندنمایی مقیاس پذیر اعمال شود. مفهوم میانگین اختلاف بین لایه‌های بر اساس نوع مقیاس پذیری تعیین می‌شود.

۶-۲- ارزیابی روش ارائه شده برای تخمین کمی کیفیت ویدئوی چندنمایی مقیاس پذیر

در این بخش، کارایی روش‌شناسی ارائه شده برای ارزیابی کیفیت ویدئوی چندنمایی مقیاس پذیر را با مقایسه نتایج استخراج شده از این روش با نتایج حاصل از ارزیابی بصری، برای دو نوع مقیاس پذیری نمایی و عمق مورد بررسی قرار می‌دهیم. برای هر مقیاس پذیری، ابتدا آزمایشی ترتیب داده‌ایم که کارایی روش ارائه شده برای ارزیابی کیفیت هر لایه را بررسی کنیم. نتایج این آزمایش کارایی روش ارائه شده در انتخاب مناسب وزن‌ها برای کیفیت واحدهای مرتبه پایین در هر لایه را نشان می‌دهد. سپس، دقت معیار ارزیابی کمی کیفیت برای ویدئوی چندنمایی مقیاس پذیر با مقایسه نتایج حاصل از این روش با نتایج ارزیابی بصری مورد بررسی قرار گرفته است. در مقیاس پذیری نمایی، اثر نماهای مختلف و اختلاف بین نمایی و بین لایه‌ای در ارزیابی کیفیت نهایی مورد بررسی قرار گرفته است. پنج ویدئوی چندنمایی مختلف، Ballet, Break-dancer, Kendo, Balloons و Tunnel برای ارزیابی این روش استفاده شده است [116] [110] [98] از آنجایی که ویدئوی چندنمایی سه بعدی استاندارد در دسترس عموم برای آزمایش کردن وجود ندارد، به منظور آزمایش کردن ویژگی چندنمایی تنها از ویدئوهای دوبعدی چندنمایی استفاده شده است، اما به منظور ارزیابی تجربه سه بعدی برای روش ارائه شده ارزیابی کیفیت، از ویدئوهای برجسته‌نما استفاده شده و ویدئوها روی صفحه نمایش سه بعدی با استفاده از عینک آبی - قرمز دیده شده‌اند. جدول

۱-۶ ویژگی‌های ویدئوهای آزمایش‌شده را نشان می‌دهد. نکته قابل ذکر اینکه در همه ویدئوهای آزمایش‌شده در این جدول، دوربین‌ها ثابت هستند. نتایج با استفاده از نرم‌افزار مرجع کدکردن ویدئوی چندنمایی برای استاندارد H.264/MVC نسخه ۸/۵ استخراج شده است [105].

نکته قابل ذکر اینکه ارزیابی بصری کیفیت ویدئوهای واکدشده با استفاده از روش Double Stimulus Continuous Quality Scale که در استاندارد ITU-R Recommendation 500 توضیح داده شده است، انجام شده است [117]. جزئیات روش ارزیابی بصری در بخش ۳-۶-۳ توضیح داده شده است.

جدول ۱-۶ - خصوصیات ویدئوهای آزمایش‌شده

تعداد فریم‌ها	تعداد دوربین‌ها	نرخ فریم (فریم بر ثانیه)	اندازه فریم	ویدئو
۱۰۰	۸	۱۵	۱۰۲۴×۷۶۸	Ballet
۱۰۰	۸	۱۵	۱۰۲۴×۷۶۸	Break-dancer
۱۰۰	۷	۱۵	۱۰۲۴×۷۶۸	Kendo
۱۰۰	۸	۱۵	۱۰۲۴×۷۶۸	Balloons
۲۵۰	۲	۲۵	۷۲۰×۵۷۶	Tunnel

۳-۶-۱ - ارزیابی روش ارائه شده برای تخمین کمی کیفیت هر یک از لایه-

های ویدئوی چندنمایی مقیاس‌پذیر

همان‌طور که اشاره شد، کیفیت نهایی هر لایه از ویدئوی چندنمایی مقیاس‌پذیر با استفاده از جمع وزن‌دار کیفیت واحدهای مرتبه پایین آن و با انتخاب مناسب مقادیر وزن‌ها با استفاده از اختلاف بین آن واحدها می‌تواند با دقت بالایی تخمین زده شود. به این ترتیب، اثر مهم-ترین واحد مرتبه پایین، که بیشترین تأثیر را در کیفیت دریافتی کاربر دارد، در کیفیت نهایی بیشتر منعکس می‌شود. به‌منظور اثبات این فرضیه، ۵ نوع خطای مختلف [118][58] را به واحدهای مرتبه پایین مهم و غیرمهم اعمال کردیم و ۱۰ حالت مختلف برای اعوجاج این واحدها استخراج کردیم. این ۱۰ حالت مختلف در جدول ۳-۶ نشان داده شده است. این نویزها از دست رفتن اطلاعات در سمت گیرنده به علت محدودیت‌ها منابع مانند پهنای باند گیرنده و خطای شبکه را شبیه‌سازی می‌کنند. تفاوت بین سناریوی اعوجاج ۱ و ۲ این است که مهم‌ترین واحد مرتبه پایین در هر یک از آنها به‌طور متفاوت خراب می‌شود. سپس کیفیت نهایی ویدئو با استفاده از معیار ارزیابی کمی ارائه‌شده و نیز روشی که در [119][101] پیشنهاد شده محاسبه شده است. در روش پیشنهادی مقالات [119][101]

کیفیت واحدهای مرتبه پایین به صورت ساده با یکدیگر جمع می شود. نتایج نشان می دهد که اثر کیفیت مهم ترین واحد مرتبه پایین، در معیار ارزیابی ارائه شده در این مقاله بهتر انعکاس داده می شود. نتایج این ارزیابی برای دو نوع مقیاس پذیری مختلف در زیربخش های زیر ارائه شده است.

جدول ۶-۲ - خصوصیات نویزهای اعمال شده برای مقیاس پذیری نمایی

حالات مختلف اعوجاج	نمای ۱	نمای ۲
۱	اضافه کردن نویز گاوسی (با $\sigma = 20$)	اضافه کردن نویز گاوسی (با $\sigma = 20$)
۲		اضافه کردن نویز گاوسی (با $\sigma = 20$)
۳	اضافه کردن نویز salt-pepper (۲۰٪)	اضافه کردن نویز salt-pepper (۲۰٪)
۴		اضافه کردن نویز salt-pepper (۲۰٪)
۵	فشرده سازی بالا (با $QP = 40$)	فشرده سازی بالا (با $QP = 40$)
۶		فشرده سازی بالا (با $QP = 40$)
۷	مدل نویز Gilbert - Elliot (نرخ خطای بسته = ۰/۲)	مدل نویز Gilbert - Elliot (نرخ خطای بسته = ۰/۲)
۸		مدل نویز Gilbert - Elliot (نرخ خطای بسته = ۰/۲)
۹	گم شدن فریم (توزیع شده به صورت یکنواخت، نرخ گم شدن فریم = ۲۰٪)	گم شدن فریم (توزیع شده به صورت یکنواخت، نرخ گم شدن فریم = ۲۰٪)
۱۰		گم شدن فریم (توزیع شده به صورت یکنواخت، نرخ گم شدن فریم = ۲۰٪)

۶-۲-۱-۱- ارزیابی روش ارائه شده برای تخمین کمی کیفیت هر یک از لایه های ویدئوی

چندنمایی مقیاس پذیر برای مقیاس پذیری نمایی

به منظور بررسی کارایی معیار ارائه شده برای ارزیابی کیفیت در مقیاس پذیری نمایی، ۴ نما از ویدئوی Ballet را همان طور که در شکل ۶-۳ نشان داده شده است، انتخاب کردیم. این نماها با استفاده از ساختار پیش بینی شکل ۶-۲ کد شدند.



شکل ۳-۶ - نماهای اصلی از ویدئوی Ballet

فرض کنیم که نمای اول (نمای V_i در شکل ۲-۶)، به‌عنوان مهم‌ترین نما برای کاربر انتخاب شود. بنابراین، کیفیت دریافتی توسط بیننده به‌شدت تحت تأثیر کیفیت این نما خواهد بود، و وزن مربوط به کیفیت این نما در کیفیت نهایی باید از همه بیشتر باشد، وزن سایر نماها از روی وزن این نما و براساس ساختار پیش‌بینی شکل ۲-۶ تعیین می‌شود. سپس نویزهای مختلف اشاره شده در جدول ۲-۶ را به نماهای مهم و غیرمهم به‌صورت جداگانه اعمال کردیم. نمونه‌هایی از نویزهای اعمال شده به‌منظور ارزیابی بصری در شکل ۴-۶ نشان داده شده است.



(الف)



(ب)



(ج)

شکل ۴-۶ - حالت‌های مختلف اعمال نویز به نماهای مهم و غیرمهم (الف) حالت ۱ (ب) حالت ۲ (ج) حالت ۳ (د) حالت ۴ (ه) حالت ۵

در شبیه‌سازی‌های ما، نویزهای گاوسی و salt-pepper برای مدل کردن نویزهایی که در فرآیند ضبط ویدئو ایجاد می‌شود، استفاده شده است. بنابراین، ما این نویز را به بعضی نماها اعمال کردیم و ۴ حالت مختلف همان‌طور که در جدول ۲-۶ نشان داده شده است، استخراج کردیم. سپس هر ۴ نما را با استفاده از ساختار پیش‌بینی شکل ۲-۶ کد و واکنش کردیم.

اعوجاج فشرده سازی بالا برای مدل کردن نویزهایی که به واسطه کد کردن ایجاد می شود، استفاده شده است. خطای گم شدن فریم و Gilbert-Elliot نیز به منظور مدل کردن نویز که در زمان ارسال ایجاد می شود، به کار برده شده اند. بنابراین، در این حالت، ابتدا ۴ نمای اولیه با ساختار پیش بینی شکل ۶-۲ کد شدند، سپس این دو نوع نویز به رشته بیت کد شده اعمال شد و در نهایت با استفاده از همان ساختار پیش بینی واکد شدند. در این ارزیابی، برای هر حالت اعمال نویز، یک ویدئوی اولیه و یک ویدئوی واکد شده برای هر نما داریم، که بعضی از نماهای واکد شده به واسطه نویزهای اعمال شده، دچار اعوجاج شده اند. این دو ویدئو به منظور ارزیابی کمی و بصری کیفیت مورد استفاده قرار گرفته اند. به منظور محاسبه اختلاف بین نمایی که برای محاسبه وزن های مورد نیاز در ارزیابی کمی کیفیت در معادله (۶-۱) لازم است، همانند بخش های قبلی از روشی که در [101] پیشنهاد شده است، استفاده کردیم. جدول ۶-۳ مقادیر اختلاف بین نمایی برای ویدئوی Ballet را که با این روش استخراج شده است، نشان می دهد. در نهایت، مقادیر وزن بر اساس معادله (۶-۱) و با مقادیر محاسبه شده برای اختلاف بین-نمایی محاسبه می شود.

جدول ۶-۳ - اختلاف بین نمایی برای نماهای مختلف ویدئوی Ballet

نمای ۱	نمای ۲	نمای ۳	نمای ۴	نمای ۵	نمای ۶	نمای ۷	
۰/۱۷	۰/۱۵	۰/۱۲	۰/۹۷	۰/۱۱	۰/۱۴	۰/۱۲	نمای ۰
	۰/۱۶	۰/۱۳	۰/۸۷	۰/۱	۰/۱۵	۰/۱۱	نمای ۱
		۰/۲	۰/۹۶	۰/۱۴	۰/۱۸	۰/۱۵	نمای ۲
			۰/۱۱	۰/۱	۰/۱۷	۰/۱۳	نمای ۳
				۰/۲۳	۰/۲۳	۰/۱۷	نمای ۴
					۰/۲۳	۰/۱۷	نمای ۵
						۰/۲	نمای ۶

سپس کیفیت نهایی ۴ نما با استفاده از جمع ساده کیفیت نماهای مختلف و نیز با روش جمع وزن دار و وزن های استخراج شده در روش شناسی ارائه شده در این رساله محاسبه شده است. در نهایت ارزیابی بصری، به منظور استخراج ارتباط بین مقادیر استخراج شده از ارزیابی کمی با دریافت بصری بینندگان مورد استفاده قرار گرفته است. ضریب همبستگی^۱ بین مقادیر بصری و کمی کیفیت استخراج شده برای حالت های مختلف اعمال نویز در جدول ۶-۴ نشان داده شده است. نتایج نشان می دهد که معیار ارائه شده قادر است اثر کیفیت نمای مهم در کیفیت نهایی را بهتر انعکاس دهد.

^۱ Correlation coefficient

جدول ۴-۶ - مقایسه بین معیار ارزیابی کمی ارائه‌شده برای هر لایه و روش جمع ساده برای مقیاس‌پذیری نمایی برای حالت‌های مختلف اعمال نویز

ضریب همبستگی بین معیار ارزیابی کمی جمع ساده کیفیت نماهای مختلف و معیار ارزیابی بصری					ضریب همبستگی بین معیار ارزیابی کمی ارائه‌شده و معیار ارزیابی بصری					ویدئو
گم- شدن فریم	Gilbert- Elliot	فشرده- سازی بالا	نویز salt- pepper	نویز گاوسی	گم- شدن فریم	Gilbert- Elliot	فشرده- سازی بالا	نویز salt- pepper	نویز گاوسی	
۰/۸۴	۰/۳	۰/۵۲	۰/۸	۰/۷۲	۰/۹۲	۰/۷۶	۰/۸۹	۰/۹	۰/۸۸	Ballet
۰/۵۲	۰/۹۲	۰/۴۸	۰/۸	۰/۹	۰/۷	۰/۹۵	۰/۸۸	۰/۹	۰/۹	Break- dancer
۰/۸۴	۰/۸	۰/۷۶	۰/۸	۰/۷۱	۰/۹۷	۰/۸	۰/۹۳	۰/۹	۰/۷۴	Balloons
۰/۸۳	۰/۷۵	۰/۷۵	۰/۸	۰/۷	۰/۹۶	۰/۸	۰/۹۲	۰/۹	۰/۷	Kendo

۲-۱-۲-۶- ارزیابی روش ارائه‌شده برای تخمین کمی کیفیت هر یک از لایه‌های ویدئوی چندنمایی مقیاس‌پذیر برای مقیاس‌پذیری عمق

به‌منظور ارزیابی روش ارزیابی کیفیت ارائه‌شده برای مقیاس‌پذیری عمق، سه واحد مرتبه پایین با سطوح عمق متفاوت از ویدئوهای آزمایش‌شده به‌عنوان لایه پایه انتخاب شدند. شکل ۵-۶ واحدهای مرتبه پایین برای نزدیک‌ترین قسمت‌های صحنه تصویربرداری شده به بینندگان را برای ویدئوهای Balloons و Kendo نشان می‌دهد.



(الف)

(ب)

شکل ۵-۶ - واحدهای مرتبه پایین مربوط به قسمت‌هایی از صحنه که نزدیک به بیننده هستند. (الف) ویدئوی Balloons (ب) ویدئوی Kendo

فرض کنیم که بیننده ترجیح می‌دهد که قسمت‌های نزدیک‌تر از صحنه تصویربرداری شده را ببیند. بنابراین این قسمت‌ها باید با کیفیت بالاتری نمایش داده شوند. واحدهای مرتبه پایین مطابق با این نواحی که قطعاً در لایه پایه هستند، به‌عنوان مهم‌ترین واحدهای مرتبه پایین این لایه در نظر گرفته می‌شوند. در روش پیشنهاد شده باید وزن ۱ به کیفیت این نواحی اختصاص داده شود تا کیفیت آنها به‌طور مؤثر در کیفیت نهایی انعکاس داده شود. سپس، مقادیر وزن برای سایر واحدهای مرتبه پایین بر اساس معادله (۲-۶) و به کمک فاصله بین سطوح عمق واحدهای مرتبه پایین مختلف تعیین می‌شود. مجدداً حالت‌های

مختلف اعمال نويز جدول ۶-۲ به منظور شبیه‌سازی محدودیت‌های منابع مورد استفاده قرار گرفته است.

کیفیت نهایی این سه واحد مرتبه پایین با استفاده از جمع کردن ساده و نیز با روش پیشنهادی و با استفاده از مقادیر وزن که از معادله (۶-۲) استخراج شده‌اند، محاسبه شد. سپس از ارزیابی بصری برای مقایسه دو روش ارزیابی کیفی با جمع ساده و جمع وزن دار استفاده شد.

به منظور ارزیابی بصری، باید واحدهای مرتبه پایینی که برای لایه پایه در نظر گرفته نشده و ارسال نشده‌اند، در سمت واگردکننده بازتولید شوند. این فرآیند باید با استفاده از اطلاعات جانبی که از ویدئوهای بافت استخراج می‌شود، انجام شود. برای هر لایه خاص، مانند لایه پایه، نقاطی که مربوط به واحدهای مرتبه پایین انتخاب شده است، از روی اطلاعات نقشه عمق به دست می‌آید. در نهایت، قسمت‌هایی از ویدئوی بافت که منطبق با این نقاط هستند، استخراج شده و برای بازتولید نما مورد استفاده قرار می‌گیرند. این فرآیند برای هر فریم از هر نما انجام می‌شود تا کل ویدئوی چندنمایی بازتولید شود. نتایج حاصل از مقایسه ضریب همبستگی بین ارزیابی بصری با روش جمع ساده و جمع وزن دار در جدول ۶-۵ نشان داده شده است. همان‌طور که می‌بینیم ضریب همبستگی معیار ارائه شده و ارزیابی بصری بیشتر از ضریب همبستگی روش جمع ساده و ارزیابی بصری است.

جدول ۶-۵ - مقایسه بین معیار ارزیابی کمی ارائه شده برای هر لایه و روش جمع ساده برای مقیاس‌پذیری عمق برای حالت‌های مختلف اعمال نويز

ضریب همبستگی بین معیار ارزیابی کمی جمع ساده کیفیت نماهای مختلف و معیار ارزیابی بصری					ضریب همبستگی بین معیار ارزیابی کمی ارائه شده و معیار ارزیابی بصری					ویدئو
گم- شدن فریم	Gilbert- Elliot	فشرده- سازی بالا	نويز salt- pepper	نويز گاوسی	گم- شدن فریم	Gilbert- Elliot	فشرده- سازی بالا	نويز salt- pepper	نويز گاوسی	
۰/۹	۰/۷۷	۰/۶۶	۰/۷	۰/۷۷	۰/۹۷	۰/۹۳	۰/۸۵	۰/۹۱	۰/۹۳	Ballet
۰/۷	۰/۷۳	۰/۶۲	۰/۶۶	۰/۷۳	۰/۹۴	۰/۹۴	۰/۸۶	۰/۹۱	۰/۹۴	Break- dancer
۰/۷۱	۰/۶۳	۰/۵۲	۰/۸۳	۰/۶۹	۰/۹۲	۰/۹	۰/۸۲	۰/۹۶	۰/۹	Balloons
۰/۸۵	۰/۸۵	۰/۷۵	۰/۸۵	۰/۸۵	۰/۸۶	۰/۸۶	۰/۸۷	۰/۸۶	۰/۸۶	Kendo

۶-۲-۲-۲ - ارزیابی روش ارائه شده برای تخمین کمی کیفیت نهایی ویدئوی

چندنمایی مقیاس پذیر

در این بخش، کارایی معیار ارزیابی کیفیت ارائه شده برای کل ویدئوی چندنمایی مقیاس-پذیر را از طریق مقایسه نتایج با نتایج ارزیابی بصری بررسی می‌کنیم. یک معیار ارزیابی

کمی خوب باید بتواند معیار ارزیابی بصری را به صورت خطی دنبال کند [102]. نتایج ارائه شده در این بخش تلاش دارد این ارتباط را برای معیار ارزیابی پیشنهاد شده نشان دهد.

۶-۲-۲-۱- ارزیابی روش ارائه شده برای تخمین کمی کیفیت نهایی ویدئوی چندنمایی مقیاس‌پذیر برای مقیاس‌پذیری نمایی

روش ارائه شده برای ارزیابی کیفیت ویدئوی چندنمایی سه بعدی مقیاس‌پذیر به صورت زیر است. برای هر ویدئو، ۴ نما به لایه پایه و بقیه نماها به لایه‌های افزایشی اختصاص داده می‌شوند. دو حالت مختلف در آزمایش‌های ما مورد بررسی قرار گرفته است. در حالت اول، تنها یک لایه افزایشی با ۴ نما و در حالت دوم، دو لایه افزایشی هر کدام با ۲ نما در نظر گرفته شده است. در رابطه (۶-۷) نشان داده شده است که از مفهوم اختلاف بین لایه‌ای می‌توان برای محاسبه کارآمد کیفیت نهایی استفاده کرد. به منظور روشن شدن اهمیت این مفهوم در ارزیابی کیفیت نهایی ویدئوی چندنمایی مقیاس‌پذیر، دو حالت مختلف برای تخصیص نماها به لایه‌های پایه و افزایشی در نظر گرفته شده است. در حالت اول، لایه‌های پایه و افزایشی اختلاف بین لایه‌ای کمی با هم دارند و در حالت دوم، اختلاف بین لایه‌ای زیاد است. جدول ۶-۶ نماهای انتخاب شده برای این دو حالت مختلف برای ویدئوی Break-dancer را نشان می‌دهد.

جدول ۶-۶ - نماهای انتخاب شده برای لایه‌های پایه و افزایشی برای حالت‌های اختلاف بین لایه‌ای کم و زیاد

حالت ۲ یک لایه پایه و دو لایه افزایشی			حالت ۱ یک لایه پایه و یک لایه افزایشی		اختلاف بین لایه-ای	ویدئو
لایه افزایشی ۲	لایه افزایشی ۱	لایه پایه	لایه افزایشی	لایه پایه		
۷-۶	۵-۳	۴-۲-۱-۰	۷-۶-۵-۳	۴-۲-۱-۰	کم	Break-dancer
۵-۴	۳-۲	۷-۶-۱-۰	۵-۴-۳-۲	۷-۶-۱-۰	زیاد	

شکل ۶-۶ نماهای انتخاب شده برای ویدئوی Break-dancer با کمترین اختلاف بین‌نمایی برای دو حالت مختلف را به منظور وضوح بیشتر نشان می‌دهد.

سپس در سمت گیرنده، بعضی از لایه‌های افزایشی به منظور شبیه‌سازی محدودیت منابع حذف می‌شوند. برای بازتولید نماهای حذف شده از نرم‌افزار مرجع بازتولید نما مربوط به گروه MPEG نسخه ۳ [102] استفاده می‌کنیم. نکته قابل توجه اینکه کارایی روش ارزیابی ارائه شده کاملاً از روش استفاده شده برای بازتولید نماها مستقل است و این آزمایش می‌تواند با روش‌های دیگر بازتولید نماها نیز انجام شود. از آنجایی که روش یکسانی برای

بازتولید نماهای از دست رفته برای هر دو روش کمی و بصری استفاده شده است، اثر آن روی کارایی کیفیت هر دو روش یکسان خواهد بود.



شکل ۶-۶ - نماهای انتخاب شده ویدئوی Break-dancer برای لایه های پایه و افزایشی با اختلاف بین لایه های کم برای دو حالت مختلف تخصیص دادن نماها به لایه های مختلف

نماهای از دست رفته در هر دو حالت توضیح داده شده در بالا، بازتولید می شوند. معیار ارزیابی ارائه شده در رابطه (۶-۷) به منظور محاسبه کیفیت نهایی برای هر دو نوع مقیاس-پذیری استفاده می شود. به این منظور، W_e و W_b و میانگین اختلاف بین لایه های به عنوان پارامترهای این معادله باید از روابط (۶-۳) و (۶-۶) به دست آیند.

جدول ۶-۷ نتایج استخراج شده برای ارزیابی بصری در مقایسه با نتایج استخراج شده از معیار کمی ارائه شده را برای ویدئوهای آزمایش شده نشان می دهد. علاوه بر این، جدول ۶-۸ ضریب همبستگی بین معیار ارزیابی کمی ارائه شده و ارزیابی کیفیت بصری را نشان می دهد. مقایسه این نتایج در دو حالت مختلف نشان می دهد که ضریب همبستگی معیار ارائه شده و معیار بصری در دو حالت به اندازه کافی بالا است.

این نتایج کارایی استفاده از مفهوم اختلاف بین لایه های در ارزیابی کیفیت نهایی ویدئوی چندنمایی مقیاس پذیر را نشان می دهد. شایان ذکر است که از آنجایی که معیار ارزیابی

دیگری برای بررسی کیفیت ویدئوی چندنمایی مقیاس پذیر وجود ندارد، نتایج حاصل از روش پیشنهادی تنها با نتایج ارزیابی بصری مقایسه شده است.

جدول ۶-۷ - مقایسه نتایج معیار ارزیابی کمی ارائه شده و نتایج ارزیابی بصری برای مقیاس پذیری نمایی

ویدئو	اختلاف بین لایه‌ای	حالت ۱ (یک لایه پایه و یک لایه افزایشی)			حالت ۲ (یک لایه پایه و دو لایه افزایشی)		
		نتایج ارزیابی بصری	نتایج کمی معیار پیشنهادی (PSNR)	نتایج کمی معیار پیشنهادی (SSIM)	نتایج ارزیابی بصری	نتایج کمی معیار پیشنهادی (PSNR)	نتایج کمی معیار پیشنهادی (SSIM)
Ballet	بالا	۲	۳۰/۷۶	۰/۷۲	۴	۳۸/۱۳	۰/۸۸
	پایین	۲	۳۱/۲۷	۰/۷۳	۴	۳۸/۷۷	۰/۸۸
Break-dancer	بالا	۲	۳۱/۳۲	۰/۷۸	۴	۴۱/۰۴	۰/۹۵
	پایین	۲	۳۱/۲۶	۰/۷۵	۵	۴۰/۱۵	۰/۹۲
Balloons	بالا	۳	۳۲/۹۸	۰/۸۴	۴	۳۳/۱۲	۰/۸۵
	پایین	۳	۳۳/۰۳	۰/۸۵	۴	۳۳/۶۷	۰/۸۶
Kendo	بالا	۳	۴۰/۷۷	۰/۸۵	۴	۴۱/۱۷	۰/۸۶
	پایین	۳	۴۲/۴۱	۰/۹۴	۴	۴۲/۷۱	۰/۹۵

جدول ۶-۸ - ضریب همبستگی بین معیار ارزیابی کمی کیفیت و معیار ارزیابی بصری کیفیت برای مقیاس پذیری نمایی

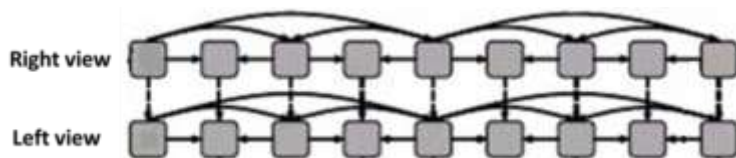
ویدئو	ضریب همبستگی
Ballet	۰/۹۷
Break-dancer	۰/۹۶
Balloons	۰/۷
Kendo	۰/۴

علاوه بر این، مقیاس پذیری نمایی را برای ویدئوی برجسته‌نما نیز اعمال کردیم تا اثر تجربه سه بعدی را نیز در کیفیت نهایی بررسی کنیم. در این حالت، همان‌طور که در شکل ۶-۷ نشان داده شده است، نماهای راست و چپ به ترتیب به عنوان لایه پایه و افزایشی در نظر گرفته شدند [122].



شکل ۶-۷ - لایه‌های پایه و افزایشی در مقیاس پذیری نمایی برای ویدئوی برجسته‌نما، فریم‌های مختلف ویدئوی Tunnel

ابتدا ویدئوی اصلی برجسته‌نما که از نماهای راست و چپ تشکیل شده و به لایه‌های پایه و افزایشی تخصیص داده شده است، با استفاده از ساختار پیش‌بینی شکل ۶-۸، کد و واگد می‌شوند. سپس ویدئوهای اصلی و واگدشده به‌منظور ارزیابی کمی و بصری کیفیت مورد استفاده قرار می‌گیرند. در این آزمایش‌ها از عینک سه‌بعدی آبی - قرمز به منظور دیدن ویدئوی سه‌بعدی استفاده شده است. جدول ۶-۹ نتایج ارزیابی بصری و نتایج حاصل از ارزیابی کمی ارائه شده و جدول ۶-۱۰ ضریب همبستگی بین این دو معیار را نشان می‌دهد. ضریب همبستگی بالا در این جدول نشان می‌دهد که معیار ارزیابی کمی پیشنهادی، نتایج حاصل از ارزیابی بصری را به‌خوبی منعکس می‌کند.



شکل ۶-۸ - ساختار پیش‌بینی استفاده شده برای کدکردن ویدئوی برجسته‌نما

جدول ۶-۹ - مقایسه نتایج ارزیابی کمی و نتایج ارزیابی بصری برای ویدئوی Tunnel برای مقیاس‌پذیری نمایی

نتایج ارزیابی کمی با استفاده از معیار ارائه شده (SSIM)	نتایج ارزیابی کمی با استفاده از معیار ارائه شده (PSNR)	نتایج ارزیابی بصری	ویدئو
۰/۹۸	۴۱/۷۳	۴	Tunnel

جدول ۶-۱۰ - ضریب همبستگی بین معیار ارزیابی کمی و معیار ارزیابی بصری برای ویدئوی Tunnel برای مقیاس‌پذیری نمایی

ضریب همبستگی	ویدئو
۰/۹	Tunnel

۶-۲-۲-۲-۲- ارزیابی روش ارائه شده برای تخمین کمی کیفیت نهایی ویدئوی چندنمایی مقیاس پذیر برای مقیاس‌پذیری عمق

به‌منظور بررسی کارایی معیار ارائه‌شده برای ارزیابی کمی کیفیت ویدئوی چندنمایی مقیاس‌پذیر برای مقیاس‌پذیری عمق، یک لایه پایه و یک لایه افزایشی در نظر می‌گیریم. لایه پایه از واحدهای مرتبه پایینی تشکیل شده‌اند که حاوی قسمت‌هایی از صحنه هستند که به بیننده نزدیک‌تر هستند و لایه افزایشی شامل نواحی باقی‌مانده می‌باشد. سپس لایه‌های افزایشی را به‌منظور شبیه‌سازی محدودیت منابع حذف می‌کنیم.

مجدداً به‌منظور بازتولید نماها در سمت گیرنده، از اطلاعات جانبی حاوی اطلاعات ویدئوی بافت استفاده می‌کنیم. هنگامی که لایه‌ها مشخص شدند، نقاط مرتبط با واحدهای مرتبه پایینی متعلق به لایه‌های افزایشی استخراج می‌شوند. سپس، قسمت‌هایی از ویدئوی بافت

متناظر با این نقاط استخراج شده و به عنوان اطلاعات جانبی ارسال می‌شوند. از آنجایی که لایه پایه از قسمت‌هایی از صحنه تصویربرداری شده تشکیل شده است که برای بیننده مهم‌تر هستند، قسمت‌های متناظر با این نقاط با کیفیت بالاتر کد می‌شوند. برای ارزیابی بصری، هر ویدئو متشکل از لایه‌های پایه و افزایشی به صورت زیر بازتولید می‌شوند. ابتدا، برای لایه پایه، واحدهای مرتبه پایین مرتبط با استفاده از اطلاعات عمق و نیز اطلاعات جانبی شامل ویدئوی بافت با کیفیت بالا بازتولید می‌شوند. از آنجایی که لایه‌های افزایشی ممکن است به واسطه محدودیت منابع، مانند خطای شبکه و محدودیت پهنای باند به مقصد نرسند، واحدهای مرتبه پایین مرتبط به آنها که شامل اطلاعات عمق هستند، از بین می‌روند. بنابراین، قسمت‌های از دست رفته لایه افزایشی تنها با استفاده از قسمت‌های متناظر در ویدئوی بافت از اطلاعات جانبی با کیفیت کم بازتولید می‌شوند.

در نهایت کیفیت نهایی ویدئوی مقیاس‌پذیر با استفاده از رابطه (۶-۷) محاسبه شده و با نتایج حاصل از ارزیابی بصری مقایسه می‌شود. جدول ۶-۱۱ و جدول ۶-۱۲ نتایج آزمایش‌ها برای مقیاس‌پذیری عمق را نشان می‌دهد. همانند نتایج قبلی ارائه شده، نتایج این جداول نیست همبستگی بالا بین معیار ارزیابی کمی ارائه شده و معیار ارزیابی بصری کیفیت را نشان می‌دهد.

جدول ۶-۱۱ - مقایسه نتایج ارزیابی کمی و بصری برای مقیاس‌پذیری عمق

نتایج ارزیابی کمی (PSNR)	نتایج ارزیابی بصری	ویدئو
۱۹/۶۱	۲	Balloons
۱۵/۵۸	۲	Kendo

جدول ۶-۱۲ - ضریب همبستگی بین معیار ارزیابی کمی و بصری کیفیت برای مقیاس‌پذیری عمق

ضریب همبستگی	ویدئو
۰/۸۷	Balloons
۰/۸۳	Kendo

۶-۳ - فرآیند ارزیابی بصری

فرآیند ارزیابی بصری کیفیت برای ویدئوهای واگد شده با استفاده از روش Double Stimulus Continuous Quality Scale (DSCQS) که در استاندارد [117] ITU-R Recommendation 500 شرح داده شده است، انجام شده است. ۲۵ نفر شرکت‌کننده که همگی غیرحرفه‌ای بوده و تجربه‌ای در زمینه پردازش ویدئو و ارزیابی کیفیت نداشته‌اند، در فرآیند ارزیابی بصری کیفیت مشارکت کرده‌اند. ویدئوها توسط همه بیننده‌ها دیده شد و مدت زمان انجام آزمایش

برای هر بیننده ۳۰ دقیقه بود. یک نمایشگر ۱۹ اینچ برای نمایش دادن ویدئوها انتخاب و ویدئوها با وضوح تصویر ۱۰۲۴×۷۶۸ پیکسل نشان داده شدند. وضوح تصویر در ویدئوها همان مقدار اصلی و اولیه آن است تا خطایی به واسطه مقیاس کردن در ویدئوها ایجاد نشود. فاصله بینندگان تا صفحه نمایشگر همان طور که در استاندارد ITU-R 812 توضیح داده شده، به اندازه ۴ برابر ارتفاع صفحه نمایش است. ویدئوهای اصلی (قبل از کد شدن) و ویدئوهای واکدشده به صورت تصادفی به فاصله ۳ ثانیه‌ای به بینندگان نشان داده شد، در فاصله ۳ ثانیه بین دو تصویر صفحه نمایش خاکستری رنگ نمایش داده می‌شود. در آخر، بینندگان کیفیت بصری دو ویدئوی نمایش داده شده را با شماره‌های ۱ تا ۵ نمره‌دهی می‌کنند که به ترتیب نشان‌دهنده "بد"، "ضعیف"، "متوسط"، "خوب" و "عالی" است. نتیجه ارزیابی بصری تفاوت نمره‌هایی است که به ویدئوهای اصلی و واکدشده اختصاص داده شده است. مقدار نهایی (MOS) Mean Opinion Score به صورت میانگین مقادیر عددی که به هر نما تخصیص داده شده، محاسبه می‌شود.

به منظور پردازش نمره‌های ارزیابی بصری، از فرآیند شرح داده شده در استاندارد استفاده شده است. ابتدا، میانگین نمرات از طریق فرمول زیر محاسبه می‌شود:

$$\bar{u}_k = \frac{1}{N} \sum_{i=1}^N u_{ik} \quad (۸-۶)$$

که u_{ik} نمره‌ای است که بیننده i به ویدئوی k داده است و N تعداد کل بینندگان است. میانگین نمرات باید یک فاصله اطمینان^۱ داشته باشد که به صورت $[\bar{u}_k - \delta_k, \bar{u}_k + \delta_k]$ تعریف می‌شود که در آن $\delta_k = 1.96 \frac{S_k}{\sqrt{N}}$ است. S_k انحراف معیار است که از رابطه $S_k = \sqrt{\frac{\sum_{i=1}^N (\bar{u}_k - u_{ik})^2}{(N-1)}}$ به دست می‌آید.

بر اساس استاندارد ITU-R Recommendation، نتایج بینندگانی که همه ارزیابی‌های آنها بسیار خوش-بینانه یا بسیار بدبینانه است در ارزیابی نهایی مورد استناد قرار نمی‌گیرد. سپس، همبستگی بین معیارهای کمی و بصری محاسبه می‌شود. در آمار، ضریب همبستگی معیاری برای نشان دادن وابستگی خطی بین دو متغیر است. ضریب همبستگی بین معیارهای کمی و بصری با استفاده از رابطه زیر محاسبه می‌شود:

$$R = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2 \times \sum_{i=1}^N (Y_i - \bar{Y})^2}} \quad (۹-۶)$$

^۱ Confidence interval

که X_i و Y_i به ترتیب نمرات ارزیابی بصری و کمی هستند. \bar{X} و \bar{Y} میانگین نمونه‌های ارزیابی بصری و کمی و N تعداد کل آزمایش‌هایی است که در فرآیند ارزیابی در نظر گرفته شده است.

۴-۶- جمع‌بندی

در این رساله، یک روش‌شناسی جدید برای استخراج معیار ارزیابی کمی کیفیت برای ویدئوی چندنمایی سه‌بعدی مقیاس‌پذیر پیشنهاد شده است. این روش دو مرحله را برای ارزیابی کیفیت ویدئوی مقیاس‌پذیر در نظر می‌گیرد. در مرحله اول، روشی جدید برای ارزیابی کیفیت هر لایه پیشنهاد می‌شود که از روش جمع وزن‌دار کیفیت واحدهای مرتبه پایین استفاده می‌کند. در این روش، مفهوم اختلاف درون‌نمایی برای هر نوع مقیاس‌پذیری برای محاسبه وزن‌ها مورد استفاده قرار می‌گیرد. در مرحله دوم، روشی برای ترکیب کردن کیفیت لایه‌ها و محاسبه کیفیت نهایی ویدئوی چندنمایی مقیاس‌پذیر معرفی می‌شود. اثر بعضی عوامل مانند تعداد لایه‌های دریافت‌شده در گیرنده و مفهوم اختلاف بین لایه‌ای به-عنوان خصوصیات ذاتی ویدئوی چندنمایی مقیاس‌پذیر در این روش در نظر گرفته می‌شود. نتایج ارزیابی‌ها نشان می‌دهد که معیار ارائه‌شده برای ارزیابی کمی کیفیت که با این روش-شناسی استخراج شده است، دریافت بصری بیننده را به نحو مناسبی منعکس می‌کند.

فصل، هفتم

روش ارائه شده برای بهینه‌سازی نرخ بیت -

اعوجاج در ویدئوی چندنمایی معیاس پذیر

در این فصل به ارائه روش پیشنهادی برای بهینه‌سازی نرخ بیت - اعوجاج برای ویدئوی چندنمایی مقیاس‌پذیر خواهیم پرداخت. ابتدا مبانی بهینه‌سازی نرخ بیت - اعوجاج در ویدئوی چندنمایی مقیاس‌پذیر در بخش ۷-۱- معرفی می‌شود. سپس روش پیشنهادی و ارزیابی آن در بخش‌های ۷-۲- و ۷-۳- ارائه می‌شود.

۷-۱- بهینه‌سازی نرخ بیت - اعوجاج در ویدئوی چندنمایی مقیاس‌پذیر

در بهینه‌سازی نرخ بیت - اعوجاج در کدکننده‌های غیرمقیاس‌پذیر، مسأله بهینه‌سازی مقید نرخ بیت - اعوجاج به صورت حداقل کردن اعوجاج نسبت به نرخ بیت محدود و به صورت زیر تعریف می‌شود:

$$\min D \text{ subject to } R \leq R_{\text{target}} \quad (1-7)$$

که D و R به ترتیب اعوجاج و نرخ بیت کل هستند. این معادله می‌تواند با استفاده از ضریب لاگرانژ، λ ، به مسأله بهینه‌سازی غیرمقید تبدیل شود. در این حالت به جای حداقل کردن اعوجاج، تابع هزینه J به صورت زیر حداقل می‌شود:

$$\min J = D + \lambda \times R \quad (2-7)$$

یک روش مستقیم برای تعمیم این روش به حالت مقیاس‌پذیر این است که اگر اعوجاج بین لایه‌ها مستقل از یکدیگر باشد، هر لایه می‌تواند به صورت مستقل بهینه شود. در این حالت مسأله بهینه‌سازی لاگرانژ می‌تواند به هر لایه اعمال شود به نحوی که برای یک نرخ بیت معین برای هر لایه، تابع هزینه لاگرانژ آن حداقل شود [123]. بنابراین، تابع لاگرانژ برای این مسأله بهینه‌سازی به صورت زیر خواهد بود [123]:

$$\min J_{BL} = D_{BL} + \lambda_{BL} \times R_{BL} \quad R_{BL} \leq R_{\text{target}_{BL}} \quad (3-7)$$

$$\min J_{EL} = D_{EL} + \lambda_{EL} \times R_{EL} \quad R_{EL} \leq R_{\text{target}_{ELn}} \quad (4-7)$$

که D_{BL} و R_{BL} به ترتیب اعوجاج و نرخ بیت لایه پایه و D_{EL} و R_{EL} اعوجاج و نرخ بیت لایه افزایشی n است. برای هر لایه، پارامتر کوانتیزاسیون متناظر، Q_{BL} و Q_{EL} ، با حداقل شدن تابع هزینه آنها به طور مستقل، تعیین می‌شود.

ساختار پیش‌بینی معمول ویدئوی چندنمایی، از مفهوم پیش‌بینی بین‌نمایی برای کدکردن بهینه ویدئوی چندنمایی استفاده می‌کند. از آنجایی که در این ساختار پیش‌بینی، بعضی از

نماها به‌عنوان مرجع برای سایر نماها استفاده می‌شوند، اعوجاج نماهای مرجع می‌تواند روی کیفیت سایر نماها اثر بگذارد. این ارتباط قوی بین اعوجاج لایه‌ها حاکی از آن است که کارایی بهتر برای بهینه‌سازی نرخ بیت - اعوجاج می‌تواند با در نظر گرفتن توأم لایه‌ها به دست آید. در این حالت، نرخ بیت هر لایه با نرخ بیت تخصیص داده شده به همان لایه محدود می‌شود. به این ترتیب، مسأله بهینه‌سازی نرخ بیت - اعوجاج در این حالت، می‌تواند به صورت زیر در نظر گرفته شود:

$$\min J_{BL} = D_{BL} + \lambda_{BL} \times R_{BL} \quad R_{BL} \leq R_{target_{BL}} \quad (5-7)$$

$$\min J_{total} = D_{total} + \lambda_{BL} \times R_{BL} + \lambda_{EL1} \times R_{EL1} + \dots + \lambda_{ELn} \times R_{ELn}$$

$$R_{BL} + R_{EL1} + \dots + R_{ELn} \leq R_{target_{total}} \quad (6-7)$$

در این معادله D_{total} اعوجاج نهایی و $R_{target_{total}}$ ، نرخ بیت مجاز برای کل لایه‌های پایه و افزایشی است.

۲-۷ - روش پیشنهادی

در این بخش، روش پیشنهادی برای بهینه‌کردن نرخ بیت - اعوجاج در ویدئوی مقیاس‌پذیر چندنمایی را معرفی می‌کنیم. روش اختصاص نرخ بیت بهینه به ویدئوی مقیاس‌پذیر از دو مرحله تشکیل شده است. در گام نخست، نرخ بیت هر لایه باید به صورت بهینه تعیین شود. این کار با در نظر گرفتن لایه پایه به عنوان مهم‌ترین لایه انجام می‌شود. سپس، اهمیت هر لایه افزایشی نسبت به لایه پایه، به عنوان معیاری برای بهینه‌کردن نرخ بیت لایه افزایشی در نظر گرفته می‌شود. در این رساله از مفهوم اختلاف بین لایه‌ای به منظور اختصاص بهینه نرخ بیت به لایه‌های افزایشی نسبت به لایه پایه استفاده شده است.

در گام بعدی، با توجه به اهمیت هر نما در هر لایه، باید ابتدا مهم‌ترین نمای هر لایه را با توجه به کاربرد مورد نظر انتخاب کنیم. به دلیل اهمیت این نما برای کاربران در این کاربرد خاص، حداکثر نرخ بیت موجود باید به این نما اختصاص یابد. سپس نرخ بیت سایر نماها بر اساس اهمیت آنها نسبت به این نمای مهم تخصیص داده می‌شود. در این رساله از مفهوم اختلاف درون لایه‌ای برای تعیین اهمیت نماها نسبت به یکدیگر استفاده شده است.

در این رساله، روش پیشنهادی بر اساس مقیاس‌پذیری نمایی به عنوان یکی از پرکاربردترین و مشهورترین انواع مقیاس‌پذیری در ویدئوی چندنمایی مورد بررسی قرار گرفته است. اما از آنجایی که مفاهیم اختلاف درون لایه‌ای و بین لایه‌ای می‌تواند بر اساس ویژگی‌های خاص هر

نوع مقیاس‌پذیری تعریف شود، روش پیشنهادی می‌تواند به سایر مقیاس‌پذیری‌ها نیز اعمال شود. جزئیات روش پیشنهادی با وضوح بیشتر در زیربخش‌های ۷-۲-۱ و ۷-۲-۲- توضیح داده خواهد شد. نکته قابل توجه اینکه با استفاده از روش پیشنهادی موجود می‌توان مقدار اولیه برای پارامترهای کوانتیزاسیون نماهای مختلف لایه‌های پایه و افزایشی را به‌منظور کدکردن بهینه آنها استخراج کرد. با تغییر شرایط کدکردن در هر لحظه می‌توان با اجرای مجدد الگوریتم بهینه‌سازی نرخ بیت - اعوجاج، پارامتر کوانتیزاسیون مناسب را به‌دست آورد.

۷-۲-۱- بهینه‌سازی نرخ بیت - اعوجاج برای لایه‌های پایه و افزایشی

ویدئوی چندنمایی مقیاس‌پذیر در مقیاس‌پذیری نمایی

در مقیاس‌پذیری نمایی، لایه پایه حداقل تعداد نماهای مورد نیاز را بر اساس حداقل پهنای باند و توان پردازشی موجود گیرنده‌ها دارد. سپس هر گیرنده می‌تواند بر اساس قابلیت‌ها پهنای باند و توان پردازشی خود، با دریافت یک یا چند لایه افزایشی، تعداد نماهای دریافتی و زاویه دید خود را بهبود بخشد.

فرض کنید که تعداد n گیرنده با پهنای باندهای مختلف BW_1, BW_2, \dots, BW_n داریم به نحوی که $BW_1 < BW_2 < \dots < BW_n$. به این ترتیب، تعداد کل نماهای لایه پایه بر اساس BW_1 ، تعداد کل نماهای لایه افزایشی اول بر اساس BW_2 و تعداد نماهای آخرین لایه افزایشی بر اساس BW_n تعیین می‌شود.

ساختار پیش‌بینی معمول پیشنهاد شده برای ویدئوی چندنمایی در شکل ۱-۵ نشان داده شده است.

در کدکردن ویدئوی چندنمایی، ارتباط بین نماهای مختلف به‌منظور فشرده‌سازی با کارایی بالاتر استخراج می‌شود [115]. اگر اختلاف بین نمای مرجع و نمای پیش‌بینی شده کم باشد، نمای پیش‌بینی شده می‌تواند در کیفیت یکسان، با نرخ بیت کمتری کد شود. در مقیاس‌پذیری نمایی، نماهای لایه پایه، به‌صورت مستقل از نماهای لایه افزایشی کد و واکنش می‌شوند. اما نماهای لایه افزایشی می‌توانند از نماهای لایه پایه به‌عنوان مرجع و به‌منظور فشرده‌سازی بیشتر استفاده کنند. بنابراین، اختلاف بین‌نمایی بین نماهای مختلف لایه پایه و افزایشی اثر مستقیمی روی کارایی فشرده‌سازی و نرخ بیت نهایی لایه افزایشی دارد. به‌دلیل وجود وابستگی زیاد بین نرخ بیت لایه‌های مختلف، به‌منظور بهینه‌سازی کارآمد نرخ بیت - اعوجاج، همان‌طور که در معادله‌های (۷-۵) و (۷-۶) نشان داده شده است، لایه‌های

پایه و افزایشی باید به صورت توأم در نظر گرفته شوند. همان‌طور که در فصل ۶ توضیح داده شد، کیفیت نهایی در ویدئوی چندنمایی مقیاس‌پذیر از رابطه (۷-۶) به دست می‌آید. نرخ بیت لایه‌های مختلف می‌تواند به صورت جمع نرخ بیت نماهای متناظر آنها در نظر گرفته شود:

$$R_{BL} = \sum R_{v_i} \quad v_i \in \text{Base Layer} \quad (7-7)$$

$$R_{EL} = \sum R_{v_j} \quad v_j \in \text{Enhancement Layer } n \quad (8-7)$$

در این مسأله، رابطه (۷-۶) نسبت به نرخ بیت، یکنواخت کاهشی است، بنابراین تابعی محدب^۱ است. بنابراین با محدب بودن تابع هدف، مسأله بهینه‌سازی نیز مسأله‌ای محدب خواهد بود. مزیت مسأله بهینه‌سازی محدب این است که نقطه حداقل محلی، حداقل عمومی نیز خواهد بود [124].

برای کنترل نرخ بیت، پارامتر کوانتیزاسیون به عنوان یک متغیر قابل کنترل در نظر گرفته می‌شود. بنابراین، باید مدل‌های ریاضی اعوجاج - پارامتر کوانتیزاسیون و نرخ بیت - پارامتر کوانتیزاسیون را در روابط (۷-۶)، (۷-۷) و (۸-۷) جایگزین کنیم تا از حل مسأله بهینه‌سازی، پارامترهای کوانتیزاسیون بهینه برای لایه‌های پایه و افزایشی به دست آید. برای حل مسأله بهینه‌سازی از مدل نرخ بیت - اعوجاج ارائه شده در فصل ۵ استفاده کردیم. مدل معرفی شده در رابطه (۲۵-۵) برای تخمین زدن نرخ بیت برای هر نما در ویدئوی چندنمایی استفاده می‌شود.

به منظور استخراج کردن مدل اعوجاج - پارامتر کوانتیزاسیون، میانگین مربعات خطا به عنوان تخمینی از اعوجاج در نظر گرفته شده است. رابطه بین میانگین مربعات خطا و نسبت سیگنال به نویز در معادله زیر نشان داده شده است [125]:

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2}{\text{MSE}} \right) \quad (9-7)$$

نسبت سیگنال به نویز با پارامتر کوانتیزاسیون یک رابطه خطی دارد که در معادله (۷-۱۰) نشان داده شده است، در این معادله، α و β پارامترهایی هستند که از شبیه‌سازی استخراج می‌شوند [125]:

$$\text{PSNR} = \alpha \text{QP} + \beta \quad (10-7)$$

^۱ Convex

بنابراین، مدل اعوجاج - پارامتر کوانتیزاسیون با استفاده از رابطه (۷-۹) و (۷-۱۰) می‌تواند استخراج شود. این مدل در نهایت در رابطه (۶-۷) جایگزین می‌شود تا اعوجاج نهایی بر حسب پارامتر کوانتیزاسیون برای استفاده در مسأله بهینه‌سازی استخراج شود. در این مرحله، به‌منظور سادگی و بدون اینکه کلیات مسأله تغییری کند، فرض می‌کنیم که همه نماهای داخل هر لایه، پارامتر کوانتیزاسیون یکسان دارند. با این حال، حل کردن مسأله بهینه‌سازی با مقادیر مختلف پارامتر کوانتیزاسیون برای لایه‌های پایه و افزایشی همچنان از نظر محاسباتی پیچیده است. بنابراین، در این رساله پیشنهاد می‌شود که از مفهوم اختلاف بین لایه‌ای به‌منظور ساده‌کردن مسأله بهینه‌سازی استفاده شود. همان‌طور که قبلاً توضیح داده شد، در مقیاس‌پذیری نمایی، به‌منظور افزایش کارایی فشرده‌سازی، نماهای لایه‌های افزایشی باید از روی نماهای لایه پایه و بر اساس ساختار پیش‌بینی شکل ۱-۵ پیش‌بینی شوند. بنابراین، وابستگی بیشتر بین نماهای لایه‌های پایه و افزایشی می‌تواند به فشرده‌سازی بهینه لایه‌های افزایشی منجر شود. به این معنی که، در این حالت، نماهای لایه افزایشی می‌توانند با کیفیت یکسان، از روی نماهای لایه پایه و با نرخ بیت کمتر پیش‌بینی شوند. بر اساس این بحث، از مفهوم اختلاف بین لایه‌ای به‌صورت زیر و برای یافتن ارتباط مناسب بین پارامترهای کوانتیزاسیون لایه‌های پایه و افزایشی استفاده می‌شود:

$$Q_{PEL} = Q_{PBL} + \frac{1}{\text{Inter_LD between base and enhancement layer } n} \quad (11-7)$$

بر اساس این معادله، وقتی که اختلاف بین لایه پایه و لایه افزایشی n کم باشد، پارامتر کوانتیزاسیون لایه افزایشی n بیشتر از پارامتر کوانتیزاسیون لایه پایه شده و نرخ بیت لایه افزایشی n کمتر خواهد بود. اما، به‌دلیل اختلاف بین لایه‌ای کمتر، این لایه افزایشی می‌تواند از روی لایه پایه، به‌صورت کارآمد پیش‌بینی شود و بنابراین کاهش نرخ بیت روی کیفیت نهایی اثر چندان بدی نخواهد گذاشت.

به این ترتیب، معادله (۷-۶) را می‌توان تنها با یک متغیر یعنی Q_{PBL} و تنها یک پارامتر λ به‌صورت زیر بازنویسی کرد:

$$J_{\text{total}} = D_{\text{total}} + \lambda(R_{BL} + R_{EL1} + \dots + R_{ELn}) \quad (12-7)$$

پارامتر کوانتیزاسیون بهینه برای لایه پایه از حل این معادله به‌دست می‌آید. برای حل مسأله بهینه‌سازی از شرایط بهینگی (Karush-Kuhn-Tucker (KKT در مسأله محدب استفاده

شده است [126]. شرایط بهینگی KKT شامل معادله گرادیان^۱، شرایط دوگانی^۲ و چک کردن شرایط علامت^۳ است. فرض کنید که QP^* حداقل‌کننده محلی باشد، در این صورت، معادله گرادیان به صورت زیر خواهد بود:

$$\frac{\partial J}{\partial QP^*} = 0 \quad (13-7)$$

برای همه قیده‌های مسأله بهینه‌سازی که به صورت نامعادله هستند، معادلات متناظر شرایط دوگانی نامیده می‌شوند. از روی معادله گرادیان و شرایط دوگانی، نقاطی به‌عنوان کاندید برای نقطه حداقل‌کننده، (λ^*, QP^*) استخراج می‌شود. از بین این نقاط، نقاطی که در شرایط علامت و شرایط اولیه صدق کنند، به‌عنوان نقاط حداقل‌کننده نهایی در نظر گرفته می‌شوند. این شرایط شامل موارد زیر است: $\lambda^* > 0$ و $QP^* > 0$ و $QP^* < q_{max}$ که حداکثر مقدار ممکن برای پارامتر کوانتیزاسیون است که در کدکننده تعریف شده است. به‌منظور ساده‌سازی پیچیدگی محاسباتی، به‌جای قرار دادن این شرایط در مسأله بهینه‌سازی، ابتدا نقاط کاندید برای جواب مسأله را تعیین کردیم. آنگاه اگر $QP^* < 0$ ، حداقل مقدار ممکن برای پارامتر کوانتیزاسیون و اگر $QP^* > q_{max}$ ، حداکثر مقدار ممکن برای پارامتر کوانتیزاسیون را به‌عنوان جواب مسأله انتخاب کردیم. این رویه در کارهای قبلی مانند [127] نیز استفاده شده‌است.

وقتی که مقدار بهینه برای پارامتر کوانتیزاسیون لایه پایه انتخاب شد، مقدار پارامتر کوانتیزاسیون بهینه برای لایه‌های افزایشی می‌تواند با استفاده معادله (۷-۱۱) استخراج شود.

۷-۲-۲ - بهینه‌سازی نرخ بیت - اعوجاج برای هر یک از نماهای لایه‌های

پایه و افزایشی در مقیاس‌پذیری نمایی

بر اساس بحثی که در بخش ۷-۱-۱ انجام شد، می‌توان از بهینه‌سازی لاگرانژ برای بهینه کردن نرخ بیت نماهای لایه پایه و افزایشی استفاده کرد. تابع لاگرانژ برای این بهینه‌سازی به صورت زیر خواهد بود:

$$\min J = D_L + \lambda_L \times (R_L - R_{opt.Layer}) \quad (14-7)$$

¹ Gradient equation

² Complementary slackness

³ Sign condition

که D_L و R_L مقدار اعوجاج و نرخ بیت نهایی لایه مورد نظر و $R_{opt.Layer}$ نرخ بیت بهینه‌شده لایه مورد نظر است که از مرحله قبل به‌دست آمده است. همان‌طور که در فصل ۶ توضیح داده شد، D_L مقدار اعوجاج هر لایه از جمع وزن‌دار اعوجاج نماهای لایه موردنظر به‌دست می‌آید. R_L نیز نرخ بیت نهایی لایه مورد نظر است که از جمع نرخ بیت نماهای مختلف آن لایه به‌دست می‌آید.

در این مرحله نیز از مدل نرخ بیت - اعوجاج که در فصل ۵ پیشنهاد شد، به‌منظور تخمین نرخ بیت هر نما، استفاده شده است. همانند مرحله قبل، حل این مسأله بهینه‌سازی با مقادیر متفاوت برای پارامتر کوانتیزاسیون برای نماهای مختلف لایه موردنظر از نظر محاسباتی بسیار پیچیده است. بنابراین، در این رساله پیشنهاد می‌شود که به‌منظور ساده‌سازی فرآیند بهینه‌سازی نرخ بیت - اعوجاج، ارتباط بین پارامترهای کوانتیزاسیون نماهای مختلف استخراج شده و به‌این ترتیب تعداد متغیرهای مسأله بهینه‌سازی کاهش یابد.

در این رساله پیشنهاد شده است که از مفهوم اختلاف بین‌نمایی به‌منظور یافتن ارتباط مناسب بین پارامترهای کوانتیزاسیون نماهای مختلف و ساده‌سازی مسأله بهینه‌سازی استفاده شود. از آنجایی‌که نماهای داخل هر لایه از ساختار پیش‌بینی بین‌نمایی به‌منظور افزایش کارایی فشرده‌سازی استفاده می‌کنند، اختلاف بین نمای اصلی (نمای مرجع اصلی در ساختار پیش‌بینی) و سایر نماها می‌تواند روی کارایی فشرده‌سازی اثر بگذارد. اگر اختلاف بین نمای اصلی (V_0 در شکل ۱-۵) و سایر نماهای پیش‌بینی شده کم باشد، نمای پیش‌بینی شده با کارایی بهتری از روی نمای مرجع پیش‌بینی و فشرده می‌شود. بنابراین، نرخ بیت سایر نماها می‌تواند به‌طور قابل توجهی کاهش یابد، بدون اینکه روی کیفیت نهایی تأثیر بگذارد. بنابراین، اختلاف بین نماهای مختلف هر لایه اثر مستقیمی روی کارایی پیش‌بینی بین‌نمایی و نرخ بیت نهایی نماهای هر لایه دارد.

به‌این ترتیب، پیشنهاد می‌شود که از مفهوم اختلاف بین‌نمایی به‌منظور یافتن ارتباط مناسب بین پارامترهای کوانتیزاسیون نماهای مختلف و به‌منظور ساده‌سازی مسأله بهینه‌سازی به‌صورت زیر استفاده شود:

$$QP_{vn} = QP_{v0} + \frac{1}{\text{Intra-LD between view 0 and view n}} \quad (15-7)$$

به‌این ترتیب می‌توان معادله (۷-۱۴) را تنها با استفاده از یک متغیر QP_{v0} و یک پارامتر لاگرانژ λ بازنویسی کرد. سپس، مقدار بهینه برای پارامتر کوانتیزاسیون برای نمای اصلی با حل مسأله بهینه‌سازی و با استفاده از شرایط KKT همان‌طور که در مرحله قبل توضیح داده

شد، استخراج می‌شود. در نهایت، مقدار بهینه برای پارامتر کوانتیزاسیون برای سایر نماها با استفاده از معادله (۷-۱۵) استخراج می‌شود.

۷-۳ - ارزیابی روش پیشنهادی

به منظور ارزیابی کارایی روش پیشنهادی برای بهینه‌سازی نرخ بیت - اعوجاج، آزمایش‌های مختلفی انجام شده است. نتایج این آزمایش‌ها کارایی روش ارائه شده در اختصاص نرخ بیت مناسب به لایه‌های پایه و افزایشی و محتوای آنها را نشان می‌دهد. به این منظور، ویدئوهای مختلفی با کدکننده ویدئوی چندنمایی استاندارد H.264/AVC نسخه ۸/۵ کد شده است [105]. جدول ۷-۱ خصوصیات این ویدئوها را نشان می‌دهد. نکته قابل ذکر اینکه در همه ویدئوهای آزمایش شده در این جدول، دوربین‌ها ثابت هستند. برای انجام آزمایش‌ها یک لایه پایه و یک لایه افزایشی در نظر گرفته شده و ساختار پیش‌بینی شکل ۱-۵ برای کدکردن ویدئوها استفاده شده است.

برای ارزیابی‌ها نرخ بیت - اعوجاج از معیار Bjøntegaard-Delta bitrate (BD-bitrate) [101] استفاده شده است. در نتایج گزارش شده، اعداد منفی نشان‌دهنده کاهش مقدار نرخ بیت و اعداد مثبت نشان‌دهنده افزایش مقدار نرخ بیت نسبت به حالت مرجع است.

جدول ۷-۱ - خصوصیات ویدئوهای آزمایش شده

تعداد نماها	تعداد فریم‌ها	وضوح تصویر	ویدئو
۸	۱۰۰	۱۰۲۴×۷۶۸	Ballet
۸	۱۰۰	۱۰۲۴×۷۶۸	Break-dancer
۷	۱۰۰	۱۰۲۴×۷۶۸	Balloons
۷	۱۰۰	۱۰۲۴×۷۶۸	Kendo

روش پیشنهادی روی نرم‌افزار مرجع کدکننده ویدئوی چندنمایی استاندارد H.264/MVC پیاده‌سازی شده است. در این نرم‌افزار از بهینه‌سازی نرخ بیت - اعوجاج استاندارد H.264 استفاده شده است. بنابراین، نتایج استخراج شده نشان‌دهنده کاهش نرخ بیت (در کیفیت یکسان) نسبت به بهینه‌سازی نرخ بیت - اعوجاج استفاده شده در استاندارد H.264 به-عنوان مرجع است. به منظور استخراج نتایج مرجع، روش بهینه‌سازی پیشنهادی ما غیرفعال بوده و تنها بهینه‌سازی نرخ بیت - اعوجاج مرجع استاندارد H.264 فعال است. در این حالت، پارامترهای کوانتیزاسیون یکسان برای نماهای لایه‌های پایه و افزایشی در نظر گرفته شده است. برای استخراج نتایج روش پیشنهادی نیز از پارامترهای کوانتیزاسیون استخراج شده از حل مسأله بهینه‌سازی در روش پیشنهادی برای کدکردن نماهای مختلف لایه‌های

پایه و افزایشی استفاده شده است. تعداد نماهای اختصاص داده شده به لایه های پایه و افزایشی و مقادیر پارامتر کوانتیزاسیون برای نماهای آنها در روش پیشنهادی و روش مرجع برای ویدئوهای آزمایش شده و به ازای یک نرخ بیت نهایی معین در جدول ۷-۲ نشان داده شده است.

جدول ۷-۲ - نماهای اختصاص داده شده به لایه های پایه و افزایشی و مقادیر پارامتر کوانتیزاسیون برای روش پیشنهادی و روش مرجع برای ویدئوی Ballet

لایه پایه		لایه افزایشی		نرخ بیت	نرخ بیت	ویدئو
نما (۰-۱-۲-۳)		نما (۴-۵-۶-۷)				
پارامتر کوانتیزاسیون		پارامتر کوانتیزاسیون				Ballet
حالت مرجع	روش پیشنهادی	حالت مرجع	روش پیشنهادی	۱۲۰۰ (کیلوبیت بر ثانیه)	۲۲۰۰ (کیلوبیت بر ثانیه)	
نمای ۰ = ۲۵	نمای ۰ = ۱۹	نمای ۴ = ۲۵	نمای ۴ = ۲۶			
نمای ۱ = ۲۵	نمای ۱ = ۲۵	نمای ۵ = ۲۵	نمای ۵ = ۳۰			
نمای ۲ = ۲۵	نمای ۲ = ۲۶	نمای ۶ = ۲۵	نمای ۶ = ۳۰			
نمای ۳ = ۲۵	نمای ۳ = ۲۷	نمای ۷ = ۲۵	نمای ۷ = ۳۲			

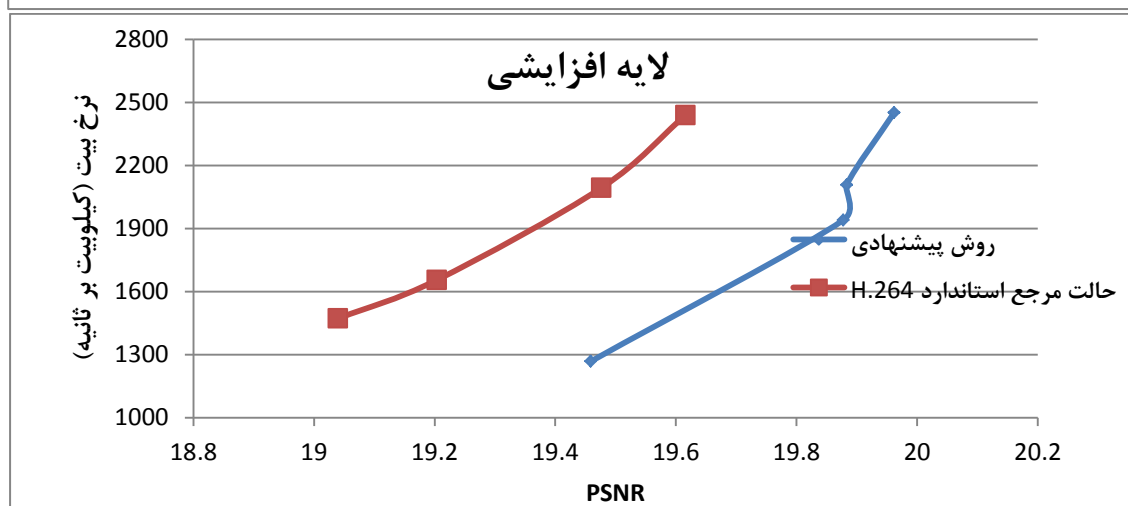
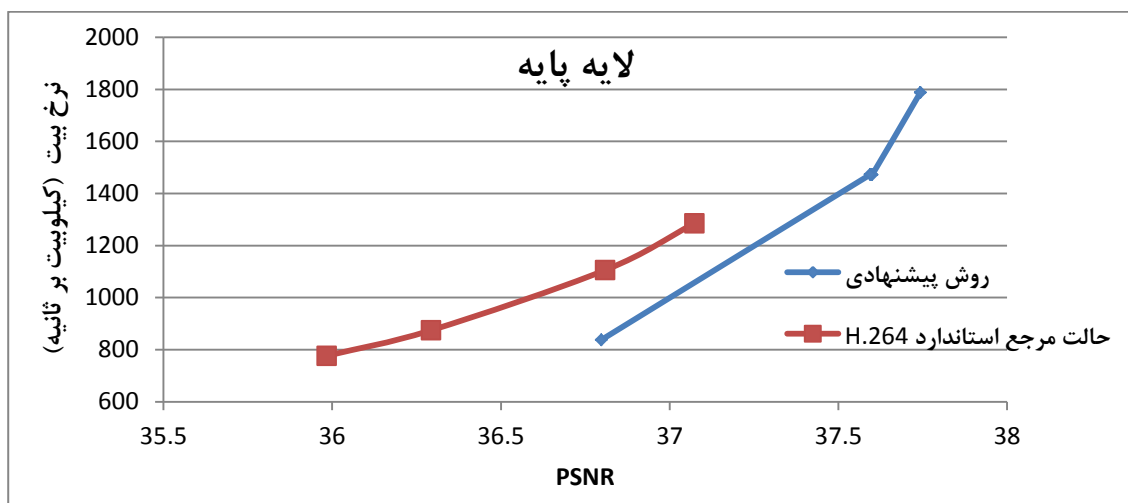
مقادیر مناسب برای پارامترهای کوانتیزاسیون به ازای ۴ نرخ بیت نهایی متفاوت استخراج شد. سپس، هر نما از لایه پایه و افزایشی با پارامترهای کوانتیزاسیون مناسب برای نماهای آنها در حالت مرجع و حالت پیشنهادی کد شدند. جدول ۷-۳ کارایی فشرده سازی برای روش بهینه سازی نرخ بیت - اعوجاج پیشنهادی نسبت به حالت مرجع را نشان می دهد. بر اساس نتایج این جدول، نرخ بیت نماها با استفاده از روش پیشنهادی به طور محسوسی کاهش می یابد.

جدول ۷-۳ - کارایی فشرده سازی روش بهینه سازی نرخ بیت - اعوجاج پیشنهادی در مقایسه با حالت مرجع

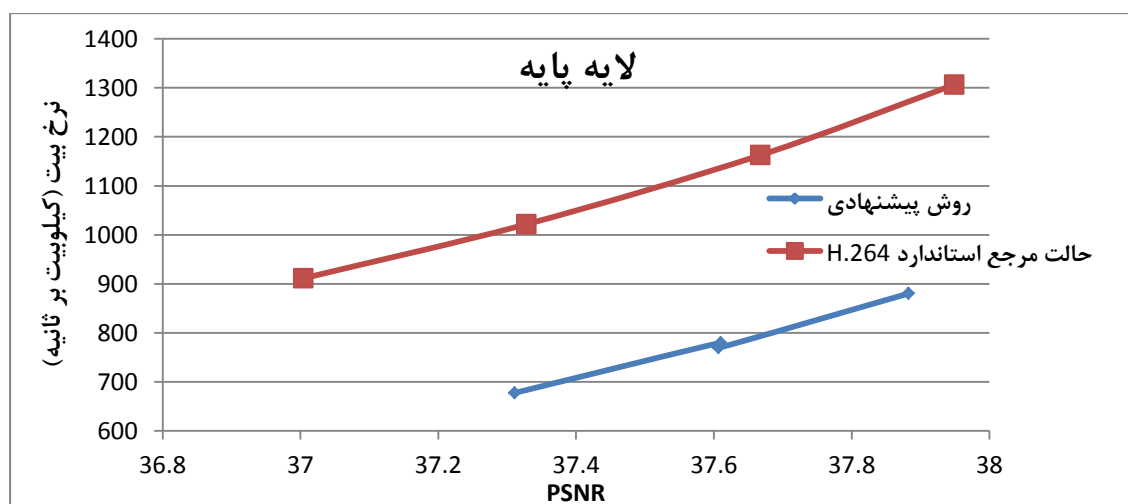
استاندارد H.264

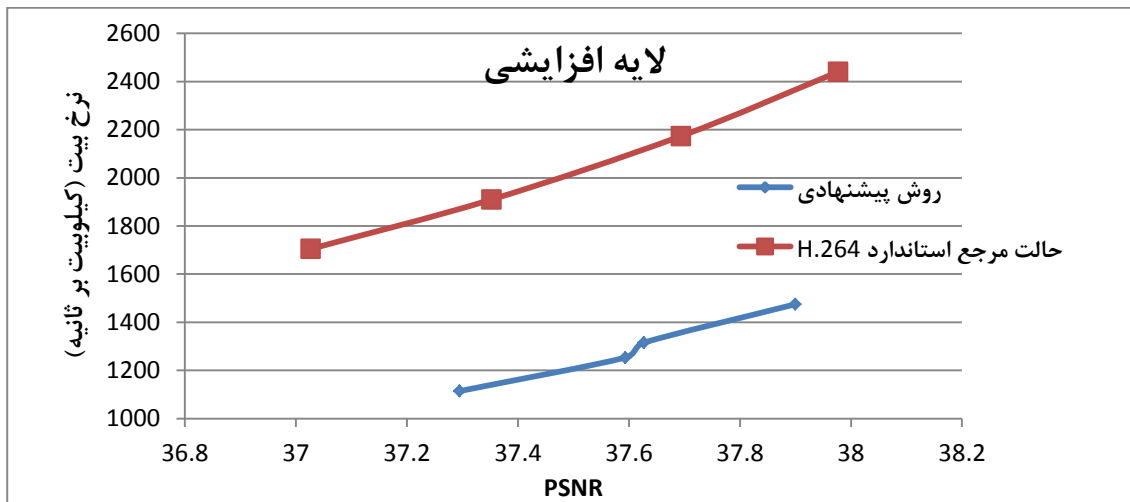
ویدئو	لایه پایه	لایه افزایشی
	BD-bitrate	BD-bitrate
Ballet	-۱۴/۹٪	-۴۲/۶٪
Break-dancer	-۲۹/۵٪	-۳۹/۱٪
Balloons	-۲۸/۱٪	-۵۸/۷٪
Kendo	-۵/۴٪	-۴۶/۶٪
میانگین	-۱۹/۴۷	-۴۶/۷۵

به منظور وضوح بیشتر، نمودارهای متناظر برای نرخ بیت - اعوجاج لایه پایه و افزایشی برای ویدئوهای مختلف در شکل ۷-۱ نشان داده شده است.

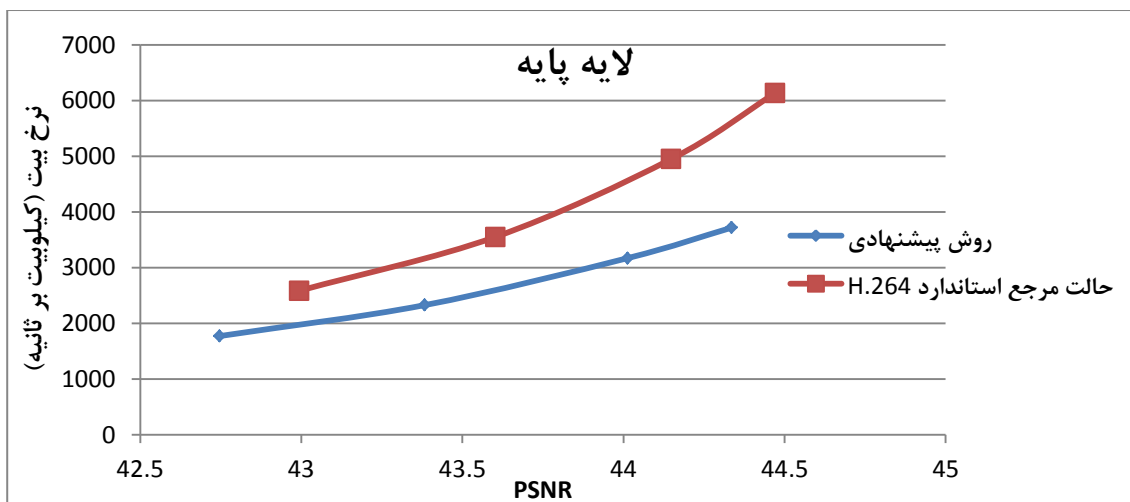


(الف)

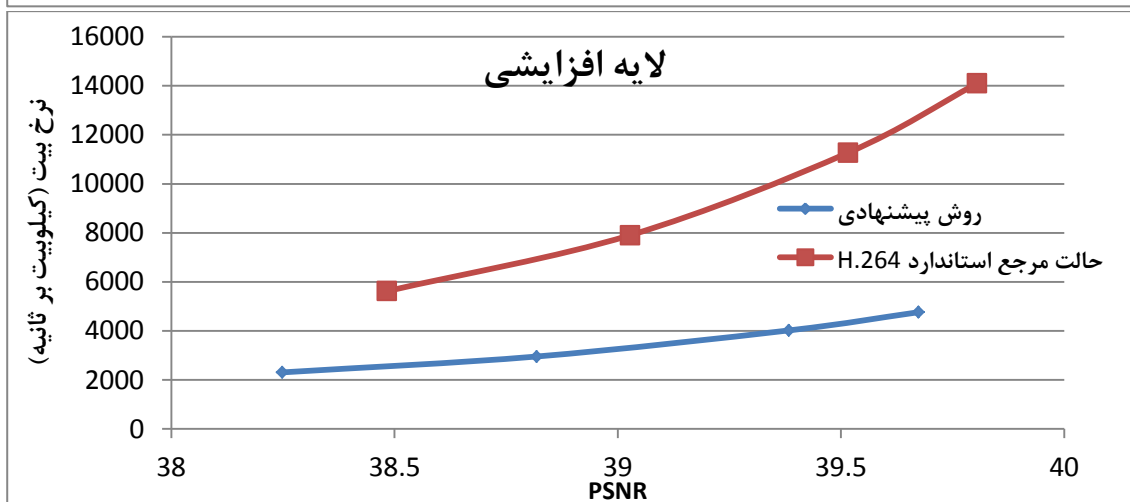


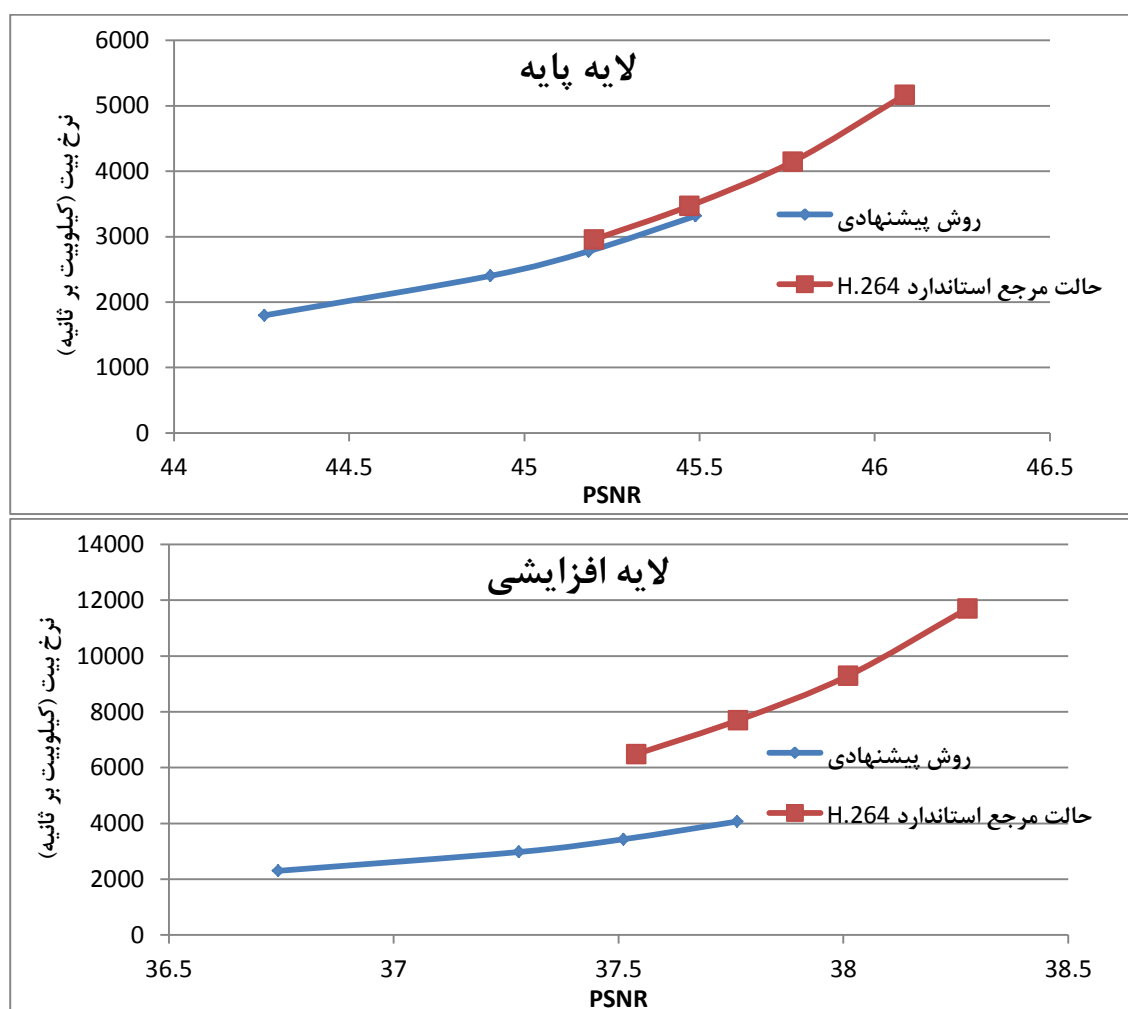


(ب)



(ج)





(د)

شکل ۷-۱ - منحنی نرخ بیت - اعوجاج برای روش بهینه‌سازی نرخ بیت - اعوجاج پیشنهادی و حالت مرجع استاندارد H.264 برای لایه‌های پایه و افزایشی (الف) Ballet (ب) Break-dancer (ج) Kendo (د) Balloons

همان‌طور که در بخش قبل توضیح داده شد، مفاهیم اختلاف بین‌لایه‌ای و درون‌لایه‌ای، ویژگی‌های اصلی روش ارائه‌شده هستند. بنابراین، میزان بهبود کارایی فشرده‌سازی به مقادیر اختلاف بین‌لایه‌ای و درون‌لایه‌ای بستگی دارد. تغییرات زیاد کارایی فشرده‌سازی برای ویدئوهای مختلف که در جدول ۷-۳ گزارش شده است، به دلیل تفاوت در مقادیر اختلاف بین‌لایه‌ای و درون‌لایه‌ای ویدئوهای مختلف است. بنابراین، به منظور نشان دادن دقیق‌تر میزان بهبود کارایی فشرده‌سازی روش پیشنهادی، نماهای مختلف را به صورت تصادفی به لایه‌های پایه و افزایشی تخصیص دادیم. سپس، روش پیشنهادی را به منظور استخراج مقادیر بهینه پارامتر کوانتیزاسیون برای نماهای لایه‌های پایه و افزایشی اعمال کردیم. به این ترتیب، مقادیر متفاوت اختلاف بین‌لایه‌ای و درون‌لایه‌ای در تحلیل میزان کارایی فشرده‌سازی روش ارائه شده، مورد بررسی قرار گرفت. با این روش کارایی روش

ارائه‌شده می‌تواند به صورت دقیق تری اثبات شود. چهار ترتیب مختلف برای اختصاص دادن نماهای متفاوت هر ویدئو به لایه‌های پایه و افزایشی و مقادیر اختلاف بین لایه‌ای و درون-لایه‌ای متناظر با آنها در جدول ۴-۷ نشان داده شده است.

جدول ۴-۷ - نماهای اختصاص داده شده به لایه‌های پایه و افزایشی و مقادیر اختلاف بین لایه‌ای و درون لایه‌ای متناظر

اختلاف بین لایه‌ای	لایه افزایشی		لایه پایه		ویدئو
	اختلاف درون لایه-ای	نماها	اختلاف درون لایه-ای	نماها	
۰/۸۹	۰/۱۳	۷-۶-۵-۳	۰/۱۵	۳-۲-۱-۰	Ballet
۰/۱۶	۰/۱۵	۶-۵-۳-۰	۰/۱۳	۲-۴-۱-۰	
۰/۱۵	۰/۱۱	۷-۴-۲-۱	۰/۱۲	۶-۵-۳-۰	
۰/۱۳	۰/۱۵	۳-۲-۱-۰	۰/۱۸	۴-۵-۶-۷	
۰/۸۷	۰/۱۱	۷-۶-۵-۴	۰/۰۷	۳-۲-۱-۰	Break-dancer
۰/۸۶	۰/۱۱	۷-۶-۵-۳	۰/۰۸	۴-۲-۱-۰	
۰/۹۷	۰/۰۹	۵-۴-۳-۲	۰/۰۶	۷-۶-۱-۰	
۰/۹۰	۰/۰۸	۶-۴-۲-۰	۰/۰۷	۷-۵-۳-۱	
۰/۱۰	۰/۱۲	۶-۵-۳-۴	۰/۱۳	۲-۱-۰	Balloons
۰/۱۳	۰/۱۳	۶-۵-۳-۱	۰/۱۱	۴-۲-۰	
۰/۱۲	۰/۱۳	۶-۵-۲-۱	۰/۰۹	۰-۳-۴	
۰/۱۱	۰/۱۲	۶-۵-۴-۱	۰/۱۰	۰-۲-۳	
۰/۲۰	۰/۰۲	۶-۵-۳-۴	۰/۵۲	۲-۱-۰	Kendo
۰/۲۲	۰/۰۶	۶-۵-۳-۱	۰/۵۱	۴-۲-۰	
۰/۱۹	۰/۰۶	۶-۵-۲-۱	۰/۲۶	۰-۳-۴	
۰/۱۸	۰/۰۶	۶-۵-۴-۱	۰/۲۶	۰-۲-۳	

جدول ۴-۷ میانگین کاهش نرخ بیت را برای روش بهینه‌سازی نرخ بیت - اعوجاج پیشنهادی نشان می‌دهد. برای ویدئوهای آزمایش‌شده روش پیشنهادی به طور میانگین ۲۴/۴۲٪ کاهش نرخ برای لایه پایه و ۴۱/۸۲٪ کاهش نرخ بیت برای لایه افزایشی نسبت به حالت مرجع را نشان می‌دهد.

جدول ۵-۷ - مقایسه کارایی فشرده‌سازی برای روش بهینه‌سازی نرخ بیت - اعوجاج پیشنهادی نسبت به حالت مرجع استاندارد H.264 برای چهار نمای متفاوت اختصاص داده شده به لایه‌های پایه و افزایشی

لایه پایه + لایه افزایشی	لایه پایه	ویدئو
BD-bitrate	BD-bitrate	
-%۴۱/۸۷	-%۲۲/۱۲	Ballet
-%۴۳/۵۲	-%۲۹/۲۵	Break-dancer
-%۳۴/۳۰	-%۱۲/۸۰	Kendo
-%۴۷/۶	-%۳۳/۵۲	Balloons
-۴۱/۸۲	-۲۴/۴۲	میانگین

در بهینه‌سازی نرخ بیت - اعوجاج برای ویدئوی چندنمایی مقیاس‌پذیر، افزایش پیچیدگی محاسباتی بیشتر به دلیل فرآیند جستجوی پارامتر کوانتیزاسیون بهینه برای لایه‌های مختلف و نماهای آنها است، به نحوی که شرایط KKT برقرار باشد. در روش پیشنهادی در این رساله، فرآیند بهینه‌سازی یک‌بار برای یافتن پارامتر کوانتیزاسیون برای لایه پایه انجام می‌شود، مقادیر پارامتر کوانتیزاسیون برای سایر لایه‌های افزایشی با استفاده از این مقدار به دست آمده و مفهوم اختلاف بین لایه‌ای تعیین می‌شود. همچنین، برای هر لایه، مسأله بهینه‌سازی تنها یک‌بار برای نمای اول آن لایه حل شده و پارامتر کوانتیزاسیون بهینه تعیین می‌شود. پارامتر کوانتیزاسیون برای سایر نماها با استفاده از این مقدار به دست آمده و مفهوم اختلاف درون لایه‌ای تعیین می‌شود.

۴-۷ - جمع‌بندی

در این رساله، یک روش جدید برای بهینه‌سازی نرخ بیت - اعوجاج برای ویدئوی چندنمایی مقیاس‌پذیر ارائه شده است. این روش، دو مرحله متفاوت برای انتخاب مقدار مناسب برای پارامتر کوانتیزاسیون برای هر لایه و برای محتوای لایه‌ها در نظر می‌گیرد. نتایج شبیه‌سازی نشان می‌دهد که کارایی فشرده‌سازی روش پیشنهادی نسبت به حالت مرجع استفاده شده در استاندارد H.264، به ترتیب برای لایه پایه و لایه پایه به علاوه افزایشی در مقیاس‌پذیری نمایی، ۲۴/۴۲٪ و ۴۱/۸۲٪ است.

فصل هشتم

نتیجه‌گیری و کارهای آینده

با رشد تقاضا برای ویدئوهای سه‌بعدی، تلاش‌ها برای گنجاندن آن در نسل بعدی استانداردهای فشرده‌سازی ویدئو و کاربردهای جریان‌سازی و پخش داده‌ها رو به گسترش است. مقیاس‌پذیری یکی از روش‌های مرسوم برای غلبه بر مشکلات ارسال داده‌ها با حجم بالا روی کانال‌های مستعد خطا و برای گیرنده‌هایی با قابلیت‌های متفاوت پهنای باند و توان پردازشی است. البته هنوز مشکلات زیادی در راه عملی ساختن استفاده از مقیاس‌پذیری برای کاربردهای گوناگون ویدئوی چندنمایی و سه‌بعدی وجود دارد. در این رساله، یک چارچوب قابل انطباق برای استفاده از مقیاس‌پذیری برای ویدئوی چندنمایی سه‌بعدی برای کاربردهای متفاوت آن ارائه شده است. قسمت‌های مختلف چارچوب پیشنهادی ویژگی‌های کاربرد موردنظر و نیز خصوصیات ویدئوی مقیاس‌پذیر چندنمایی سه‌بعدی مانند اختلاف بین لایه‌ای و درون لایه‌ای را به صورت توأم در نظر می‌گیرند. در نظر گرفتن این موارد منجر به بالا رفتن کیفیت تجربی و دریافتی کاربر بسته به کاربرد موردنظر شده و چارچوب پیشنهادی نیز قابل انطباق با کاربردهای مختلف ویدئوی چندنمایی سه‌بعدی خواهد بود.

قسمت‌های اصلی این چارچوب شامل تعیین نوع مناسب مقیاس‌پذیری بر حسب کاربرد موردنظر، اختصاص داده‌های مناسب به لایه‌های پایه و افزایشی، استخراج مدل نرخ بیت - اعوجاج برای ویدئوی چندنمایی و معیار ارزیابی کمی کیفیت برای ویدئوی چندنمایی مقیاس‌پذیر و در نهایت تعیین نرخ بیت بهینه برای هر یک از این لایه‌ها و محتوای آنها است. روش‌شناسی ارائه‌شده برای هر یک از این قسمت‌ها از مفاهیم اختلاف بین لایه‌ای و درون لایه‌ای که بر حسب نوع مقیاس‌پذیری انتخاب‌شده تعریف می‌شود، استفاده می‌کنند.

در این چارچوب، ابتدا انواع مناسب مقیاس‌پذیری بر حسب کاربردهای مختلف ویدئوی چندنمایی با استفاده از نظریه مبنایی استخراج شد. روش‌شناسی ارائه‌شده برای این بخش، مفاهیم مرتبط با مقیاس‌پذیری ویدئوی چندنمایی سه‌بعدی ارائه‌شده در منابع را جمع‌آوری کرده و با استفاده از آنها به تعاریف جدید برای مقیاس‌پذیری بر حسب کاربرد مورد نظر ویدئوی چندنمایی سه‌بعدی می‌رسد. اهمیت مقیاس‌پذیری‌های ارائه‌شده با ارائه مثال‌های عملی از کاربردهای ویدئوی چندنمایی مورد بررسی قرار گرفته و و لزوم استفاده از آنها اثبات شده است. با استفاده از مقیاس‌پذیری انتخاب‌شده در این قسمت، مفاهیم اختلاف بین لایه‌ای و درون لایه‌ای تعریف شده و برای اختصاص مناسب داده‌ها به لایه‌های پایه و افزایشی در قسمت بعدی چارچوب پیشنهادی استفاده می‌شوند. نتایج شبیه‌سازی این قسمت از چارچوب ارائه‌شده نشان می‌دهد که روش پیشنهادی در مقایسه با روش‌های موجود کیفیت بالاتری داشته و محدودیت‌های کاربردهای خاص ویدئوی چندنمایی و گیرنده‌های آنها را نیز مورد توجه قرار می‌دهد.

در ادامه و به‌منظور حل مسأله تخصیص نرخ بیت بهینه به لایه‌های مختلف و محتوای آنها، ابتدا یک مدل نرخ بیت - اعوجاج برای ویدئوی چندنمایی و بر حسب ویژگی‌های آن ارائه شد. این مدل از مفاهیم پیش‌بینی بین‌نمایی و درون‌نمایی که در ویدئوی چندنمایی و به‌منظور افزایش کارایی فشرده‌سازی استفاده می‌شود، بهره می‌گیرد. نتایج شبیه‌سازی‌ها برای این قسمت از رساله نشان می‌دهد که نرخ بیت پیش‌بینی‌شده توسط این مدل، با دقت بالایی به نرخ بیت واقعی نماهای کد شده نزدیک است.

از سوی دیگر و به‌منظور ارزیابی کیفیت ویدئوی چندنمایی مقیاس‌پذیر، نیاز به معیاری برای ارزیابی کمی کیفیت این ویدئوها داریم. از آنجایی که چنین معیاری وجود نداشت، ارائه این معیار به‌عنوان قسمت بعدی چارچوب پیشنهادی مورد توجه قرار گرفت. برای ارزیابی معیار ارزیابی کمی کیفیت ارائه‌شده، نتایج ارزیابی کمی با نتایج حاصل از ارزیابی بصری مقایسه شد. آزمایش‌های انجام‌شده نشان می‌دهد که همبستگی بالایی بین نتایج حاصل از ارزیابی بصری و نتایج حاصل از معیار ارزیابی کمی کیفیت پیشنهادی وجود دارد.

در نهایت به‌عنوان آخرین بخش چارچوب پیشنهادی، با استفاده از مدل نرخ بیت - اعوجاج پیشنهادی برای ویدئوی چندنمایی و معیار ارزیابی کمی کیفیت ارائه شده، مسأله بهینه‌سازی نرخ بیت - اعوجاج حل شده و نرخ بیت بهینه به لایه‌ها و محتوای آنها تخصیص داده می‌شود. نتایج شبیه‌سازی این قسمت نیز نشان می‌دهد که روش ارائه‌شده نسبت به روش بهینه‌سازی استفاده‌شده در نرم‌افزار مرجع H.264، کارایی فشرده‌سازی بالاتری دارد.

۸-۱ - کارهای آینده

استاندارد کد کردن ویدئوی HEVC که به تازگی و بعد از استاندارد H.264/AVC ارائه شده است، قابلیت‌های جدیدی برای کد کردن ویدئوی تک‌نمایی و چندنمایی و سه بعدی ارائه کرده است. در زمینه کد کردن ویدئوی چندنمایی، علاوه بر مفهوم پیش‌بینی بر اساس اختلاف بین‌نمایی، پارامترهای حرکت بین‌نمایی و پیش‌بینی بین‌نمایی باقی‌مانده‌ها نیز برای کد کردن نماها از روی یکدیگر در این استاندارد اضافه شده است. علاوه بر این، برای کد کردن ویدئوی عمق، مودهای جدید برای کد کردن intra، الگوریتم‌های جبران حرکت و کد کردن بردار حرکت تغییر یافته نیز به‌منظور افزایش کارایی کد کردن در این استاندارد در نظر گرفته شده است.

ویدئوهایی با وضوح تصویر بالا از جمله کاربردهای مهم استاندارد HEVC هستند. استفاده از این ویدئوها در حالت چندنمایی و سه‌بعدی حجم داده‌های موجود را به شدت افزایش می‌دهد. علاوه بر این، فرمت رشته‌بیت تولید شده توسط این استاندارد به‌گونه‌ای است که

می‌توان بخشی‌هایی از رشته بیت مرتبط با فرمت‌های ویدئوی مرسوم دوبعدی، ویدئوی برجسته‌نما و ویدئوی چندنمایی به‌علاوه عمق را از روی یک رشته بیت استخراج و واکنش کرد. به این ترتیب استفاده از مقیاس‌پذیری در کاربردهای متفاوت استاندارد HEVC و به‌منظور کاهش حجم داده‌های ارسالی بسیار مفید خواهد بود.

از آنجایی که تغییرات اضافه‌شده به این استاندارد نسبت به استاندارد H.264/AVC در جهت استخراج دقیق‌تر اختلاف بین‌نمایی است و با توجه به اینکه اختلاف بین‌نما و لایه‌ها از مفاهیم اصلی بخش‌های مختلف چارچوب ارائه شده در این رساله است، اعمال چارچوب پیشنهادی برای استاندارد جدید کدکردن ویدئوی چندنمایی می‌تواند کارایی بالایی برای کاربردهای خاص این استاندارد در حالت مقیاس‌پذیر داشته باشد.

بر اساس توضیحات بالا، از جمله کارهای آتی این رساله می‌توان به تعمیم روش‌های ارائه شده در بخش‌های مختلف رساله به استاندارد HEVC باشد.

همان‌طور که توضیح داده شد، مفاهیم اختلاف بین‌نمایی و درون‌نمایی در بخش‌های مختلف رساله از جمله روش‌شناسی پیشنهادی برای اختصاص داده‌های مناسب برای لایه‌های مختلف، استخراج مدل نرخ بیت - اعوجاج، استخراج معیار ارزیابی کمی کیفیت برای ویدئوی چندنمایی مقیاس‌پذیر و نیز برای بهینه‌سازی نرخ بیت لایه‌های مختلف و محتوای آنها استفاده شده است. دلیل استفاده از این مفاهیم، اثر آنها روی کارایی فشرده‌سازی و نیز کیفیت نهایی دریافتی توسط کاربر است. از آنجایی که در استاندارد HEVC، علاوه بر این مفاهیم از مفاهیم دیگری همچون پیش‌بینی حرکت بین‌نمایی و پیش‌بینی بین‌نمایی باقی‌مانده‌ها نیز به‌منظور بالابردن کارایی فشرده‌سازی و کیفیت دریافتی استفاده می‌کند، این مفاهیم می‌تواند در قسمت‌های مختلف چارچوب پیشنهادی وارد شود.

به‌عنوان مثال، در روش‌شناسی ارائه‌شده در این رساله به‌منظور استخراج مدل نرخ بیت - اعوجاج در سطح نما برای ویدئوی چندنمایی سه‌بعدی عنوان شده است که نرخ بیت هر نما با مفاهیم وابستگی بین‌نمایی و دورن‌نمایی ارتباط دارد. سپس پارامترهای مناسب برای توصیف این مفاهیم و رابطه نرخ بیت هر نما با آنها استخراج شد. به‌منظور تعمیم دادن روش‌شناسی پیشنهادی برای استاندارد HEVC، می‌توان پارامترهای مرتبط با مفاهیم پیش‌بینی حرکت بین‌نمایی و پیش‌بینی باقی‌مانده‌ها و نیز رابطه نرخ بیت - اعوجاج هر نما با این پارامترها را استخراج کرد. به‌طریق مشابه، این مفاهیم می‌تواند در سایر قسمت‌های چارچوب پیشنهادی نیز مورد استفاده قرار گیرند.

فصل نهم

مراجع

- [1] L. Do, S. Zinger, P. H. N. de With, "Conversion of free-viewpoint 3D multi-view video for stereoscopic displays", Proceedings of the 2010 IEEE International Conference on Multimedia and Expo (ICME), Suntec City, Singapore, July 2010.
- [2] K. Muller, P. Merkle, T. Wiegand, "3-D video representation using depth maps", Proceedings of the IEEE, vol. 99, issue 4, pp. 643-656, April 2011.
- [3] S. Cho, K. Yun, C. Ahn, S. Lee, "Disparity-compensated stereoscopic video coding using the MAC in MPEG-4", ETRI Journal, vol. 27, pp. 326-329, June 2005.
- [4] M. Tanimoto, "Overview of FTV (free-viewpoint television)", Proceedings of the 2009 IEEE international conference on Multimedia and Expo (ICME), New York, NY, July 2009.
- [5] "Introduction to multi-view video coding", ISO/IEC JTC1/SC29/WG11, N7328, July 2005.
- [6] N. A. Dodgson, "Autostereoscopic 3D Displays", Computer Journal Published by the IEEE Computer Society, vol 38, issue 8, pp. 31-36, August 2005.
- [7] ITU-T Rec. H.264 / ISO/IEC, Annex H, "Multiview video coding", March 2010.
- [8] A. Vetro, T. Wiegand and G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard", Proceedings of the IEEE, vol. 99, issue 4, pp. 626-642, April 2011.
- [9] <http://www.movidius.com/best-mobile-3d/>, last access on September, 2014.
- [10] <http://research.nokia.com/page/4988>, last access on September, 2014.
- [11] <http://www.htc.com/www/smartphones/htc-evo-3d>, last access on September, 2014.
- [12] <http://www.nintendo.com/3ds>, last access on September, 2014.
- [13] <http://www.movidius.com/movidius-brings-hd-3d-to-mobile-devices>, last access on September, 2014.
- [14] N. Ozbek, A.M. Tekalp, "Scalable Multi-View Video Coding for Interactive 3DTV", IEEE International Conference on Multimedia and Expo, Toronto, Ca, July 2006.
- [15] Yo-Sung Ho, Kwan-Jung Oh, "Overview of multi-view video coding", 14th International Workshop on Systems Signals and Image Processing and 6th EURASIP Conference focused on Speech and Image Processing Multimedia Communications and Services, Maribor, June 2007.
- [16] M. Drose, C. Clemens, T. Sikora, "Extending Single-View Scalable Video Coding to Multi-View Based on H.264/AVC", IEEE International Conference on Image Processing, Atlanta, GA, October 2006.
- [17] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard", IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, issue 9, pp. 1103 – 1112, 2007.
- [18] D. Grois, E. Kaminsky, and O. Hadar, "Dynamically adjustable and scalable ROI video coding", IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Shanghai, China, March 2010.
- [19] Sh. Shimizu, M. Kitahara, H. Kimata, K. Kamikura, and Y. Yashima, "View scalable multiview video coding using 3-D warping with depth map", IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, issue 11, pp. 1485 – 1495, November 2007.

- [20] A. Vetro, A.M. Tourapis, K. Muller, and T. Chen, "3D-TV content storage and transmission", *IEEE Transactions on Broadcasting*, vol. 57, issue 2, pp. 384 – 394, 2011.
- [21] M. Hefeeda, and C.H. Hsu, "Rate-Distortion optimized streaming of fine-grained scalable video sequences", *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 4, issue 1, January 2008.
- [22] L.S. Karlsson, M. Sjostrom, "Layer assignment based on depth data distribution for multiview-plus-depth scalable video coding", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, issue 6, pp. 742 - 754 , June 2011.
- [23] A. Deng, W.J. Tsaur, J.H Li, H.C. Tsai, "Basic unit layer rate control for video security", *International Computer Symposium*, Taiwan, December 2010.
- [24] P. An, L. Shen, Q. Zhang, Z. Zhang, "Rate control algorithm for multi-view video coding based on correlation analysis", *Symposium on Photonics and Optoelectronics*, China, April 2009.
- [25] M. Cordina, C.J. Debono, "A novel view-level target bit rate distribution estimation technique for real-time multi-view video plus depth", *International Conference on Multimedia and Expo*, Australia, July 2012.
- [26] S. Park, D. Sim, "An efficient rate-control algorithm for multi-view video coding", *International Symposium on Consumer Electronics*, Japan, May 2009.
- [27] Pei-Jun Lee, Yu-Chen Lai, "Vision perceptual based rate control algorithm for multi-view video coding", *International Conference on System Science and Engineering*, Macao, June 2011.
- [28] Y. Liu, Q. Huang, S. Ma, D. Zhao, W. Gao, S. Ci, and H. Tang, "A novel rate control technique for multiview video plus depth based 3D video coding", *IEEE Transactions on Broadcasting*, vol. 57, issue 2, pp. 562 - 571, June 2011.
- [29] T. Yan, L. Shen, P. An, H. Wang, Z. Zhang, "Frame-layer rate control algorithm for multi-view video coding", *World Summit on Genetic and Evolutionary Computation*, China, June 2009.
- [30] Q. Zheng, M. Yu, G. Jiang, F. Shao, Z. Peng, "Rate control for multi-view video coding based on statistical analysis and frame complexity estimation", *Journal of Software*, vol. 7, pp. 2591-2598, Nov. 2012.
- [31] A. Fiandrotti, J. Chakareski, P. Frossard, "Popularity-aware rate allocation in multiview video", *Proceedings of SPIE*, vol. 7744, July 2010.
- [32] L. Feng, X. Jie, F. Jingjing, L. Qiongjie, "Efficient rate control algorithm for multi-view video coding", *China Communications*, vol. 8, issue 3, pp. 83-89, 2011.
- [33] B.B. Vizzotto, B. Zatt, M. Shafique, S. Bampi, J. Henkel, "A model predictive controller for frame-level rate control in multiview video coding", *International Conference on Multimedia and Expo*, Australia, July 2012.
- [34] Lili Zhou , Gang Wu, Yan He, Tao Wang, QingWen Chen, Xiaopeng Fan, Wen Gao, "A new just-noticeable-distortion model combined with the depth information and its application in multi-view video coding", *Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, Piraeus, July 2012.
- [35] Tao Yan, Anyuan Deng , Changshou Deng, Guochao Li, "A joint bit allocation scheme for MV", *4th International Congress on Image and Signal Processing (CISP)*, Shanghai, October 2011.

- [36] V. Davidoiu, T. Maugey, B. Pesquet-Popescu, P. Frossard, "Rate distortion analysis in a disparity compensated scheme", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, May 2011.
- [37] Feng Shao, Gangyi Jiang, Weisi Lin, Mei Yu, Qionghai Dai, "Joint bit allocation and rate control for coding multi-view video plus depth based 3D video", IEEE Transactions on Multimedia, vol. 15, issue 8, pp. 1843 – 1854, December 2013.
- [38] Pei-Jun Lee, Yu-Chen Lai, "Perceptual awareness rate control for multi-view video Encoder in Stereoscopic Display", Journal of Display Technology, vol. 9, issue: 7, pp. 552 – 560, July 2013.
- [39] Yi Liao, Mei Yu, Xiaodong Wang, Gangyi Jiang, Zongju Peng, Feng Shao, "Rate control for multi-view video coding based on visual perception", Journal of Theoretical and Applied Information Technology, vol. 49, no.3, pp. 959 – 964, March 2013.
- [40] Bruno Boessio Vizzotto, Bruno Zatt, Muhammad Shafique, Sergio Bampi, Jorg Henkel, "Model predictive hierarchical rate control with markov decision process for multiview video coding", IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 12, pp. 2090 – 2104, December 2013.
- [41] C.T.E.R Hewage and M.G. Martini, "Reduced-reference quality assessment for 3D video compression and transmission", IEEE Transactions on Consumer Electronics, vol. 57, issue. 3, pp. 1185 – 1193, September 2011.
- [42] Q. Li and Z. Wang, "Reduced-Reference image quality assessment using divisive normalization-based image representation", IEEE Journal of Selected Topics in Signal Processing, vol. 3, issue. 2, pp. 202 – 211, April 2009.
- [43] W. Lin, and C.-C. Kuo, "Perceptual visual quality metrics: a survey", Journal of Visual Communication and Image Representation, vol. 22, issue. 4, pp. 297-312, May 2011.
- [44] C. Li, X. Yang, B.Chu, W. Lu and L. Pang, "A new image fusion quality assessment method based on contourlet and SSIM", 3rd IEEE International Conference on Computer Science and Information Technology, Chengdu, China, July 2010.
- [45] M. Barkowsky, K. Wang, R. Cousseau, K. Brunnstrom, R. Olsson, and P. Le Callet, "Subjective quality assessment of error concealment strategies for 3DTV in the presence of asymmetric transmission errors", 18th International Packet Video Workshop (PV), Hong Kong, China, December 2010.
- [46] G. Saygili, C.G. Gurler, and A.M. Tekalp, "Quality assessment of asymmetric stereo video coding", 17th IEEE International Conference on Image Processing (ICIP), Hong Kong, September 2010.
- [47] N. Ozbek, G. Ertan and O. Karakus, "Interactive quality assessment for asymmetric coding of 3D video", 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), Antalya, Turkey, May 2011.
- [48] A.S. Umar, R.M. Swash, and A.H. Sadka, "Subjective quality assessment of 3D videos", IEEE Africon, Livingstone, September 2011.
- [49] P. Hanhart, P. Korshunov, T. Ebrahimi, "Crowd-based quality assessment of multiview video plus depth coding", International Conference on Image Processing (ICIP), Paris, October 2014.

- [50] D. Kim, D. Min, J. Oh, S. Jeon, and K. Sohn, "Depth map quality metric for three-dimensional video", *Proceedings of SPIE*, vol. 7237, February 2009.
- [51] G. Leon, H. Kalva, and B. Furht, "3D video quality evaluation with depth quality variations", *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, Istanbul, Turkey, May 2008.
- [52] P. Joveluro, H. Malekmohamadi, W.A.C. Fernando and A.M. Kondoz, "Perceptual video quality metric for 3D video quality assessment", *The True Vision - Capture, Transmission and Display of 3D Video*, Tampere, Finland, June 2010.
- [53] H. Shao, X. Cao, and G. Er, "Objective quality of depth image based rendering in 3DTV system", *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, Potsdam, Germany, May 2009.
- [54] G. Nur, H.K. Arachchi, S. Dogan, and A.M. Kondoz, "Extended VQM model for predicting 3D video quality considering ambient illumination context", *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, Antalya, Turkey, May 2011.
- [55] L. Jin, A. Boev, A. Gotchev, K. Egiazarian, "3D-DCT based perceptual quality assessment of stereo video", *18th IEEE International Conference on Image Processing (ICIP)*, Brussels, Belgium, September 2011.
- [56] K. Ha, and M. Kim, "A perceptual quality assessment metric using temporal complexity and disparity information for stereoscopic video", *18th IEEE International Conference on Image Processing (ICIP)*, Brussels, Belgium, September 2011.
- [57] K. Yamagishi, T. Kawano, T. Hayashi, "Effect of difference in 2D video quality for left and right views on overall 3D video quality", *19th IEEE International Conference on Image Processing (ICIP)*, Orlando, FL, October 2012.
- [58] Yan Zhang, Ping An, Yanfei Wu, Zhaoyang Zhang, "A multiview video quality assessment method based on disparity and SSIM", *IEEE 10th International Conference on Signal Processing (ICSP)*, Beijing, October 2010.
- [59] K. Yamagishi, L. Karam, J. Okamoto, and T. Hayashi, "Subjective characteristics for stereoscopic high definition video", *third International Workshop on Quality of Multimedia Experience (QoMEX)*, Mechelen, September 2011.
- [60] J. Gutierrez, P. Perez, F. Jaureguizar, J. Cabrera, and N. Garcia, "Subjective assessment of the impact of transmission errors in 3DTV compared to HDTV", *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, Antalya, Turkey, May 2011.
- [61] E.J. Rodriguez, T. Nunome, and S. Tasaka, "Assessment of user behavior and QoE in multi-view video and audio IP transmission", *15th Asia-Pacific Conference on Communications (APCC)*, Shanghai, China, October 2009.
- [62] X. Liyuan, Y. Junyong, T. Ebrahimi, and A. Perkis, "Objective metrics for quality of experience in stereoscopic images", *18th IEEE International Conference on Image Processing (ICIP)*, Brussels, Belgium, September 2011.
- [63] E. Jimenez Rodriguez, T. Nunome, S. Tasaka, "Multidimensional QoE Assessment of Multi-view Video and Audio (MVV-A) IP Transmission: The Effects of User Interfaces and Contents", *26th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, Fukuoka, March 2012.

- [64] T. Nunome, S. Tasaka, "Estimation of multidimensional QoE of Multi-View Video and Audio (MVV-A) IP transmission", 18th Asia-Pacific Conference on Communications (APCC), Jeju Island, October 2012.
- [65] M. Yamamoto, T. Nunome, S. Tasaka, "The effects of camera arrangements and contents on QoE in multi-view video and audio IP transmission", IEEE Region 10 Conference TENCON, Fukuoka, November 2010.
- [66] T. Ishida, T. Nunome, "Multidimensional QoE assessment of multi-view video and selectable audio (MVV-SA) IP transmission", IEEE International Symposium on Multimedia (ISM), Anaheim, CA, December 2013.
- [67] M.G. Martini, and C.T.E.R. Hewage, "Objective quality assessment for 3-D video delivery", IEEE ComSoc MMTC E-letter, vol. 6, issue. 8, pp. 8-22, September 2011.
- [68] A. Mittal, A.K. Moorthy, J. Ghosh, and A.C. Bovik, "Algorithmic assessment of 3D quality of experience for images and videos", Digital Signal Processing and IEEE Signal Processing Education Workshop, Sedona, Arizona, January 2011.
- [69] A. Maalouf, and M.-C. Larabi, "A no-reference color video quality metric based on a 3D multispectral wavelet transform", Second International Workshop on Quality of Multimedia Experience (QoMEX), Trondheim, Norway, June 2010.
- [70] C.T.E.R. Hewage, and M.G. Martini, "Reduced-reference quality evaluation for compressed depth maps associated with colour plus depth 3D video", 17th IEEE International Conference on Image Processing (ICIP), Hong Kong, September 2010.
- [71] C.T.E.R. Hewage, and M.G. Martini, "Reduced-reference quality metric for 3D depth map transmission", 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), Tampere, Finland, June 2010.
- [72] X. Li, N. Oertel, A. Hutter and A. Kaup, "Laplace distribution based Lagrangian rate distortion optimization for hybrid video coding", IEEE Transactions on Circuits and Systems for Video Technology, vol. 19, no. 2, pp. 193 – 205, February 2009.
- [73] N.K. Sharma, M.K. Nema and S. Rakshit, "Scalable video coding using Symmetric Residue Pyramid and rate-distortion optimization", Conference on Signal Processing and Communications, India, July 2012.
- [74] R. Atta, "Optimal bit allocation for subband video coding", Image Processing, IET, vol. 4, issue 5, November 2010.
- [75] M. Koziri and A. Eleftheriadis, "Joint quantizer optimization for scalable coding", IEEE Conference on Image Processing, Hong Kong, September 2010.
- [76] M. Winken, H. Schwarz, T. Wiegand, "Joint rate-distortion optimization of transform coefficients for spatial Scalable Video Coding using SVC", 15th IEEE International Conference on Image Processing, San Diego, CA, October 2008.
- [77] M. Shoaib, A. Cai, "LA-RDO based error resilient coding using multi-layer Lagrange multiplier selection in SVC", 12th International Conference on Advanced Communication Technology (ICACT), vol. 2, pp. 1600 – 1605, February 2010.
- [78] Jing Yang, Bowen An, "Low-complexity rate-distortion optimization for robust H.264 video coding", IEEE International Conference on Multimedia and Expo (ICME), Barcelona, July 2011.

- [79] Xi Wang, Li Su, Qingming Huang, Chunxi Liu, “Visual perception based Lagrangian rate distortion optimization for video coding”, 18th IEEE International Conference on Image Processing (ICIP), Brussels, September 2011.
- [80] Shiqi Wang, A. Rehman, Zhou Wang, Ma, Siwei, Wen Gao, “SSIM-motivated rate-distortion optimization for video coding”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, issue 4, pp. 516 – 529, April 2012.
- [81] H.H. Chen, Yi-Hsin Huang, Po-Yen Su, Tao-Sheng Ou, “Improving video coding quality by perceptual rate-distortion optimization”, IEEE International Conference on Multimedia and Expo (ICME), Suntec City, July 2010.
- [82] N. Ozbek and A. M. Tekalp, “Viterbi-like joint optimization of stereo extraction for on-line rate adaptation in scalable multiview video coding”, International Conference on Image Processing, Hong Kong, September 2010.
- [83] P. Merkte, J.B. Singla, K. Muller and T. Wiegand, “Stereo video encoder optimization for mobile applications”, 3DTV Conference: True Vision - Capture, Transmission and Display of 3D Video, Turkey, May 2011.
- [84] Q. Zhang, “Joint source-channel rate-distortion optimization algorithm for multiview video coding”, IEEE International Conference on Computer and Information Technology, China, October 2012.
- [85] Y. Zhang, S. Kwong, L. Xu, S. Hu, C.-C. Jay Kuo and G. Jiang, “Regional bit allocation and rate distortion optimization for multiview depth video coding with view synthesis distortion model”, IEEE Transactions on Image Processing, vol. 22, issue 9, pp. 3497 – 3512, September 2013.
- [86] J. Chakareski, V. Velisavljevic and V. Stankovic, “User-action-driven view and rate scalable multiview video coding”, IEEE Transactions on Image Processing, vol. 22, issue. 9, pp. 3473 – 3484, June 2013.
- [87] A. Fiandrotti, J. Chakareski and P. Frossard, “Popularity-aware rate allocation in multi-view video”, Visual Communications and Image Processing, China, July 2010.
- [88] M. Cordina, C.J. Debono, “An adaptive Lagrange multiplier technique for multi-view video plus depth coding”, Picture Coding Symposium (PCS), San Jose, CA, December 2013.
- [89] Qian Ma, Yongbing Zhang, Qiong Liu, Xiangyang Ji, Qionghai Dai, “Enhanced rate-distortion optimization for stereo interleaving video coding”, International Conference on 3D Imaging (IC3D), Liege, December 2011.
- [90] Y. Peng, G. Kou, Y. Shi, and Z. Chen, “A systemic framework for the field of data mining and knowledge discovery”, IEEE International Conference on Data Mining Workshops, Hong Kong, China, December 2006.
- [91] A. Strauss, J. Corbin, “Grounded Theory Methodology: An Overview”, Sage Publications, Thousand Oaks, CA, US, 1998.
- [92] S. Looso, R. Borner, M. Goeken, “Using grounded theory for method engineering”, Fifth International Conference on Research Challenges in Information Science (RCIS), Gosier, Guadeloupe, May 2011.
- [93] S. Chakraborty, J. Dehlinger, “Applying the grounded theory method to derive enterprise system requirements”, 10th ACIS International Conference on Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing, Daegu, May 2009.

- [94] M. Raad, R. Raad, and F. Safaei, "Media rate control for large scale immersive communications", IEEE International Conference on Multimedia and Expo, Barcelona, Spain, July, 2011.
- [95] N. Ozbek, and A. M. Tekalp, "Quality layers in scalable multi-view video coding", IEEE international conference on Multimedia and Expo, New York, NY, June 28 - July 2, 2009.
- [96] M. Shafique, B. Zatt, S. Bampi, J. Henkel, "Power-aware complexity-scalable multiview video coding for mobile devices", Picture Coding Symposium (PCS), Nagoya, Japan, December 2010.
- [97] J. Lim, K. Ngan, W. Yang and K. Sohn, "A multiview sequence CODEC with view scalability", Signal Processing: Image Communication, vol. 19, issue 3, pp. 239–256, March 2004.
- [98] C.L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, R. Szeliski, "High-quality video view interpolation using a layered representation", ACM Transactions on Graphics, CA, 2004. <http://research.microsoft.com/en-us/um/people/sbkang/3dvideodownload>, last access on September, 2014.
- [99] <ftp://ftp.merl.com/pub/avetro/mvc-testseq/orig-yuv/>, last access on September, 2014.
- [100] http://www.tanimoto.nuee.nagoya-u.ac.jp/mpeg/mpeg_ftv.html, last access on September, 2014.
- [101] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," ITU-T VCEG, VCEG-M33, Austin, USA, April 2001.
- [102] M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima and Y. Mori, "Reference softwares for depth estimation and view synthesis", ISO/IEC JTC1/SC29/WG11 MPEG2008/M15377, Archamps, 2008.
- [103] Hwangjun Song, C.-C.J. Kuo, "A region-based H.263+ codec and its rate control for low VBR video", IEEE Transactions on Multimedia, vol. 6, issue 3, pp. 489 – 500, June 2004.
- [104] Z. Iravani, H. Roodaki, M.R. Hashemi, "An Efficient Parameter Selection Scheme For View Level Rate-Distortion Control In Multi-View/3d Video Coding", Seventh International Symposium on Telecommunication, Tehran, Iran, September 2014.
- [105] S.P. Pandit, Y. Chen, S. Ye, Text of ISO/IEC 14496- 5:2001/PDAM 15 Reference Software for Multiview Video Coding, ISO/IEC JTC1/SC29/WG11 MPEG2008/W9974, Germany, 2008.
- [106] Jing Hu, H. Wildfeuer, "Use of content complexity factors in video over IP quality monitoring", International workshop on Quality of Multimedia Experience (QoMEx), San Diego, CA, July 2009.
- [107] B.W. Micallef, C.J. Debono, R.A. Farrugia, "Exploiting depth information for fast multi-view video coding", Picture Coding Symposium, Japan, December 2010.
- [108] M. Rezaei, S. Wenger, M. Gabbouj, "Analyzed rate distortion model in standard video codecs for rate control", IEEE Workshop on Signal Processing Systems, Greece, November 2005.
- [109] J. Korhonen, J. You, "Improving objective video quality assessment with content analysis", International Workshop on Video Processing and Quality Metrics for Consumer Electronics, USA, January 2010.

- [110] Tanimoto Laboratory, <http://www.tanimoto.nuee.nagoya-u.ac.jp/~fukushima/mpegftv/yuv/>, last access on September, 2014.
- [111] Yanwei Liu, Qingming Huang, Siwei Ma, Debin Zhao, Wen Gao, Song Ci, and Hui Tang, "A novel rate control technique for multiview video plus depth based 3D video coding", *IEEE Transactions on Broadcasting*, vol 57, issue 2, pp. 562 – 571, June 2011.
- [112] MVC Extension for Inclusion of Depth Maps Draft Text 6, document JCT3V-C1001.doc, JCT-3V, Mar. 2013.
- [113] P. Merkle, A. Smolic, K. Muller, T. Wiegand, "Multi-View Video Plus Depth Representation and Coding", *IEEE International Conference on Image Processing*, San Antonio, TX, 2007.
- [114] L. Yixia, C. Yilin, and Z. Na, "Method of ensuring eye-to eye contact in multi-view video conference", *IEEE International Conference on Broadband Network & Multimedia Technology*, Beijing, China, October 2009.
- [115] Ka-Man Wong, Lai-Man Po, Kwok-Wai Cheung, Ka-Ho Ng, and Xuyuan Xu. "Stretching, compression and shearing disparity compensated prediction techniques for stereo and multiview video coding", *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011.
- [116] Merl, <ftp://ftp.merl.com/pub/avetro/mvc-testseq/stereo-interlaced/420/>, last access on September, 2014.
- [117] ITU-R Recommendation BT.500-10, Methodology for the subjective assessment of the quality of television picture. <http://www.itu.int/rec/R-REC-BT.500/en>, 2000.
- [118] H. Koumaras, A. Kourtis, Cheng-Han Lin, and Ce-Kuen Shieh, "A theoretical framework for end-to-end video quality prediction of MPEG-based sequences", *3rd International Conference on Networking and Services*, Athens, Greece, June 2007.
- [119] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", *IEEE Transactions on Image Processing*, vol. 13, issue 4, pp. 600 – 612, April 2004.
- [120] C.T.E.R Hewage, S.T. Worrall, S. Dogan, and A.M. Kondo, "Prediction of stereoscopic video quality using objective quality models of 2-D video", *Electronics Letters*, vol. 44, issue 16, pp. 963 – 965, August 2008.
- [121] R. Feghali, F. Speranza, D. Wang, and A. Vincent, "Video quality metric for bit rate control via joint adjustment of quantization and frame rate", *IEEE Transactions on Broadcasting*, vol. 53, issue 1, pp. 441 – 446, March 2007.
- [122] H. Jia, W. Gao, and Y. Lu, "Stereoscopic video coding based on global displacement compensated prediction", *Information, Communications and Signal Processing and the Pacific Rim Conference on Multimedia*, Singapore, December 2003.
- [123] H. Roodaki, H.R. Rabiee and M. Ghanbari, "Rate-distortion optimization of scalable video codecs", *Signal Processing: Image Communication*, vol. 25, issue 4, pp. 276–286, April 2010.
- [124] Qing Wu, Shing-Chow Chan, Heung-Yeung Shum, "A Convex Optimization-Based Object-Level Rate Control Algorithm for MPEG-4 Video Object Coding", *IEEE Asia Pacific Conference on Circuits and Systems*, Singapore, December 2006.

-
- [125] Sudeng Hu, Sam Kwong, Yun Zhang and C.-C. Jay Kuo, "Rate-distortion optimized rate control for depth map-based 3-D video coding", IEEE Transactions on Image Processing, vol. 22, issue 2, pp. 585 – 594, February 2013.
- [126] M. S. Bazaraa, H. D. Sherali and C. M. Shetty, "Nonlinear programming: theory and application", 2nd ed., New York: Wiley, 1993.
- [127] Y. Chang and M. Kim, "A joint rate control scheme in a hybrid stereoscopic video codec system for 3DTV broadcasting", IEEE Transactions on Broadcast, vol. 59, no. 2, pp. 265 – 280, June 2013.

Abstract

With the growing demand for 3D video, efforts are underway to incorporate it in the next generation of broadcast and streaming applications and standards. Scalability is one possible solution to reduce the amount of data in multi-view/3D video in heterogeneous environments. But using Scalable Multi-view Video Coding (SMVC) for multi-view/3D video still has many unresolved challenges.

In this thesis, we propose an adaptive framework to use SMVC in various 3D video applications effectively. For this issue, first, the proper scalable modality should be selected according to the application at hand, its related features and requirements. To the best of our knowledge, no work has systematically defined new and proper scalable modalities specifically for multi-view 3D video, so far. Hence, at the first step of the proposed framework, we will suggest a methodology to extract the proper scalable modalities for multi-view/3D video. In addition, while SMVC can help support heterogeneous receivers, the question becomes: how to scale the 3D video content in a given type of scalability and a specified application in order to achieve the highest performance and satisfy the receivers' constraint as much as possible? In other words, the proper mechanism to assign SMVC data to various layers should be clearly determined. This issue is considered as the second step of our proposed framework. This method uses the inter-layer and intra-layer disparity concepts. Note that specific features of any given scalable modality should be used to define these concepts in that specific scalable modality. Simulation results indicate that the proposed method achieves relatively better compression rate for each layer, with much less overhead. At the next step of our proposed framework, we propose an analytical view-level rate model for multi-view video coding. Our rate model takes into account both previous theoretical results as well as new results specifically obtained for multi-view video and confirmed by comprehensive practical experiments. Simulation results show that our model can predict the rate of each view with relatively high precision and a low estimation error of 12% on average for tested sequences. However, the evaluation of the overall visual quality of scalable multi-view video requires a new objective perceptual quality measure specifically designed for scalable multi-view/3D video. Although several subjective and objective quality assessment methods have been proposed for multi-view/3D sequences, no comparable attempt has been made for quality assessment of scalable multi-view/3D video so far. Hence, in this framework, we propose a new methodology to build suitable objective quality assessment metrics for different scalable modalities in multi-view/3D video. Our proposed methodology considers the importance of each layer and its content as a quality of experience factor in the overall quality. Furthermore, in addition to the quality of each layer, the concept of inter-layer and intra-layer disparity is considered as an effective feature to evaluate overall perceived quality more accurately. Our simulation results indicate that the correlation coefficient between our extracted objective quality evaluation metric and subjective quality assessment is 0.8 on average for tested video sequences. At the last step of our proposed framework, we present a novel method for rate-distortion optimization in scalable multi-view video that tries to minimize the perceptual distortion of decoded video under the conditions that the sum of bits generated from different views is constrained within a given bit budget. Since the constraint-based optimization problem is usually computational intensive, our proposed approach considers the concept of intra-layer and inter-layer disparity to reduce this computational complexity. Experimental results show that the proposed approach uses on average 24% and 42% less bitrate than the H.264/AVC rate-distortion optimization for base and base plus enhancement layers, respectively.



University of Tehran

Faculty of Engineering

School of Electrical and Computer Engineering

An Adaptive Framework for Scalable Multi-view Video Coding in H.264/AVC Standard

By
Hoda Roodaki Lavasani

Supervisor
Dr. Mahmoud Reza Hashemi

Advisor
Dr. Shervin Shirmohammadi

September 2014