

دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده مهندسی برق و کامپیوتر

طراحی یک کنترل کننده‌ی نرخ بیت آگاه از محتوا برای بازی‌های مبتنی بر رایانش ابری

نگارش:
حامد احمدی

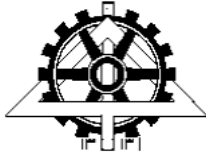
استاد راهنما:
دکتر محمودرضا هاشمی

استاد مشاور:
دکتر شروین شیرمحمدی

رساله برای دریافت درجه دکتری در رشته
مهندسی کامپیوتر - گرایش فناوری اطلاعات

اسفند ۱۳۹۴

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده مهندسی برق و کامپیوتر

طراحی یک کنترل کننده نرخ بیت آگاه از محتوا برای بازی‌های مبتنی بر رایانش ابری

نگارش:
حامد احمدی

استاد راهنما:
دکتر محمودرضا هاشمی

استاد مشاور:
دکتر شروین شیرمحمدی

رساله برای دریافت درجه دکتری در رشته
مهندسی کامپیوتر - گرایش فناوری اطلاعات

اسفند ۱۳۹۴



دانشگاه تهران



پردیس دانشکده های فنی

دانشکده مهندسی برق و کامپیوتر

پایان نامه برای دریافت درجه دکتری در رشته مهندسی کامپیوتر – فناوری اطلاعات

عنوان: طراحی یک کنترل کننده ی نرخ بیت آگاه از محتوا برای بازی های مبتنی بر رایانش ابری

نگارش: حامد احمدی

این پایان نامه در تاریخ ۱۳۹۴/۱۲/۱۲ در مقابل هیأت داوران دفاع گردید و مورد تصویب قرار گرفت.

معاون آموزشی و تحصیلات تکمیلی پردیس دانشکده های فنی: دکتر جلیل آقا راشد محصل

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 محمدرضا پاکروان

عضو هیأت داوران: دکتر شهره کسایی

عضو هیأت داوران: دکتر امید فاطمی

عضو هیأت داوران: دکتر هادی مرادی

تعهد نامه اصالت اثر

اینجانب حامد احمدی 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 می باشد.

نام و نام خانوادگی دانشجو : حامد احمدی

امضای دانشجو :

تقدیم به

عزیزانم که مراقبت از آنها در طول این سال ها به تعویق افتاد.

تشکر و قدردانی:

بر خود لازم می‌دانم از همراهی جناب آقای دکتر محمودرضا هاشمی و جناب آقای دکتر شروین شیرمحمدی نه تنها در کلیه مراحل شکل‌گیری این رساله، بلکه در طول پستی‌ها و بلندی‌های دوره دکتری صمیمانه تشکر نمایم. همچنین از کمک همیشگی دوست گرامی جناب آقای سامان توت‌آغاج بسیار سپاسگزارم.

چکیده

فراگیر شدن کاربرد بازی‌های مبتنی بر رایانش ابری^۱ (CG) و رسیدن به هدف نهایی آن که امکان بازی در هر زمان، در هر مکان، و بر روی هر وسیله است، بدون کاهش پهنای‌بند موردنیاز آن (حدود ۵ مگابیت بر ثانیه) عملی نخواهد بود. کدکننده‌های ویدئوی معمول، اگرچه سعی می‌کنند هر دو نوع افزودنی آماری^۲ و ادراکی^۳ را حذف نمایند، اما به دلیل پیچیدگی بالای سامانه بینایی چشم انسان^۴ (HVS) و همچنین تنوع محتوا در کاربردهای مختلف، شناخت و حذف افزودنی‌های ادراکی همچنان به تحقیق بیشتری نیاز دارد. ارائه مدل‌های ادراکی، به خصوص مدل‌های توجه بصری^۵، برای کاربرد بازی‌های مبتنی بر رایانش ابری نیز به دلیل محتوای خاص ویدئوهای بازی، مانند خیالی بودن شخصیت‌ها، نوع خاص چیدمان اشیا، و پیچیدگی طراحی و منطق بازی به منظور افزایش غوطه‌وری و رضایت بازیکن‌ها، چالش‌برانگیز است. تفاوت سطح مهارت، عادت‌ها، و راهبردهای بازیکن‌ها که در الگوی توجه آنها به ویدئوی بازی تاثیرگذار است، از دیگر چالش‌های توسعه چنین مدل‌هایی محسوب می‌شود. بنابراین اولین گام برای به‌کارگیری کدکردن ادراکی در کاهش پهنای‌بند و کنترل نرخ بیت کاربرد بازی‌های مبتنی بر رایانش ابری، توسعه مدل‌هایی است که بتوانند بر چالش‌های مذکور فائق آیند. این کار خود نیز مستلزم داشتن مجموعه‌داده‌ای است که نقطه توجه کاربر و سایر ویژگی‌های نامبرده شده را در حین بازی بازیکن‌ها ضبط کرده باشد. از آنجا که تاکنون مجموعه‌داده و مدل توجه مناسبی برای بازی‌های مبتنی بر رایانش ابری ارائه نشده است، بخشی از این رساله به این موارد می‌پردازد. در این راستا، دو مجموعه‌داده و دو مدل ادراکی مبتنی بر توجه ارائه گردید. مجموعه‌داده اول شامل ویدئوی بازی‌های گوناگون و اشیا موجود در آنها می‌باشد. مدلی که از روی این مجموعه‌داده ساخته می‌شود، مبتنی بر سازوکار توجه در انسان است و محل توجه بازیکن‌ها را بر اساس ترکیب ویژگی‌های سطح پایین سیگنال و اولویت‌دهی اشیا موجود در بازی در شرایط گوناگون، پیش‌بینی می‌کند. نتایج آزمایش‌های کیفی نشان می‌دهد که با تنظیم پارامتر پیمانه‌بندی^۶ کدکننده‌ی H.264/AVC بر اساس این مدل، می‌توان بطور متوسط تا ۲۵٪ نرخ بیت ویدئوی بازی را کاهش داد، بدون آنکه کیفیت ادراکی بازیکن تغییر زیادی داشته باشد. مدل ادراکی دوم، تفاوت الگوهای توجه بین بازیکن‌ها با مهارت‌های مختلف را مدل می‌کند. برای توسعه این مدل، ابتدا الگوهای جمع‌آوری شده از دستگاه ردیاب چشم^۷ خوشه‌بندی شده و سپس همبستگی بالای خوشه‌ها با مهارت بازیکن‌ها از طرق آماری و آزمایشگاهی نشان داده می‌شود. برای پیش‌بینی سطح مهارت بازیکن‌ها نیز از امتیاز استفاده می‌شود. استفاده از این مدل بسته به سطح مهارت بازیکن‌ها به طور متوسط تا ۱۵٪ نرخ بیت را کاهش می‌دهد. گام بعدی، به‌کارگیری مدل‌های ادراکی در کدکننده ویدئو است. برای این کار، از مدل‌های نرخ-اعوجاج ادراکی استفاده می‌شود تا اختصاص بیت‌ها به نواحی مختلف ویدئو بر اساس اهمیت ادراکی آنها برای HVS کنترل گردد. از آنجا که مدل‌های اختصاص بیت مبتنی بر توجه فعلی سایر ویژگی‌های ادراکی HVS را در نظر نمی‌گیرند، در بعضی مواقع مقدار اعوجاج تحمیل شده به نواحی کم‌اهمیت ویدئو از حدی فراتر می‌رود که حواس بازیکن را پرت کرده و باعث افت کیفیت ادراکی وی می‌شود. در همین راستا، مدل جدیدی ارائه می‌شود که با در نظر گرفتن هر دو سازوکار توجه و گودی مرکزی شبکه^۸ مقدار اعوجاج در نواحی مختلف را بر اساس فاصله آن‌ها از نواحی مهم ویدئو کنترل کرده و متوسط امتیاز نظر^۹ را حدود ۲۰٪ افزایش دهد. از دیگر دستاوردهای این رساله، توسعه یک سامانه آزمون بازی‌های مبتنی بر رایانش ابری برای گسترش پژوهش‌های مرتبط با CG می‌باشد.

واژه‌های کلیدی: بازی‌های مبتنی بر رایانش ابری، کدکردن ویدئو، کدکردن ادراکی ویدئو، مدل‌های توجه بصری

^۱ Cloud Gaming (CG)

^۲ Encoder

^۳ Statistical

^۴ Perceptual

^۵ Human Visual System (HVS)

^۶ Visual Attention Model

^۷ Quantization

^۸ Eye Tracker Device

^۹ Fovea

^{۱۰} Mean Opinion Score (MOS)

فهرست مطالب

۱- مقدمه.....	۱
۱-۱- بیان مسأله.....	۳
۱-۱-۱- توسعه مدل توجه بازی.....	۱۲
۱-۱-۲- توسعه مدل توجه مبتنی بر مهارت.....	۱۴
۱-۱-۳- استخراج مدل نرخ-اعوجاج مبتنی بر توجه و گودی مرکزی شبکه.....	۱۵
۲-۱- ساختار رساله.....	۱۷
۳-۱- جمع‌بندی.....	۱۷
۲- پیشینه تحقیق.....	۱۹
۱-۲- مدل‌های ادراکی.....	۲۰
۱-۱-۲- انتخاب دستی نواحی مورد علاقه.....	۲۰
۲-۱-۲- انتخاب نواحی مورد علاقه بر اساس ورودی کاربر.....	۲۰
۳-۱-۲- حساسیت بصری.....	۲۱
۴-۱-۲- گودی مرکزی شبکه.....	۲۲
۵-۱-۲- توجه بصری.....	۲۳
۲-۲- کد کردن ادراکی.....	۳۱
۱-۲-۲- پیش‌پردازش ویدئو.....	۳۲
۲-۲-۲- جفت کدکننده/واکدکننده.....	۳۲
۳-۲-۲- تنظیم پارامترهای کدکننده.....	۳۳
۳-۲- معیارهای کیفیت ادراکی ویدئو.....	۳۴
۱-۳-۲- معیارهای کیفی.....	۳۴
۲-۳-۲- معیارهای کمی.....	۳۴
۴-۲- جمع‌بندی.....	۳۵
۳- مدل‌های ادراکی پیشنهادی.....	۳۷
۱-۳- مدل توجه بازی.....	۴۰
۱-۱-۳- پیاده‌سازی.....	۴۲
۲-۱-۳- ارزیابی مدل توجه بازی.....	۴۵
۳-۱-۳- جمع‌بندی.....	۵۵
۲-۳- مدل توجه مبتنی بر مهارت.....	۵۶
۱-۲-۳- توسعه مدل.....	۵۶

۶۸	۲-۲-۳- پیش بینی الگوی توجه بازیکن
۷۰	۳-۲-۳- اعمال مدل در کدکننده ویدئو
۷۵	۴-۲-۳- ارزیابی دقت مدل توجه
۷۸	۵-۲-۳- بحث
۸۰	۳-۳- جمع‌بندی
۸۱	۴- مدل نرخ-اعوجاج پیشنهادی
۸۳	۱-۴- ساخت مدل
۸۶	۱-۱-۴- تهیه مجموعه داده
۸۶	۲-۱-۴- بدست آوردن نقشه توجه
۸۶	۳-۱-۴- اختصاص بیت بر اساس توجه
۸۷	۴-۱-۴- بدست آوردن کیفیت ادراکی بهتر
۸۸	۵-۱-۴- استخراج تابع f
۹۹	۲-۴- جمع‌بندی
۱۰۰	۵- نتیجه‌گیری و کارهای آینده
۱۰۲	۱-۵- کارهای آینده
۱۰۴	ضمیمه الف- سامانه آزمون بازی‌های مبتنی بر رایانش ابری
۱۰۶	الف-۱- سامانه GAMINGANYWHERE
۱۰۶	الف-۱-۱- طراحی
۱۰۷	الف-۱-۲- معماری
۱۰۸	الف-۱-۳- پیاده‌سازی
۱۱۰	الف-۲- سامانه‌ی ارائه شده
۱۱۰	الف-۲-۱- ساختار سامانه
۱۱۳	الف-۲-۲- راه‌اندازی
۱۱۴	الف-۳- جمع‌بندی
۱۱۶	مراجع
۱۲۳	واژه نامه انگلیسی به فارسی

فهرست اشکال

- شکل ۱-۱- بلوک دیاگرام یک سامانه‌ی CG ۲
- شکل ۲-۱- اجزای فشرده‌سازی ادراکی ویدئو ۴
- شکل ۳-۱- تاثیر تشخیص صحیح یا غلط اهمیت نواحی از نظر فشرده‌سازی و کیفیت تجربه ۸
- شکل ۴-۱- جایگاه تمرکز این رساله در کاربرد CG ۹
- شکل ۵-۱- زنجیره ارزش کاربرد CG ۱۰
- شکل ۶-۱- مدل‌های پیشنهادی در این تحقیق ۱۲
- شکل ۷-۱- محورهای موثر در تشخیص ROI برای کاربرد CG ۱۳
- شکل ۸-۱- روند انجام کارها در رساله به همراه دستاوردها ۱۷
- شکل ۱-۲- قدرت بینایی چشم چپ نسبت به زاویه از محور دید ۲۳
- شکل ۲-۲- اعمال ویژگی‌های پیشنهادی ایتی ۲۵
- شکل ۳-۲- اعمال مدل ایتی بر روی مجموعه داده‌ی Toronto-120 ۲۸
- شکل ۱-۳- اعمال مدل‌های مختلف بر روی صحنه‌هایی از بازی‌ها و نمایش عدم تطبیق با خیرگی چشم ۳۹
- شکل ۲-۳- مدل توجه بازی ارائه شده ۴۱
- شکل ۳-۳- یک نمونه پیاده‌سازی مدل توجه بازی ۴۲
- شکل ۴-۳- نتایج پیمانه‌های مدل توجه بازی برای سه نمونه فریم ۴۴
- شکل ۵-۳- اولین فریم هر یک از نه ویدئوی بازی‌های استفاده شده در ارزیابی ۴۸
- شکل ۶-۳- مقدار PSNR ویدئو COD به ازای مقادیر مختلف QP ۵۰
- شکل ۷-۳- نمودار نرخ-اعوجاج ویدئوی COD به ازای سه تایی‌های مختلف QP ۵۰
- شکل ۸-۳- درصد متوسط نواحی کم اهمیت، اهمیت متوسط و با اهمیت در نگاشتهای مختلف ۵۴
- شکل ۹-۳- اولین فریم کد شده ویدئوی BLR در سناریوی (A) تک-QP و (B) چند-QP ۵۵
- شکل ۱۰-۳- مراحل توسعه مدل توجه مبتنی بر مهارت ۵۷
- شکل ۱۱-۳- صحنه‌هایی از بازی «عروسک قشنگ من، سُمی» ۶۰
- شکل ۱۲-۳- گام‌های آزمایش ردیاب چشم ۶۱
- شکل ۱۳-۳- بافت‌نگاشت امتیازهای شرکت‌کنندگان ۶۳
- شکل ۱۴-۳- دسته‌بندی اشیا بازی سُمی ۶۳
- شکل ۱۵-۳- متوسط توجه در هر دسته برای مهارت‌های مختلف ۶۴
- شکل ۱۶-۳- متوسط توجه در هر دسته برای مهارت‌ها و حالت‌های مختلف ۶۵
- شکل ۱۷-۳- شمای بلاک‌های متناظر در نقشه‌های توجه ۶۷
- شکل ۱۸-۳- نقشه توجه هر سطح از مهارت به همراه نقشه یک بازیکن نمونه از همان سطح ۶۷
- شکل ۱۹-۳- جزئیات کنترل‌کننده نرخ در JM ۷۱
- شکل ۲۰-۳- لایه‌های کنترل‌کننده نرخ در JM و محل پیمانه تطبیق QP ۷۲

شکل ۳-۲۱ - مقدار EWPSNR برای ۱۲۰ فریم متوالی از ویدئوی یک نمونه بازیکن متوسط	۷۴
شکل ۳-۲۲ - مقایسه NSS با و بدون در نظر گرفتن مهارت	۷۷
شکل ۳-۲۳ - مقدار NSS برای مدل‌های توجه بصری و مدل پیشنهادی	۷۷
شکل ۴-۱ - جداسازی سیگنال توسط تبدیل	۸۴
شکل ۴-۲ - روال یافتن تابع f	۸۶
شکل ۴-۳ - مقادیر تابع f حاصل از ویدئوهای گام چهارم	۸۹
شکل ۴-۴ - رابطه مقادیر تابع f و وزن‌های توجه در ویدئوها	۹۰
شکل ۴-۵ - منحنی برازش شده بر روی تابع f و وزن‌های توجه	۹۱
شکل ۴-۶ - رابطه مقادیر تابع f و وزن‌های توجه در ویدئوها	۹۳
شکل ۴-۷ - رابطه مقادیر تابع f و وزن‌های گودی مرکزی شبکه در ویدئوها	۹۴
شکل ۴-۸ - دو فریم متناظر از ویدئو PES - فریم بالا از گام سوم، فریم پایین از گام چهارم	۹۶
شکل ۴-۹ - دو فریم متناظر از ویدئو SNE2 - فریم بالا از گام سوم، فریم پایین از گام چهارم	۹۷
شکل الف-۱ - سناریوی استفاده از سیستم GA	۱۰۷
شکل الف-۲ - نگاه پیمانه‌ای به کارساز و کارگزار GA	۱۰۸
شکل الف-۳ - ارتباط بین پیمانه‌های کارساز، بافرهای مشترک و ارتباطات شبکه در GA	۱۰۹
شکل الف-۴ - ارتباط بین پیمانه‌های کارخواه، بافرهای مشترک و ارتباطات شبکه در GA	۱۰۹
شکل الف-۵ - گراف فیلتر در سمت کارساز شامل چهار فیلتر	۱۱۰
شکل الف-۶ - رابطه بین API گرافیکی، کاربرد و کارت گرافیک	۱۱۱
شکل الف-۷ - نمای اصلی برنامه کارساز	۱۱۴
شکل الف-۸ - سامانه آزمون در حال اجرا	۱۱۴

فهرست جداول

جدول ۱-۲	شش مدل برتر در محک جاد	۲۹
جدول ۱-۳	اطلاعات بازبهای انتخاب شده برای ارزیابی مدل توجه بازی	۴۶
جدول ۲-۳	نام و مقدار پارامترهای کدکننده H.264/AVC استفاده شده در ارزیابی	۴۹
جدول ۳-۳	مقادیر PSNR (dB)، SSIM و نرخ بیت (Mbps) برای QP های مختلف	۵۱
جدول ۴-۳	درصد کاهش نرخ بیت برای سناریوهای چند-QP	۵۱
جدول ۵-۳	اطلاعات دموگرافیک شرکتکنندگان	۵۲
جدول ۶-۳	اطلاعات دموگرافیک شرکتکنندگان	۵۳
جدول ۷-۳	نتایج آزمایش برای دو سناریو	۵۳
جدول ۸-۳	ویژگی دیتاست تهیه شده و دیگر دیتاست‌های مرتبط با کاربرد بازی	۵۹
جدول ۹-۳	اطلاعات دموگرافیک شرکت کنندگان در آزمایش	۶۱
جدول ۱۰-۳	تقسیم‌بندی شرکت کنندگان به سه گروه	۶۲
جدول ۱۱-۳	نتایج تست ANOVA	۶۸
جدول ۱۲-۳	نتایج تست t-Test برای مبتدی-متوسط	۶۹
جدول ۱۳-۳	نتایج تست t-Test برای متوسط-حرفه‌ای	۶۹
جدول ۱۴-۳	نتایج تست t-Test برای مبتدی-حرفه‌ای	۶۹
جدول ۱۵-۳	دقت مرزهای انتخاب شده	۷۰
جدول ۱۶-۳	ماتریس درهم‌ریختگی پیش‌بینی خوشه‌ها	۷۰
جدول ۱۷-۳	پارامترهای کدکننده استفاده شده در شبیه‌سازی‌ها	۷۲
جدول ۱۸-۳	مقدار EWPSNR و نرخ بیت برای مقادیر مختلف QP	۷۳
جدول ۱۹-۳	درصد تغییر EWPSNR و نرخ بیت برای مقادیر مختلف QP	۷۴
جدول ۲۰-۳	مقدار تغییر EWPSNR و نرخ بیت با و بدون استفاده از پیمانۀ تطبیق QP	۷۵
جدول ۲۱-۳	دقت مدل‌های بصری و مدل پیشنهادی	۷۵
جدول ۱-۴	بیت‌های اختصاص داده شده به هر ناحیه بر اساس مدل پیشین	۸۷
جدول ۲-۴	اطلاعات دموگرافیک شرکتکنندگان	۸۸
جدول ۳-۴	نتایج آزمایش کیفی برای دو سناریو	۸۸
جدول ۴-۴	مقادیر تابع f برای هر ناحیه	۸۹
جدول ۵-۴	بیت‌های اختصاص داده شده به هر ناحیه در گام چهارم	۸۹
جدول ۶-۴	وزن توجه نواحی مختلف	۹۰
جدول ۷-۴	مقادیر $R2$ برای انواع منحنی‌های برازش شده بر تابع f و وزن‌های توجه	۹۱
جدول ۸-۴	مقادیر φ برای هر ناحیه	۹۲
جدول ۹-۴	مقادیر $R2$ برای انواع منحنی‌های برازش شده بر تابع f و وزن‌های گودی مرکزی شبکه	۹۳

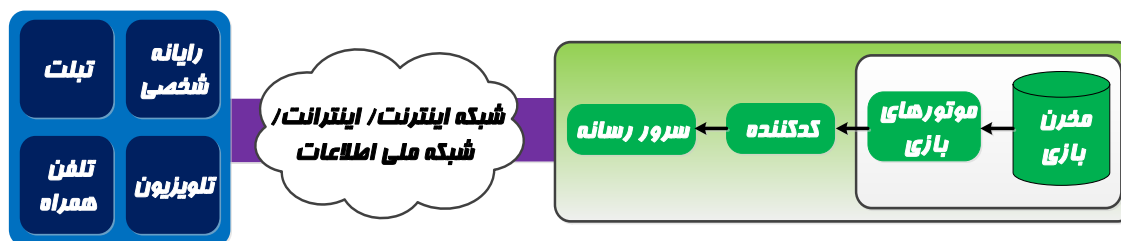
جدول ۴-۱۰- مقادیر واقعی و تخمین زده شده تابع f	۹۴
جدول ۴-۱۱- نرخ بیت‌های اختصاص داده شده به نواحی مختلف با مدل پیشنهادی و مدل [56].....	۹۸
جدول ۴-۱۲- نتایج آزمایش کیفی.....	۹۹

فصل اول

مقدمه

بازی‌های مبتنی بر رایانش ابری یک دامنه جذاب و نوظهور در صنعت سرگرمی محسوب می‌شود. با استفاده از این رویکرد جدید، بازیکن‌ها دیگر نیازی به خرید سخت‌افزارهای قوی و پرهزینه ندارند. آن‌ها همچنین نگران ناسازگاری‌های احتمالی در نصب بازی نخواهند بود. با استفاده از این رویکرد هر فرد می‌تواند در هر زمان و بر روی هر دستگاهی به انجام بازی مورد علاقه خود که در محیط رایانش ابری قرار دارد بپردازد. افزایش دسترسی به اینترنت پرسرعت و پیشرفت تکنولوژی‌های سخت‌افزاری - بخصوص در زمینه پردازنده‌های گرافیکی (GPU) و کدکننده‌های نرم‌افزاری یا سخت‌افزاری ویدئو - شرایط را برای تحقق این رویکرد فراهم ساخته است. منافع بالقوه این سرویس برای بازیکن‌ها، توسعه‌دهندگان بازی‌ها و ارائه‌دهندگان خدمات ارزش افزوده، شرکت‌های بزرگی چون NVidia و AMD، Google و Samsung را نیز به فعالیت جدی در این حوزه واداشته است.

بازی‌های مبتنی بر رایانش ابری، که از این پس در این گزارش با CG به آن اشاره می‌شود، ترکیبی از مزایای بازی‌های اینترنتی و رایانش ابری می‌باشد. مفهوم CG در واقع به تصویر کشیدن بازی در ابر و ارسال آن در قالب ویدئو بر روی شبکه می‌باشد. بر خلاف سرویس جویبارسازی ویدئو^۱ سرویس CG یک سرویس دوطرفه است که در آن ورودی‌های کاربر نیز به کارساز^۲ ارسال می‌شوند. نمودار یک سامانه CG در شکل ۱-۱ آمده است.



شکل ۱-۱- بلوک دیاگرام یک سامانه‌ی CG

سرویس CG از جنبه‌های مختلف برای کاربران، توسعه‌دهندگان بازی و ارائه‌دهندگان سرویس مزیت دارد. همانطور که ذکر شد با استفاده از این رویکرد نیاز کاربران به خرید سخت‌افزارهای قوی و گران قیمت از بین می‌رود. آن‌ها همچنین در مورد سازگاری بازی با سیستم عامل و سایر برنامه‌های رایانه خود نگرانی نخواهند داشت. بدین ترتیب با استفاده از این سرویس آن‌ها می‌توانند در هر کجا و بر روی هر سکویی از طریق شبکه به انجام بازی مورد علاقه خود بپردازند. علاوه بر این بر حسب مدل کسب و کار، این امکان برای کاربران وجود دارد تا با پرداخت یک هزینه اشتراک به مجموعه وسیعی از بازی‌ها دسترسی داشته باشند. نیازمندی‌های این سرویس برای کاربران تنها یک دستگاه با قابلیت کدگشایی برخط ویدئو و یک خط شبکه پرسرعت می‌باشد.

^۱ Video Streaming

^۲ Server

یکی از مشکلات اساسی در توزیع بازی‌ها بر روی لوح‌های فشرده، مساله برداری غیرمجاز می‌باشد. از آنجا که در CG بازی‌ها بر روی ابر اجرا می‌شوند، عملاً امکان نسخه‌برداری از آن‌ها وجود نخواهد داشت. این یکی از مزیت‌های اصلی CG برای توسعه‌دهندگان بازی‌های رایانه‌ای محسوب می‌شود. در رویکرد CG، بازی تنها بر روی یک سکو اجرا می‌شود؛ اما در عین حال ویدئوی بازی پس از تولید بر روی انواع سکوها قابل جویبارسازی است. بنابراین توسعه‌دهندگان بازی‌ها با استفاده از CG می‌توانند انرژی‌ای که تا قبل از این صرف سازگارسازی بازی خود با انواع سیستم‌های عامل و سکوها می‌کردند، را بر روی بهبود خود بازی (از جمله بخش‌های گرافیک و هوش مصنوعی آن) متمرکز نمایند. به این ترتیب، رویکرد CG می‌تواند هزینه ورود به بازار را برای تازه‌واردین صنعت کاهش دهد. این تازه‌واردین، با استفاده از CG، دیگر نیازی به تهیه سرمایه اولیه برای ساخت سخت‌افزار کنسول و توزیع آن نخواهند داشت.

سود صنعت بازی بسیار زیاد است، بطوری‌که سود آن دو برابر صنعت ویدئو برآورد شده است [۱]. ارائه دهندگان سرویس‌های CG می‌توانند با تکیه بر مزایای جذاب CG برای کاربران و استفاده از یک مدل کسب و کار مناسب سهم خوبی از این بازار را بدست آورند. جلوگیری از نسخه‌برداری غیرمجاز، ارتباط مستقیم با کاربر و حذف توزیع‌کنندگان لوح‌های فشرده و همچنین مقیاس‌پذیری رایانش ابری همگی از مواردی هستند که ریسک سرمایه‌گذاری در این حوزه را کاهش می‌دهند. بنابراین شرکت‌های اپراتور مخابرات و تلویزیون می‌توانند با ارائه این سرویس به مشتریان فعلی خود، درآمد خود را افزایش داده و از منابع موجود خود استفاده موثرتری نمایند.

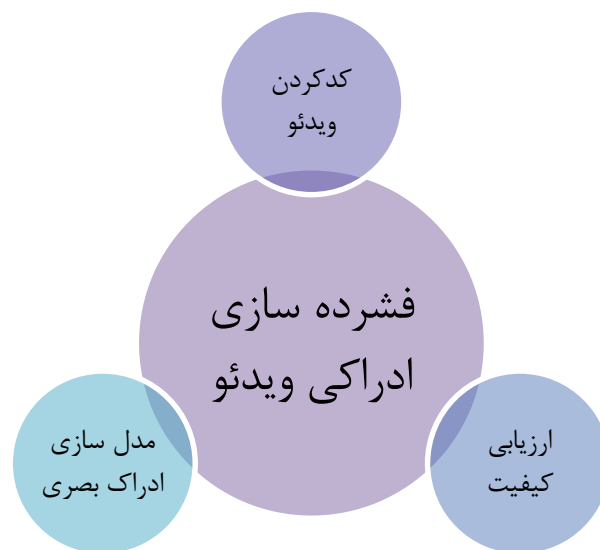
۱-۱- بیان مسأله

در سال‌های اخیر با ظهور شبکه‌های جدید و دسترسی کاربران به خطوط پرسرعت، ایده انتقال سرویس از دستگاه مشتری به کارساز تحقق یافته است. سرویس CG نیز در ادامه همین روند قرار دارد. این رویکرد اگرچه مزایای زیادی دربردارد، اما چالش‌هایی هم دارد که باید مورد توجه قرار گیرد. یکی از مهمترین چالش‌ها، پهنای باند است. در این سرویس به ازای هر کاربر یک ویدئو آن هم عموماً با کیفیت HD بر روی شبکه ارسال می‌شود. بنابراین در صورت عدم استفاده از روش‌های فشرده‌سازی مناسب شبکه‌ای که این سرویس بر روی آن ارائه می‌شود باید پهنای باند بالایی (حدود ۵ مگابیت بر ثانیه برای برخی نمونه‌های تجاری موجود) داشته باشد.

در چند سال اخیر، تحقیقات بر روی بازی‌های مبتنی بر رایانش ابری سرعت یافته است. این تحقیقات شامل ارائه معماری مناسب برای راه‌اندازی و ارائه این سرویس [۲، ۳]، مدل مناسب کسب و کار [۴]، [۵] و ارزیابی کیفیت تجربه [۶-۸] و اندازه‌گیری تاخیر [۹-۱۱] می‌باشند. با این وجود، تحقیق عمیقی در زمینه کاهش پهنای‌باند موردنیاز این سرویس انجام نشده است.

هدف از این پایان‌نامه ارائه راه‌حلی برای برخورد با چالش پهنای‌باند بالای مورد نیاز سرویس CG می‌باشد. کاهش پهنای‌باند موردنیاز نه تنها تعداد مشتریان بالقوه CG را افزایش می‌دهد، بلکه باعث کاهش هزینه‌های مشتریان فعلی نیز می‌شود. علاوه بر این نمایشگرهای با وضوح فوق-HD به بازار راه پیدا کرده‌اند و بزودی تقاضا برای جویبارسازی بازی‌ها با چنین وضوحی افزایش خواهد یافت. در واقع، اگرچه به مرور با پیشرفت و در دسترس قرار گرفتن شبکه‌های پرسرعت، پهنای‌باند بیشتری در اختیار مشتریان قرار می‌گیرد، این افزایش پهنای‌باند همراه با ارائه کاربردهای جدیدتر و تغییر الگوی مصرف مشتریان خواهد بود و نمی‌توان برای آن حساب باز کرد. بنابراین تحقیق در زمینه پهنای‌باند، امکان فراگیر شدن و موفقیت این سرویس را افزایش می‌دهد.

در این راستا، از روش‌های کدکردن ادراکی ویدئو^۱ استفاده خواهد شد. این روش‌ها سعی دارند با در نظر گرفتن ویژگی‌های سیستم بینایی چشم انسان^۲ (HVS)، علاوه بر افزونگی‌های آماری^۳، افزونگی‌های ادراکی^۴ ویدئو را نیز حذف نمایند و به این ترتیب نرخ فشرده‌سازی را افزایش دهند. به عنوان مثال، بعضی از این روش‌ها سعی می‌کنند، بر پایه سازوکار توجه^۵، نقاطی که بیننده توجه بیشتری به آنها می‌نماید را از سایر نقاط تمیز دهند و در مرحله بعد کیفیت کدکردن ویدئو در این نقاط را با توجه به سطح اهمیت آنها کنترل کنند بطوریکه نرخ بیت ویدئو کاهش یافته و در عین حال کیفیت آن حفظ شود. فشرده‌سازی ادراکی ویدئو سه جزء اساسی دارد: مدل ادراکی، اعمال مدل در کدکننده و ارزیابی کیفیت (شکل ۱-۲).



شکل ۱-۲- اجزای فشرده‌سازی ادراکی ویدئو

^۱ Perceptual Video Coding

^۲ Human Visual System (HVS)

^۳ Statistical Redundancy

^۴ Perceptual Redundancy

^۵ Attention Mechanism

مدل‌های ادراکی بر اساس ویژگی‌های مختلف چشم از جمله حساسیت بصری، توزیع سلول‌ها در شبکه و سازوکار توجه، ناحیه مورد علاقه‌ی^۱ (ROI) کاربر را پیش‌بینی می‌کنند. تشخیص ناحیه موردعلاقه از یک سمت، بر اساس ویژگی‌های سطح پایین سلول‌های چشم در درک سیگنال انجام می‌شود و از طرف دیگر، به فرامین سطح بالایی که از طرف مغز برای هدایت چشم ارسال می‌شود، وابسته است. مورد دوم بسیار به کاربرد وابسته است. به عنوان مثال، مدل‌های ادراکی که برای ویدئو کنفرانس توسعه می‌یابند، عموماً به دنبال چهره انسان می‌گردند. این مدل‌ها بطور مستقیم برای کاربرد رانندگی که در آن علائم راهنمایی و رانندگی و سایر ماشین‌ها اهمیت دارند، قابل استفاده نیستند. همچنین اهدافی که بیننده در ذهن دارد، بر روی توجه وی تاثیرگذار است. به عنوان مثال، اگر تصاویری از اشخاص مختلف را به گروهی بدهیم و از آنها بخواهیم ثروت اشخاص موجود در تصویر را حدس بزنند، آنها بیشتر به لباس‌ها توجه می‌کنند. این در حالی است که اگر از همین گروه بخواهیم، در مورد سن افراد حدس بزنند، آنها بیشتر به صورت‌ها توجه می‌کنند [۱۲]. تاکنون تقریباً تمام مدل‌های ادراکی در حوزه‌هایی غیر از دنیای بازی ارائه شده‌اند [۱۳]. این مدل‌ها به دلیل محتوای خاص ویدئوهای بازی، مانند فانتزی بودن شخصیت‌ها، نوع خاص چیدمان اشیاء و پیچیدگی طراحی و منطق بازی در CG کارآمد نیستند. به عنوان مثال، الگوریتم‌هایی که در مدل‌های ادراکی مخصوص ویدئو کنفرانس به دنبال صورت انسان هستند، در بازی‌هایی که دارای شخصیت‌های فانتزی هستند، عملکرد مطلوبی ندارند. در همین مثال، پس‌زمینه از اهمیت پایینی برخوردار است، در حالیکه در بازی‌های رایانه‌ای در بعضی حالت‌ها و بسته به منطق بازی و تنظیمات انتخابی برای افزایش غوطه‌وری و رقابت‌پذیری بخشی از اشیاء مهم بازی در پس‌زمینه مستتر می‌شود. به عنوان مثالی دیگر، در کاربرد تماشای آزاد، نگاه بیننده به مرکز تصویر تمایل دارد. بسیاری از مدل‌های توجه بصری که زیر مجموعه‌ای از مدل‌های ادراکی محسوب می‌شوند و سازوکار توجه در انسان را مدل‌سازی می‌کنند، برای مرکز تصویر وزن بیشتری قائل هستند، در حالی که این فرض در آن دسته از بازی‌ها که دوربین به صورت خودکار عمل نمی‌کند و بسته به انتخاب و ورودی‌های کاربر مکان، زاویه دید و مقدار بزرگنمایی آن تغییر می‌کند، صادق نیست. تفاوت دیگر کاربرد بازی نسبت به سایر کاربردها تغییر اولویت اشیاء در حین بازی بسته به عواملی نظیر فعالیت کنونی بازیکن و حالت بازی است. به عنوان مثال، توجه بازیکن به نواحی مختلف تصویر در حالیکه مشغول جنگ با یک دشمن است، با زمانی که در حال جستجوی یک شی خاص است، متفاوت است. حالت فعلی بازی نیز اهمیت خود را دارد. به عنوان مثال، در بازی مسابقه رانندگی وقتی مقدار بنزین کاهش می‌یابد، بازیکن توجه بیشتری به علائم راهنمایی و رانندگی موجود در صحنه می‌نماید تا در اسرع وقت یک پمپ بنزین پیدا کرده و سوخت‌گیری نماید. علاوه بر ویژگی‌های بازی، ویژگی‌های بازیکن‌ها نیز مدل‌سازی برای کاربرد بازی‌های مبتنی بر رایانش ابری را چالش‌برانگیز می‌کند. در واقع، تفاوت سطح مهارت، عادت‌ها و راهبردهای بازیکن‌ها در الگوی توجه آنها به ویدئوی بازی تاثیرگذار است. به عنوان مثال، بازیکن‌های

^۱ Region Of Interest (ROI)

مبتدی هنگام پرش از روی موانع برای اینکه مطمئن شوند به آنها برخورد نمی‌کنند، هنگام پرش توجه زیادی به شخصیت خود در بازی هنگام پرش و همچنین خود مانع می‌کنند، در حالیکه بازیکن‌های حرفه‌ای این معمولاً زمان رسیدن به مانع را پیش‌بینی کرده و در حالیکه توجه خود به موارد دیگری از معطوف کرده‌اند، وقتی زمان پرش فرا رسد، دکمه پرش را فشار می‌دهند.

به این ترتیب، مدل‌های درکی ارائه شده برای کاربردهایی غیر از CG، برای این کاربرد کارآیی مطلوبی نخواهند داشت. در واقع این کاربرد نیاز به طراحی مدل‌هایی را دارد که بتوانند بر چالش‌های مذکور فائق آیند. در این رساله این مسئله مورد بررسی قرار گرفته و دو مدل ادراکی مخصوص بازی ارائه می‌شود. هر دوی این مدل‌ها، در واقع مدل‌های توجه بصری هستند. همانطور که قبلاً گفته شد، مدل‌های توجه زیرمجموعه‌ای از مدل‌های ادراکی هستند که بر اساس سازوکار توجه در انسان، نواحی موردعلاقه بازیکن در ویدئوی بازی را پیش‌بینی می‌کنند. از میان ویژگی‌های ادراکی سیستم بینایی انسان، از آن جهت سازوکار توجه در این رساله مورد توجه قرار گرفت که سایر ویژگی‌ها وابستگی کمتری به کاربرد دارند. به طور دقیق‌تر، توجه در انسان به دو شکل تحریک‌محور و وظیفه‌محور است. بخش تحریک‌محور بر اساس ویژگی‌های سطح پایین حسگرهای چشم عمل می‌کند و تقریباً عملکرد آن در کاربردهای مختلف یکسان است. این در حالی است که بخش وظیفه‌محور بر اساس اهدافی که انسان در سر دارد، توجه وی را به بخش خاصی از صحنه معطوف می‌کند. به عنوان مثال، بازیکنی را در نظر بگیرید که شخصیت وی در بازی مجروح شده و به دنبال جعبه کمک‌های اولیه می‌گردد. هر تغییر ناگهانی روشنایی در تصویر، مستقل از هدف بازیکن، توجه پایین-به-بالای وی را جلب می‌کند. اما از آنجا که وی به دنبال جعبه کمک‌های اولیه است، سازوکار توجه بالا-به-پایین وی سعی می‌کند به دنبال نواحی سفید و یا علامت قرمز بگردد.

مدل‌های توجه بصری، سازوکار توجه در انسان را بصورت محاسباتی شبیه‌سازی می‌کنند. آن دسته از این مدل‌ها که برای کاربردهای خاصی مانند کاربرد ویدئو کنفرانس توسعه یافته‌اند، همانطور که قبلاً هم بیان شد، برای کاربرد بازی مناسب نیستند. دسته دیگری از این مدل‌ها نیز وجود دارد که با در نظر گرفتن ویژگی‌های مستقل از کاربرد مانند فعالیت بافت و حرکت در تصویر نواحی موردعلاقه بازیکن را پیش‌بینی می‌کند. این مدل‌ها نیز به دلیل پیچیدگی‌های نامبرده شده برای کاربرد بازی کارآمد نیستند. به عنوان مثال، معمولاً در پس‌زمینه بازی‌ها اشیاء متحرکی وجود دارند که لزوماً در تمام لحظات بازی در موفقیت بازیکن تاثیرگذار نیستند. بنابراین در عمل، بازیکن تنها در بعضی مواقع به آنها توجه می‌کند. این در حالی است که آن اشیاء به علت تحرکشان در همه مواقع توسط این دسته از مدل‌های توجه بصری به عنوان شیء مهم پیش‌بینی می‌شوند.

جزء بعدی فشرده‌سازی ادراکی، اعمال مدل‌ها در کدکننده است. در این زمینه نیز روش‌های گوناگونی ارائه شده است. اساس این روش‌ها، کنترل کیفیت ویدئو بر اساس مدل ادراکی است. به عنوان مثال، برخی از این روش‌ها پارامتر پیمان‌بندی کدکننده ویدئو را در بخش‌های پراهمیت، کوچک‌تر و در سایر بخش‌ها بزرگ‌تر در نظر می‌گیرد. به این ترتیب، مقدار اعوجاج در نواحی کم‌اهمیت بیشتر

می‌شود. اما از آنجا که بیننده به آن نواحی کمتر توجه می‌کند، متوجه اعوجاج موجود در آنها نمی‌شود. این باعث می‌شود تا بتوان بدون کاهش کیفیت تجربه^۱، حجم ویدئو را کاهش داد. کیفیت تجربه در واقع یک معیار کیفی از تجربه مشتری از یک کاربرد مشخص می‌باشد. کیفیت تجربه برای هر کاربرد متفاوت است. در بازی‌های مبتنی بر رایانش ابری، کیفیت را می‌توان هم از نظر بازی و هم از نظر ویدئو بررسی کرد. از نظر بازی، کیفیت تجربه با استفاده از معیارهایی مانند غوطه‌وری^۲، کشش^۳ و رقابت‌پذیری^۴ سنجیده می‌شود [۱۴]. از نظر ویدئو نیز معیارهایی مانند PSNR و SSIM وجود دارد. نکته‌ای که در اینجا باید بدان اشاره کرد رابطه کیفیت ویدئو بر معیارهای مبتنی بر بازی، مانند غوطه‌وری، می‌باشد. به عنوان مثال، کاهش کیفیت ویدئوی بازی ممکن است حواس بازیکن را پرت کند و حس غوطه‌وری را از وی بگیرد. اعوجاج در ویدئو حتی ممکن است بازیکن را عصبی نماید. علاوه بر این، اعوجاج در ویدئو ممکن است موفقیت بازیکن در بازی را نیز تحت الشعاع قرار دهد. به عنوان مثال اگر کیفیت ویدئو مناسب نباشد، ممکن است بازیکن یکی از دشمنان یا یکی از علامت‌های مهم در بازی را نبیند و امتیاز از دست دهد. بنابراین کیفیت ویدئوی بازی یکی از عوامل مهم کیفیت تجربه کاربر در سرویس بازی‌های مبتنی بر رایانش ابری می‌باشد.

کیفیت تجربه ممکن است برای بازی‌های مختلف متفاوت باشد. به عنوان مثال، در بازی‌های سریع، مثل بازی‌های شلیک‌گر اول شخص^۵، که موفقیت بازیکن در گرو عکس‌العمل سریع وی است، بازیکن کمتر به محیط پس‌زمینه دقت می‌کند. در این نوع بازی‌ها، کاهش کیفیت در اشیایی که بازیکن مستقیماً با آنها درگیر می‌شود، تاثیر منفی بیشتری بر روی کیفیت تجربه کاربر می‌گذارد. مستقل از نوع بازی، کیفیت تجربه به انتظار و هدف کاربر از بازی نیز وابسته است. به عنوان مثال، بازیکن هنگام شرکت در یک مسابقه، تمام تلاش خود را برای بدست آوردن یک رتبه بهتر انجام می‌دهد و در این حالت توجه کمتری به کیفیت ویدئوی بازی خواهد داشت.

در حال حاضر گروه‌های طراح بازی همکاری موثری با گروه‌های فشرده‌سازی ویدئو ندارند. این در حالی است که نتایج آزمایشات این رساله نشان می‌دهد با برقراری چنین ارتباطی می‌توان نرخ فشرده‌سازی در سرویس بازی‌های مبتنی بر رایانش ابری را با دقت و سرعت خیلی بیشتری کاهش داد. بعد از تحقیق بر روی محتوای بازی‌ها و پیدا کردن ویژگی‌های موثر برگرفته از محتوا در فشرده‌سازی ویدئو، طراح بازی می‌تواند با ارسال اطلاعات جانبی از طریق واسط‌های نرم‌افزاری مشخص در استخراج دقیق و سریع این ویژگی‌ها کمک نماید. اما تا شکل‌گیری این ارتباط، باید سعی کنیم با استفاده از ویژگی‌هایی مانند تغییرات رنگ، نور، بافت و حرکت در فضاهای زمانی، مکانی و

^۱ Quality of Experience (QoE)

^۲ Immersion

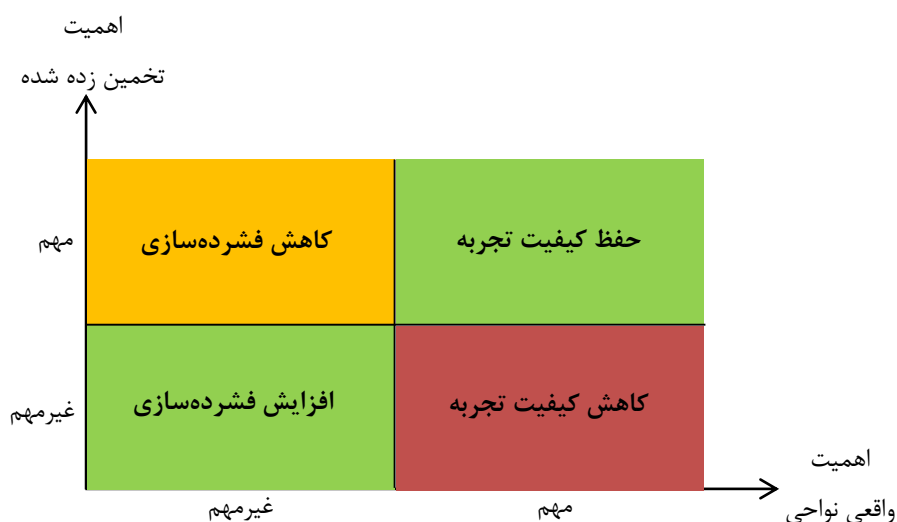
^۳ Tension

^۴ Competence

^۵ First Person Shooter (FPS)

فرکانسی نواحی مورد علاقه بازیکن را تخمین بزنییم و از آن در افزایش نرخ فشرده‌سازی ویدئوی بازی بهره ببریم.

در صورتیکه دقت تشخیص نواحی مورد علاقه پایین باشد، ممکن است نواحی مهم به عنوان نواحی غیرمهم تشخیص داده شده و کیفیت آنها به اشتباه کم شود. این افت کیفیت در نواحی مهم، علاوه بر آنکه لذت بازی را از بازیکن می‌گیرد حتی ممکن است موفقیت او در بازی را تحت شعاع قرار دهد. حالت بد دیگر، تشخیص نواحی غیرمهم به عنوان نواحی مهم است که مانع افزایش نرخ فشرده‌سازی می‌شود. بنابراین تشخیص صحیح ناحیه موردعلاقه تاثیر مستقیم بر روی کارایی کنترل‌کننده نرخ بیت خواهد داشت. شکل ۳-۱ تاثیر تشخیص صحیح یا غلط اهمیت نواحی از نظر فشرده‌سازی و کیفیت تجربه را نشان می‌دهد.

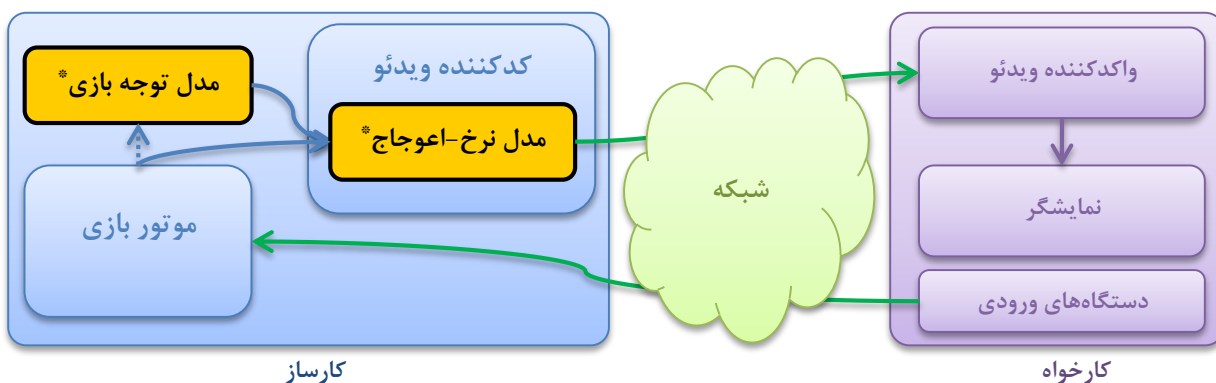


شکل ۳-۱ - تاثیر تشخیص صحیح یا غلط اهمیت نواحی از نظر فشرده‌سازی و کیفیت تجربه

بعد از بدست آمدن اطلاعات در مورد نواحی موردعلاقه، مسئله بعدی نحوه استفاده از این اطلاعات در کنترل‌کننده نرخ بیت می‌باشد. این کنترل‌کننده باید بر اساس نرخ بیت هدف، پارامترهای کدکننده‌ی ویدئو را بگونه‌ای تنظیم نماید که میزان اعوجاج در ویدئوی گذشته به کمترین مقدار ممکن برسد. در اینجا باید توجه کرد که درک اعوجاج در نواحی مختلف ویدئو از دید کاربر متفاوت است. در واقع، بازیکن در هر لحظه از بازی، بر حسب فعالیت فعلی خود، بر روی قسمتی از تصویر تمرکز کرده و به سایر نواحی تصویر توجه کمتری می‌کند. اعوجاج در نواحی مورد توجه بازیکن به مراتب بیشتر از اعوجاج در سایر نواحی، چشم و فکر بازیکن را آزار می‌دهد. بنابراین استفاده از کنترل‌کننده‌های نرخ بیت فعلی که مدل‌های نرخ-اعوجاج‌شان بر اساس این نوع از اعوجاج نمی‌باشد، برای کاربرد CG مناسب نیست. کنترل‌کننده ویدئو برای کاربرد CG باید در تنظیم پارامترهای کدکننده این مسئله را مدنظر قرار دهد. یکی از بخش‌های مهم کنترل‌کننده ویدئو، الگوریتم تخصیص بیت در آن می‌باشد.

تخصیص بیت می‌تواند در سه لایه‌ی گروه تصاویر^۱، فریم و یا واحدهای کدشونده کوچکتر مانند ماکروبلوک انجام گیرد. در هر لایه کل بیت‌های موجود بین واحدهای آن لایه تقسیم می‌شود. الگوریتم‌های معمول بدون در نظر گرفتن محتوا و عموماً بطور یکنواخت بیت‌های باقیمانده را بین واحدها تقسیم می‌کنند. بنابراین صرف‌نظر از میزان اهمیت واحدها، تعداد یکسانی بیت به آنها اختصاص داده می‌شود. این در حالی است که الگوریتم‌های فشرده‌سازی ادراکی سعی می‌کنند با اختصاص بیت کمتر به نواحی کم اهمیت و استفاده از بیت‌های بیشتر برای نواحی پراهمیت، کیفیت نهایی ویدئو را بهبود بخشند. چالش این الگوریتم‌ها استفاده از مدل نرخ-اعوجاجی است که مقدار اعوجاج قابل تحمیل در نواحی مختلف ویدئو را به درستی تعیین کند. از آنجا که مدل‌های اختصاص بیت مبتنی بر توجه موجود، سایر ویژگی‌های HVS را در نظر نمی‌گیرند، در بعضی مواقع مقدار اعوجاج تحمیل شده به نواحی کم‌اهمیت ویدئو از حدی فراتر می‌رود که حواس بازیکن را پرت کرده و باعث افت کیفیت ادراکی وی می‌شود. در همین راستا، در این رساله، یک مدل نرخ-اعوجاج جدید ارائه می‌شود که با در نظر گرفتن سازوکار گودی مرکزی شبکه در کنار سازوکار توجه مقدار اعوجاج در نواحی مختلف را بر اساس فاصله آنها از نواحی مهم ویدئو کنترل کرده و کیفیت ادراکی بهتری تولید می‌کند.

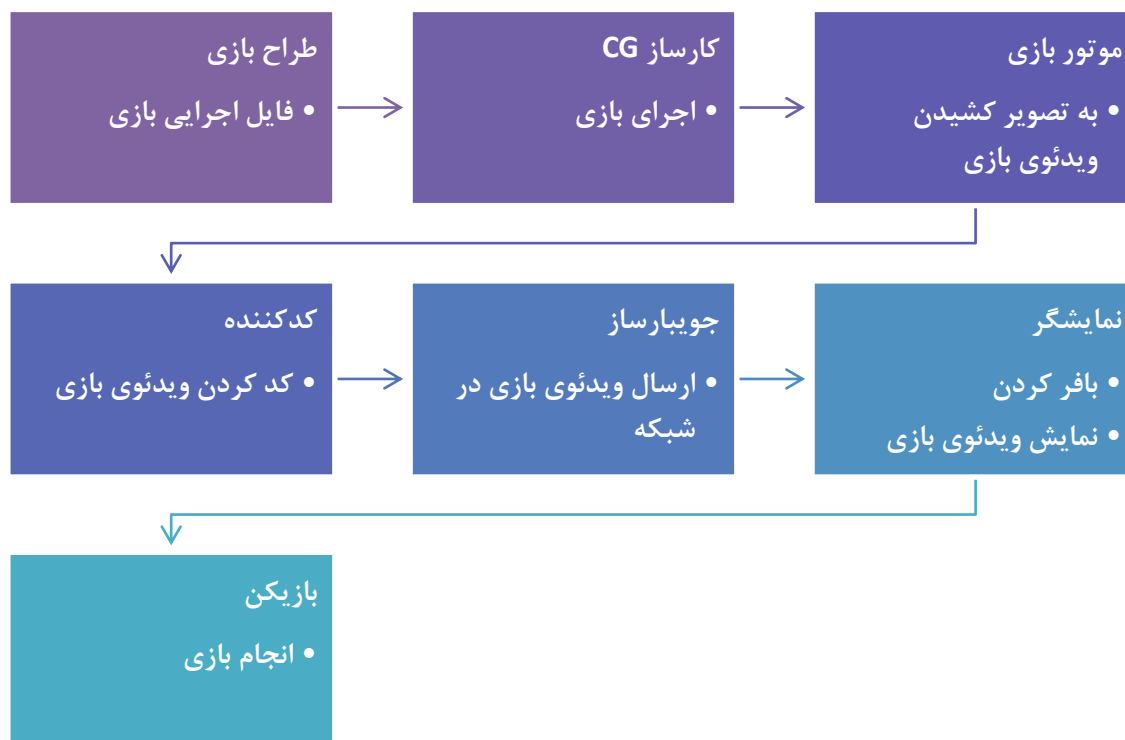
شکل ۴-۱ یک معماری عمومی از کاربرد CG را نشان می‌دهد. در این شکل، جعبه‌های مدل نرخ-اعوجاج و مدل توجه بازی که با رنگ نارنجی (ستاره‌دار) مشخص شده‌اند، محل تمرکز این رساله می‌باشند.



شکل ۴-۱ - جایگاه تمرکز این رساله در کاربرد CG

البته مدل‌های توجه بصری ارائه شده در این رساله، علاوه بر اینکه در کد کردن ویدئو کاربرد دارند، در سایر حلقه‌های زنجیره ارزش کاربرد CG نیز می‌توانند مورد استفاده قرار گیرند.

^۱ Group of pictures (GOP)



شکل ۵-۱- زنجیره ارزش کاربرد CG

شکل ۵-۱ زنجیره ارزش^۱ کاربرد CG را نشان می‌دهد. یک مدل توجه مخصوص بازی در همه حلقه‌های این زنجیره کاربرد مفید دارد. با در دست داشتن این مدل، طراحان بازی می‌توانند برای جلب توجه بازیکن برای یک امر خاص در حین بازی، از این مدل در طراحی خود استفاده کنند. به عنوان مثال، علامت شروع جنگ در یک نقطه از نقشه، در بازی‌های استراتژیک، که ممکن است یک متن، تصویر یا پویانمایی باشد، می‌تواند بر اساس این مدل ساخته شود تا احتمال جلب توجه کاربر بیشتر شود. حتی می‌توان بصورت هوشمندتری از این مدل استفاده کرد و میزان جلب توجه چنین مواردی را با توجه به مهارت بازیکن در بازی تنظیم کرد. به عنوان مثال، بازیکن‌های با مهارت پایین عادت ندارند هر از گاهی به نواحی مختلف تصویر نگاه کنند و ممکن است اطلاعات مهمی از بازی را که در موفقیت و خشنودی آنها موثر است، از دست بدهند. بنابراین می‌توان میزان جلب توجه چنین اطلاعاتی را برای چنین بازیکنانی افزایش داد. برعکس، با کاهش میزان جلب توجه چنین اطلاعاتی می‌توان مقدار رقابت‌پذیری بازی برای افراد ماهر را افزایش داد.

کارساز CG که وظیفه مجازی‌سازی پردازنده گرافیکی و اجرای موتور بازی بر روی آن را بر عهده دارد، نیز می‌تواند از مدل توجه بازی برای کنترل بار پردازشی استفاده نماید. به عنوان مثال، موتور بازی

^۱ Value Chain

می‌تواند بگونه‌ای طراحی شود که با تعیین مناسب سطح جزییات^۱ بافت بر اساس اهمیت اشیا مختلف بازی بار پردازشی را تعدیل نماید.

چنین مدلی در لایه شبکه نیز می‌تواند مفید باشد. یکی از کاربردهای آن، افزایش مقاومت در برابر خطای جویبارسازی ویدئو در شبکه می‌باشد. روش‌هایی که در این زمینه استفاده می‌شوند از جمله روش نشانه‌گذاری و یا پارتیشن کردن رشته بیت، معیارهای خاص خود را دارند. مدل توجه بازی با مشخص کردن درجه اهمیت ماکرو بلاک‌ها می‌تواند معیار مناسبی برای این قبیل الگوریتم‌ها باشد. در واقع مدل توجه بازی می‌تواند به عنوان معیاری برای برقراری تعادل بین میزان مقاوم‌سازی و میزان افزونگی، مورد استفاده قرار گیرد.

کاربرد دیگر مدل توجه مخصوص بازی می‌تواند در کد کردن مقیاس‌پذیر ویدئو^۲ باشد. بطور کلی سه نوع مقیاس‌پذیری برای کد کردن ویدئو وجود دارد: زمانی^۳، مکانی^۴ و کیفی^۵. ویدئوی مقیاس‌پذیر، یا از روش کد کردن چند-لایه ویدئو^۶ و یا از طریق کد کردن چند-توصیف^۷ بر روی شبکه ارسال می‌شود. روش اول از نظر پهنای باند مصرفی، بهتر از روش دوم می‌باشد. اما در این روش با از بین رفتن لایه‌ی پایه سایر لایه‌های بهبود، حتی در صورت سالم رسیدن به مقصد، باز هم قابل استفاده نخواهند بود. این در حالی است که در روش چند-توصیف هر یک از توصیف‌ها به تنهایی قابل واکد کردن است.

با در نظر گرفتن معایب و مزایای این دو روش و توجه به اینکه سرویس CG در حال حاضر نیز پهنای باند مصرفی بالایی دارد، به نظر می‌رسد روش چند-لایه برای این سرویس مناسب‌تر باشد. در این روش، ویدئو با حداقل کیفیت قابل قبول کد شده و در قالب لایه‌ی پایه برای مقصد ارسال می‌شود. سپس اطلاعات اضافی در قالب یک یا چند لایه‌ی بهبود برای افرادی که پهنای باند بیشتری در اختیار دارند، ارسال می‌شود. واکدکننده‌ی مقیاس‌پذیر با دریافت لایه‌های بهبود، ویدئوی ساخته شده از روی لایه‌ی پایه را بازسازی کرده و ویدئویی با کیفیت بالاتر تولید می‌نماید.

اینکه در روش کد کردن ویدئو چند-لایه، چه تعداد لایه بهبود مورد استفاده قرار گیرد و به هر یک از لایه‌ها (لایه‌ی پایه و لایه‌های بهبود) چه میزان نرخ بیت اختصاص داده شود، از چالش‌های بکارگیری این روش می‌باشد. همچنین باید معیاری برای تفکیک ویدئو در قالب لایه‌ها انتخاب گردد. این معیار باید بگونه‌ای اضافه‌شونده باشد که با رسیدن هر یک از لایه‌های بهبود، جزئیاتی به ویدئوی تاکنون بازسازی شده، اضافه گردد تا کیفیت آن بهبود یابد. مدل توجه مخصوص بازی می‌تواند در انتخاب تعداد لایه‌ها و تخصیص نرخ بیت به آنها مفید باشد.

^۱ Level of details (LOD)

^۲ Scalable video coding

^۳ Temporal

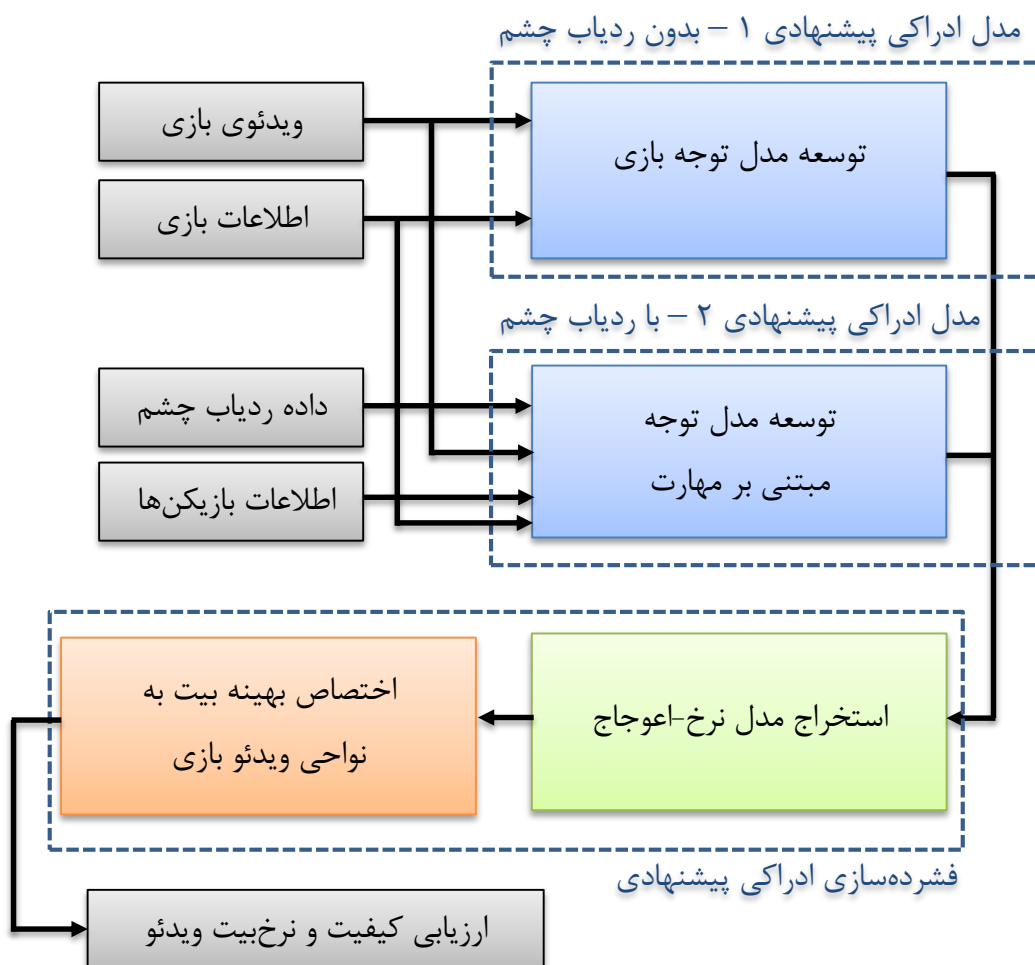
^۴ Spatial

^۵ SNR-based

^۶ Layered video coding

^۷ Multi-description video coding

مدل‌های توجه و نرخ-اعوجاج پیشنهادی و ارتباط بین آن‌ها در این تحقیق در شکل ۶-۱ نمایش داده شده است. قسمت‌های مختلف این شکل در زیربخش‌های ۱-۱-۱، ۲-۱-۱، و ۳-۱-۱ توضیح داده خواهند شد.



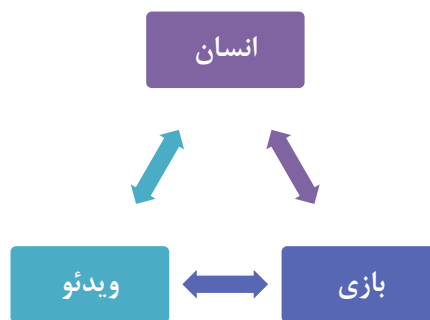
شکل ۶-۱ - مدل‌های پیشنهادی در این تحقیق و ارتباط بین آن‌ها

مدل‌های توجه و نرخ-اعوجاج پیشنهادی در این تحقیق در شکل ۶-۱ نمایش داده شده است. قسمت‌های مختلف این شکل در زیربخش‌های ۱-۱-۱، ۲-۱-۱، و ۳-۱-۱ توضیح داده خواهند شد.

۱-۱-۱ - توسعه مدل توجه بازی

در فشرده‌سازی، توزیع منبع نقش مهمی در نرخ فشرده‌سازی قابل حصول دارد. در واقع، بدون در نظر گرفتن توزیع منبع، ارائه یک روش فشرده‌سازی بهینه امکان‌پذیر نیست. در کاربرد CG، منبع، در واقع ویدئوی انواع بازی‌ها می‌باشد. از آنجا که فضای حالت بازی‌ها بسیار بزرگ و پیچیده است و کد منبع بسیاری از این بازی‌ها نیز در دسترس نیست، شناسایی توزیع منبع در کاربرد CG بسیار مشکل است. به عنوان مثال، در یک بازی مشخص با نوع معلوم، انواع مکان‌ها، اشیاء محیطی، وضعیت

آب و هوایی، جلوه‌های ویژه تصویری و غیره وجود دارد. یک عامل اصلی دیگر در فشرده‌سازی، کاربر است. حسگرهای انسان، از جمله چشم و گوش وی، ویژگی‌های خاص خود را دارند. به عنوان مثال چشم انسان به رنگ قرمز بیشتر از سایر رنگ‌ها حساس است. چشم انسان همچنین در فرکانس‌های بالا، تحمل‌پذیری بیشتری نسبت به نویز دارد. البته کارکرد چشم انسان بسیار پیچیده‌تر از این مثال‌ها می‌باشد بخصوص اینکه درک انسان فقط بر اساس اطلاعات دریافت شده از حسگرهای وی نمی‌باشد. در واقع، عوامل دیگری از جمله دانش، تجربه، مهارت و حتی وضعیت روحی انسان در درک وی تاثیرگذارند. محورهای موثر در تشخیص ROI در کاربرد CG در شکل ۷-۱ نشان داده شده است.



شکل ۷-۱ - محورهای موثر در تشخیص ROI برای کاربرد CG

همانطور که پیش‌تر نیز اشاره شد، الگوریتم‌های فعلی تشخیص ROI برای کاربرد CG مناسب نیستند. از آنجا که محتوای ویدئوها با یکدیگر فرق دارند، الگوریتم‌های ROI اغلب بصورت خاص منظوره برای یک کاربرد مشخص طراحی می‌شوند. به عنوان مثال، برای کاربرد ویدئو کنفرانس صورت افرادی که مشغول صحبت با یکدیگر هستند بسیار مهم است و معمولا افراد حرکت زیادی ندارند. بنابراین در الگوریتم‌های ROI برای این کاربرد، از تشخیص چهره استفاده می‌شود. به عنوان یک مثال دیگر، ممکن است برای یک کاربرد تحلیل دوربین ترافیک از الگوریتم‌های تشخیص ماشین یا تشخیص پلاک استفاده شود. بازی‌ها نیز ویژگی‌های خاص خود را دارند و الگوریتم‌های ROI مخصوص خود را می‌طلبند. استفاده از الگوریتم تشخیص چهره برای یک بازی فانتزی که شخصیت‌های آن از جنس حیوان یا آدم فضایی هستند، نتیجه خوبی نخواهد داشت. به علاوه فضای حالت بازی‌ها بسیار پیچیده و حجیم است. در همین مثال بازی فانتزی، صورت شخصیت‌ها همیشه مهم نیستند. در واقع، اهمیت اشیا بازی در هر لحظه به عوامل مختلفی از جمله شرایط بازی و فعالیت فعلی بازیکن بستگی دارد. علاوه بر الگوریتم‌های ROI خاص منظوره، الگوریتم‌های عمومی‌تری نیز در ادبیات پیشنهاد شده‌اند. این الگوریتم‌ها سعی می‌کنند با استفاده از ویژگی‌های سطح پایین مثل فعالیت حرکت و بافت در ویدئو نواحی مورد توجه ببینند را شناسایی کنند. اما این نوع الگوریتم‌ها نیز نمی‌توانند برای کاربرد CG عملکرد خوبی داشته باشند. در واقع بافت پیچیده یا حرکت شدید در یک ناحیه از ویدئوی بازی، نمی‌تواند لزوما نشان‌دهنده میزان اهمیت آن ناحیه از دید بازیکن باشد. به عنوان مثال، ممکن است بازیکن بخواهد از یک قسمت مین‌گذاری شده عبور کند. وی حتما با سرعت بسیار پایین از آن قسمت

عبور خواهد کرد و بیشتر توجه او معطوف به قسمت مین گذاری شده خواهد بود. این در حالی است که میزان فعالیت حرکت و بافت در سایر بخش‌ها ممکن است بیشتر باشد.

در این رساله برای تشخیص ROI از سازوکار توجه در انسان کمک گرفته می‌شود. توجه در انسان از دو بخش پایین-به-بالا و بالا-به-پایین تشکیل شده است. بخش پایین-به-بالا حجم عظیم اطلاعات دریافتی را در مسیر چشم تا مغز به روش‌های مختلف پایش می‌کند، بگونه‌ای که بیشترین اطلاعات با کمترین داده برای مغز ارسال شود. سپس مغز بر اساس وظیفه‌ی کنونی‌اش چشم را برای دریافت اطلاعات بیشتر سمتی هدایت می‌کند که توجه بالا-به-پایین نام دارد. مقالات زیادی برای مدل سازی سازوکار توجه در انسان ارائه شده است [۱۳]. بعضی مقالات فقط توجه پایین-به-بالا یا فقط توجه بالا-به-پایین را هدف قرار داده‌اند. اما آن دسته از مقالات که هر دو توجه را با هم در نظر گرفته‌اند (مدل‌های توجه ترکیبی)، موفقیت بیشتری در مدل سازی توجه انسان داشته‌اند. دلیل این امر آن است که رفتار مبتنی بر توجه در انسان توسط ترکیبی از هر دو نوع توجه هدایت می‌شود [۱۳، ۱۵].

مدل توجه بالا-به-پایین به کاربرد بسیار وابسته است. تاکنون مدل توجه بالا-به-پایینی مخصوص کاربرد بازی ارائه نشده است [۱۳]. از همین رو، یکی از اهداف این رساله ارائه یک مدل توجه ترکیبی مخصوص کاربرد بازی می‌باشد. این مدل با اولویت‌دهی اشیا بازی بر اساس فعالیت فعلی بازیکن و از جنبه‌های مختلف، توجه بالا-به-پایین را تخمین می‌زند. سپس این اطلاعات را با نتایج یک مدل توجه پایین-به-بالا ترکیب کرده و اهمیت نواحی مختلف ویدئوی بازی از دیدی بازیکن را پیش‌بینی می‌کند. نتایج آزمایش‌ها نشان می‌دهد، با تنظیم پارامتر پیمانه‌بندی در کدکننده ویدئو H.264/AVC بر اساس نقشه توجه تولید شده توسط این مدل، می‌توان با حفظ کیفیت کیفی تا ۲۵٪ نرخ بیت ویدئوی بازی را کاهش داد. دیتاست جمع آوری شده برای این ارزیابی بخش از کار شامل ویدئوهای متنوعی از نظر میزان فعالیت بافت و حرکت و همچنین نور می‌باشد و برای کارهای آتی بسیار مفید است.

۱-۱-۲- توسعه مدل توجه مبتنی بر مهارت

همانطور که در شکل ۱-۷ نشان داده شده است، یکی از محورهای موثر در تشخیص ROI برای کاربرد بازی انسان (بازیکن) است. پیش‌تر به تاثیر سیستم بینایی انسان در هدایت توجه وی به نواحی مختلف ویدئوی بازی اشاره شد. اما ویژگی‌های دیگری از انسان نیز در کاربرد بازی تاثیرگذار است. یکی از این ویژگی‌ها مهارت بازیکن است. مدل‌های معمول توجه، فرض می‌کنند نواحی پراهمیت و کم‌اهمیت در بین بینندگان مختلف تقریباً یکسان هستند [۱۶]. این فرض در کاربردهایی مانند کاربرد تماشای ویدئو که در آن کارگردان سعی می‌کند با استفاده ماهرانه از حقه‌های سینمایی توجه بینندگان به نواحی مختلف ویدئو را هدایت کند، صادق است. با این وجود، این فرض در کاربرد بازی‌های رایانه‌ای، به دلیل تعاملی بودن بازی، همواره صادق نیست. به عبارت دقیق‌تر، در آن دسته از بازی‌ها که بازیکن می‌تواند کنترل دوربین را توسط ورودی‌ها کنترل کرده و آزادانه نواحی موردنظر خود در بازی را ببیند، بسته به سطح مهارت، عادت‌ها و راهبردهای خود انتخاب‌های متفاوتی خواهد

داشت. همین مسئله باعث تفاوت الگوی توجه بین بازیکن‌ها می‌شود. حتی در بازی‌های ژانر ماشین‌های مسابقه نیز که دوربین همواره بصورت خودکار طوری تنظیم می‌شود که جاده در مرکز تصویر باشد، بازیکن‌های با مهارت بیشتر برای سبقت از دیگران و حفظ موقعیت خود توجه بیشتری به نقشه راه و آینه عقب دارند و همین مسئله باعث می‌شود، الگوهای توجه تفاوت‌هایی هرچند کم با یکدیگر داشته باشند. شناسایی این تفاوت‌ها به افزایش دقت مدل‌های توجه کمک شایانی خواهد کرد. به همین دلیل در این رساله سعی می‌شود تا با جمع‌آوری داده‌های ردیاب چشم، تاثیر مهارت بازیکن‌ها بر الگوهای توجه نشان داده شده و مدل توجه‌ای مبتنی بر مهارت توسعه داده شود.

به همین منظور، با همکاری یک گروه بازی‌ساز در دانشگاه تهران، یکی از بازی‌های این گروه به نام «عروسک قشنگ من؛ سُمی^۱» گونه‌ای تغییر داده شد تا اطلاعات اشیا بازی و رویدادهای آن را از طریق شبکه برای یک پورت خاص ارسال نماید. سپس با استفاده از یک دستگاه ردیاب چشم از ۸۰ شرکت‌کننده خواسته شد تا آن بازی را یک تا سه بار انجام دهند. در حین بازی این افراد، اطلاعات ردیاب چشم، صفحه‌کلید، موسواره و اشیا و رویدادها و ویدئوی بازی ضبط گردید. این دیتاست از دو جهت در این رساله مورد استفاده قرار گرفت. ابتدا، قسمت بالا-به-پایین مدل توجه بازی که در بخش قبل به آن اشاره شد، خودکارسازی شد. به عبارت دیگر، این قسمت یک جدول اهمیت دارد که باید توسط طراح بازی پر شود. نتایج آزمایش‌ها نشان می‌دهد با استفاده از داده‌های ردیاب چشم، این قسمت از مدل قابل خودکارسازی است. این نتایج امیدوارکننده هستند زیرا امکان استفاده از مدل در سناریوهای واقعی که در آنها لزوماً طراح بازی در دسترس نیست، عملی‌تر می‌شود.

اطلاعات ردیاب چشم از جنبه دیگری نیز مورد استفاده قرار گرفت. با تحلیل الگوهای توجه در بین بازیکن‌های مختلف با مهارت‌های مختلف نشان دادیم که بازیکن‌های با مهارت یکسان الگوهای توجه تقریباً یکسانی دارند. همچنین الگوهای توجه بازیکن‌های یک سطح از مهارت با سطح دیگری از مهارت می‌توانند به اندازه کافی متفاوت باشند. نتایج آزمایش‌ها بر روی بازی سُمی، نشان داد الگوهای بازیکن‌ها در سه مهارت مبتدی، متوسط و حرفه‌ای از هم قابل تمیز هستند. مسئله بعدی، استفاده از این الگوها در عمل است. از آنجا که در عمل، امکان استفاده از دستگاه ردیاب وجود ندارد، از روی امتیاز بازیکن‌ها سطح مهارت و متعاقباً الگوی توجه آنها پیش‌بینی گردید.

۱-۱-۳- استخراج مدل نرخ-اعوجاج مبتنی بر توجه و گودی مرکزی شبکه

بعد از توسعه مدل‌های ادراکی، باید بگونه‌ای از آنها در فرایند فشرده‌سازی استفاده کرد تا افزونگی‌های ادراکی از ویدئو حذف شود. مقالات موجود در این حوزه به سه دسته تقسیم می‌شوند. دسته اول بجای تغییر در کدکننده ویدئو، خود ویدئو را تغییر می‌دهند. به عنوان مثال، ابتدا نواحی از ویدئو که اهمیت کمتری دارند را مات کرده و سپس ویدئوی تغییریافته را برای کدکننده ارسال می‌کنند. این دسته از

^۱ My beautiful Doll, Somi

روش‌ها کمترین نرخ فشرده‌سازی را دارند. به علاوه، اعوجاج ایجاد شده در خود منبع ممکن است از دید کاربر پنهان نماند. دسته بعدی از روش‌ها، یک زوج کدکننده/واکننده طراحی می‌کنند. به عنوان مثال، ویدئو به چند قسمت سنتز شده و هر قسمت با روش خاصی کد و در سمت گیرنده واکد می‌شود. این روش‌ها به دلیل نیاز به تغییر واکدکننده‌های موجود از محبوبیت کمتری برخوردارند. دسته آخر که توجه زیادی به خود جلب کرده است، روش‌هایی هستند که سعی می‌کنند با تنظیم مقادیر پارامترهای کدکننده بر اساس مدل ادراکی، افزونگی‌های ادراکی را از بین ببرند. تنظیم مد ماروبلاک، تعیین ضریب لاگرانژ در بهینه‌سازی نرخ-اعوجاج^۱ (RDO) و تعیین تعداد بیت هدف برای واحدهای مختلف کدشونده^۲ از جمله روش‌هایی هستند که در این دسته قرار می‌گیرند. از بین این روش‌ها، آنهایی که با فرمول‌های بسته ریاضی اعوجاج را کنترل می‌کنند محبوبیت بیشتری دارند زیرا تعمیم‌پذیرتر هستند.

در قلب این الگوریتم‌ها، یک مدل نرخ-اعوجاج نهفته است که با استفاده از آن می‌توان تعادلی میان نرخ بیت و اعوجاج برقرار کرد. مدل‌های معمول نرخ-اعوجاج بر اساس میزان اختلاف ویدئوی گذشته با ویدئوی اصلی عمل کرده و ویژگی‌های سیستم بینایی چشم را در نظر نمی‌گیرند. همین مسئله باعث می‌شود به دو ناحیه با سطح اهمیت ادراکی نابرابر مقدار یکسانی اعوجاج تعلق گیرد؛ در این حالت ممکن است مقدار اعوجاج در ناحیه پراهمیت بیش‌تر از حد آستانه تحمل بوده و کاربر را متوجه خود کند. این در حالی است که با استفاده از مدل‌های نرخ-اعوجاج ادراکی می‌توان میزان اعوجاج در نواحی مختلف ویدئو را بر اساس مقدار اهمیت ادراکی آنها کنترل کرد. اما مقدار تعیین مقدار اعوجاج (نرخ بیت) برای هر ناحیه کاری چالش برانگیز است. اگر میزان اعوجاج تحمیل شده به نواحی کم‌اهمیت فراتر از حد باشد، حواس بازیکن پرت شده و کیفیت ادراکی افت خواهد کرد.

مدل‌های کنونی اختصاص بیت مبتنی بر توجه، به سایر ویژگی‌های ادراکی HVS توجه نمی‌کنند، بنابراین در بعضی مواقع مشکل مذکور برای آنها اتفاق می‌افتد. برای رفع این مشکل، تمرکز این بخش از رساله بر روی تعیین تعداد بیت‌های هدف برای نواحی مختلف ویدئو از طریق یک مدل نرخ-اعوجاج جدید می‌باشد بطوریکه که علاوه بر سازوکار توجه، سازوکار گودی مرکزی شبکه را نیز در نظر بگیرد. بدین ترتیب فاصله نواحی مختلف ویدئو از نواحی پراهمیت نیز به عنوان یک عامل کنترل‌کننده میزان اعوجاج عمل کرده و تخصیص بیت به نواحی تطبیق بیشتری با HVS خواهد داشت. مدل پیشنهادی در قالب یک مسئله بهینه‌سازی اختصاص بیت مورد استفاده قرار می‌گیرد. نتایج آزمایش‌ها نشان می‌دهد استفاده از این مدل نسبت به مدل قبلی که در آن فقط سازوکار توجه مدنظر قرار می‌گرفت، باعث بهبود کیفیت ویدئوها می‌گردد.

^۱ Rate-Distortion Optimization (RDO)

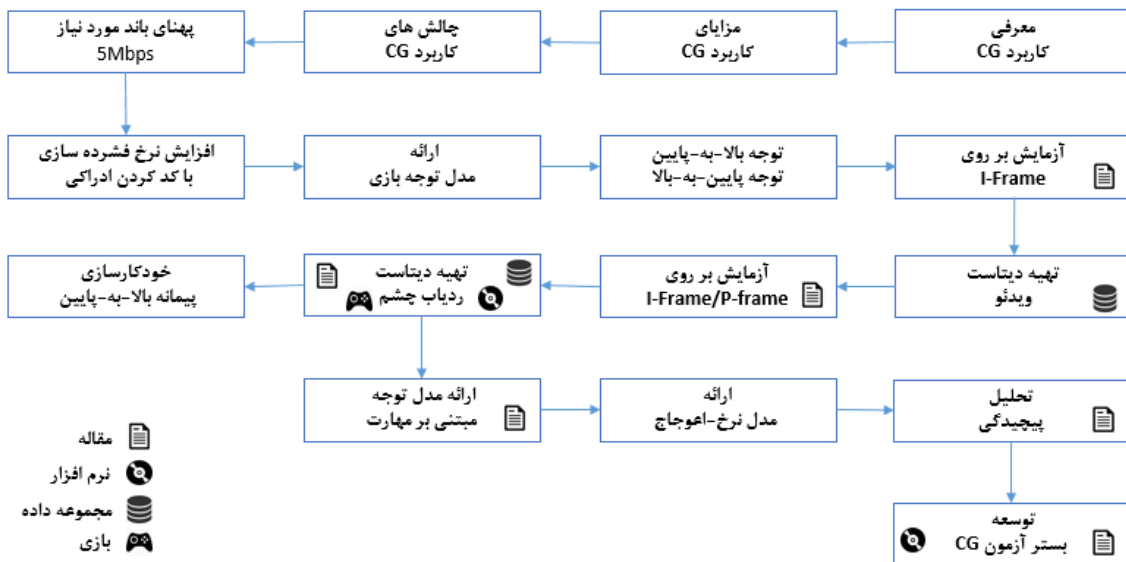
^۲ Coding Units

۲-۱- ساختار رساله

ساختار کلی رساله از چند قسمت تشکیل شده است. در فصل دوم رساله به پیشینه تحقیق روش‌های مختلف فشرده‌سازی ادراکی خواهیم پرداخت و کارهایی که در زمینه هر یک از اجزای آن 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. فصل سوم، شامل دو بخش است. در بخش اول، مدل‌های ادراکی پیشنهادی برای تشخیص نقاط مورد توجه بازیکن‌ها مورد بررسی قرار خواهد گرفت. در بخش دوم، مدل نرخ بیت-اعوجاج استخراج شده و نحوه استفاده از آن در اختصاص بیت به نواحی مختلف ویدئوی بازی ارائه خواهد شد. در نهایت، در فصل چهارم به جمع‌بندی رساله خواهیم پرداخت و کارهای آینده ممکن در رابطه با این رساله پیشنهاد خواهند شد.

۳-۱- جمع‌بندی

در این فصل فضای کلی این رساله که در ارتباط با طراحی کنترل‌کننده‌ی نرخ بیت آگاه از محتوا برای بازی‌های مبتنی بر رایانش ابری است، تبیین شد. فشرده‌سازی ادراکی معرفی و کاستی‌های روش‌های موجود برای کاربرد بازی بیان شد. توسعه دو مدل ادراکی (مدل توجه بازی و مدل توجه مبتنی بر مهارت بازیکن) و یک مدل نرخ اعوجاج مبتنی بر توجه و گودی مرکزی شبکه برای تخصیص بیت به نواحی مختلف ویدئوی بازی از بخش‌های اصلی این رساله 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. شکل ۸-۱ بطور خلاصه روند انجام کارها در رساله را نشان می‌دهد. در این شکل همچنین دستاوردهای رساله در بخش‌های مختلف آن با استفاده از نمادهای مقاله، نرم افزار، مجموعه داده و بازی نشان داده شده است.



شکل ۸-۱- روند انجام کارها در رساله به همراه دستاوردها

بر اساس این شکل، در این رساله، ابتدا کاربرد بازی‌های مبتنی بر رایانش ابری به همراه مزایا و چالش‌های آن معرفی می‌گردد. سپس بر روی چالش پهنای‌بند تمرکز و با استفاده از فشرده‌سازی ادراکی نرخ بیت مورد نیاز این کاربرد کاهش داده می‌شود. به همین منظور، یک مدل توجه بازی با الهام از سازوکارهای توجه پایین-به-بالا و بالا-به-پایین توسعه می‌شود. برای ارزیابی این مدل یک مجموعه داده از ویدئوی بازی‌های مختلف تهیه می‌شود. در ادامه برای گسترش مدل، با استفاده از دستگاه ردیاب چشم، محل توجه بازیکن‌های مختلف در حال بازی یک بازی رایانه‌ای داخلی ضبط شده و مجموعه داده دیگری تهیه می‌شود. این مجموعه داده به توسعه مدل توجه مبتنی بر مهارت کمک می‌کند. همچنین می‌توان از آن به منظور خودکارسازی پیمانه بالا-به-پایین مدل توجه بازی استفاده کرد. در ادامه یک مدل نرخ-اعوجاج توسعه داده شده و از آن برای اختصاص بیت به نواحی مختلف ویدیوی بازی بر مبنای اهمیت ادراکی‌شان استفاده می‌شود. به عنوان کارهای آتی، می‌توان عامل پیچیدگی را نیز به این مدل اضافه کرد. در این راستا تحلیل‌های اولیه از پیچیدگی کدکردن ویدئوهای بازی انجام می‌شود. در نهایت برای ارزیابی مدل‌های پیشنهادی در محیطی مشابه محیط واقعی، سامانه آزمونی مخصوص این کاربرد ارائه می‌گردد. 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 رساله شامل انواع مدل‌های ادراکی، به ویژه مدل‌های توجه بصری، انواع روش‌های فشرده‌سازی ادراکی و مدل‌های نرخ-اعوجاج، و معیارهای ارزیابی کیفیت ادراکی ویدئو مورد بررسی قرار خواهند گرفت.

۲-۱- مدل‌های ادراکی

هدف از این رساله پیدا کردن نواحی مورد علاقه‌ی بازیکن‌ها در ویدئوی بازی و متعاقباً استفاده از آن برای کاهش نرخ بیت ویدئوی ارسالی می‌باشد. در مرور ادبیات مدل‌هایی که با درک و پیش‌بینی سامانه بینایی، یا استفاده‌ی مستقیم از آن، به شناخت نواحی مورد علاقه کاربر می‌پردازند، بررسی شده‌اند. این مدل‌ها در حوزه‌ی چندرسانه‌ای و بینایی ماشینی با نام مدل‌های ادراکی شناخته می‌شوند. در این بخش، این مدل‌ها در قالب پنج گروه مورد بررسی قرار می‌گیرند. در گروه اول به صورت دستی نواحی مهم و مورد علاقه کاربر تعیین می‌شود. گروه دوم با استفاده از ورودی داده‌ی کاربر، از جمله ورودی نقاط خیرگی^۱ که از دستگاه ردیاب چشم به صورت مستقیم و برخط بدست آمده است، نواحی مورد علاقه را تعیین می‌کند. این دسته روش‌ها در کاربردهای بلادرنگ بسیار مورد استفاده است. گروه سوم از نابرابری حساسیت بصری انسان^۲ در نواحی مختلف تصویر استفاده کرده تا اعوجاج ناشی از کد کردن را در نواحی‌ای که برای انسان محسوس نیست، پنهان کند. گروه چهارم از توزیع غیریکنواخت سلول‌های بینایی در شبکه‌ی چشم انسان کمک می‌گیرد. گروه پنجم استفاده از روش‌های توجه بصری است که با استفاده از الگوریتم‌های بینایی ماشینی سعی در پیش‌بینی نواحی مورد علاقه و نواحی برجستگی^۳ تصویر دارند.

۲-۱-۱- انتخاب دستی نواحی مورد علاقه

در برخی کاربردها تعیین دقیق نواحی مورد علاقه (ROI) در ویدئو بسیار مهم و حیاتی است. بنابراین در این دسته کاربردها سعی می‌شود به صورت دستی نواحی مورد علاقه مشخص شوند. از جمله‌ی این کاربردها می‌توان به کاربردهای پزشکی اشاره کرد که کاهش کیفیت در نواحی توجه، قابل تحمل نیست. در این کاربردها، نواحی مورد علاقه توسط یک متخصص تعیین می‌گردد.

۲-۱-۲- انتخاب نواحی مورد علاقه بر اساس ورودی کاربر

یک راه دقیق برای یافتن نواحی مهم در فریم‌های تصویر بدست آوردن اطلاعات از کاربر در حین تماشای محتوا با استفاده از سخت‌افزارهایی نظیر موشواره و ردیاب چشم است. این دستگاه‌ها ناحیه مورد علاقه کاربر را در آن لحظه مشخص می‌کنند. مشخصاً اولین چالش در این روش، حضور

^۱ Gaze Point

^۲ Human Visual Sensitivity

^۳ Saliency

همیشه‌گی دستگاه مورد نظر است، که همیشه این امکان فراهم نیست. دستگاه‌های دیگری نیز برای دریافت ناحیه مورد علاقه استفاده می‌شود نظیر لمس انگشت^۱، قلم نوری، کنترل‌کننده‌ی حرکت و تشخیص دهنده‌ی حالت با استفاده از دوربین.

در مقاله [۱۷] روشی برای اصلاح نرخ بیت^۲ در دامنه‌ی DCT ارائه شده است که نقطه تمرکز^۳ به صورت صریح با استفاده از کلیک موشواره و ارسال به پیمانۀ انتقال در کانال بازخورد فراهم می‌شود. همچنین در مقاله [۱۸] یک سامانه چند کیفیت‌ی بلادرنگ foveated توسعه داده شده است که نقطه‌ی foveation را از دستگاه موشواره یا ردیاب چشم بدست می‌آورد.

استفاده از دستگاه‌های تعیین‌کننده‌ی نقطه توجه با سربرار تاخیر انتقال و پردازش بین ورودی کاربر در سمت گیرنده و تولید نقشه و کد کردن در سمت انتقال دهنده همراه است. لذا دومین چالش در این روش‌ها وجود تاخیری است که به سامانه می‌افزاید. در مقاله [۱۹] تلاش شده تا این مشکل را با اعمال پنجره‌ی توجه بصری^۴ به جای نقطه‌ی تمرکز حل کند. این پنجره در توسعه‌ی سامانه بلادرنگ بازکردن ادراکی^۵ استفاده شده است. البته افزودن پنجره توجه بصری دقت را کاهش می‌دهد که در برخی کاربردها نظیر پزشکی قابل تحمل نیست. از آنجایی که این روش مبتنی بر کاربر است، قابل پیاده‌سازی در کاربردهای پخش جمعی^۶ نیست. همچنین این روش نیاز به استفاده از کانال بازخورد^۷ دارد که خود چالشی دیگر در استفاده از این روش می‌باشد.

در کاربردهایی که بتوان تاخیرات ممکن را کاهش داد، نیاز به استفاده از پنجره توجه بصری نیز نخواهد بود، لذا دقت بالا حفظ خواهد شد؛ همچنین اگر کانال فیدبک نیز وجود داشته باشد این روش بسیار مناسب است. بنابراین در مواردی نظیر پزشکی از راه دور^۸ و ویدئوی نظارتی^۹ که تاخیر انتقال بسیار کم است و قادر به کاهش تاخیر پردازش نیز هستیم این روش مناسب و قابل اجرا است.

۲-۱-۳- حساسیت بصری

حساسیت بصری از اصل وجود آستانه‌ی حساسیت در انسان‌ها الهام گرفته است. آستانه حساسیت نسبت به عواملی نظیر خطا، اعوجاج یا سطح نویز، است که اگر از این آستانه حساسیت بالاتر نویم کاربر متوجه هیچ یک از موارد ذکر شده در تصویر و صدا نخواهد شد. لذا اعوجاج ناشی از کد کردن تا زمانی که به حد آستانه نرسد قابل اعمال است.

^۱ Finger touch

^۲ Rate-shaping

^۳ Fixation Point

^۴ Perceptual attention window

^۵ Perceptual Transcode

^۶ Broadcast

^۷ Feedback

^۸ Telemedicine

^۹ Surveillance Video

در نواحی با بافت شلوغ و پیچیده^۱ جزئیات قابل درک نیست و خطا قابل پنهان سازی در این نواحی می‌باشد. بنابراین یافتن این نواحی توسط معیارهایی نظیر بافت، روشنایی سپس کاهش کیفیت در این نواحی، یکی از راه‌های کاهش پهنای باند ویدئو می‌باشد. کد کردن براساس آنالیز بافت در مقالات [۲۰-۲۳] مورد توجه بوده است. در این روش‌ها فریم تصویر به نواحی با بافت پیچیده و نواحی با بافت ساده تقسیم می‌شود. نواحی با بافت پیچیده برای ساخت مدلی در سمت کدکننده برای کاهش نرخ بیت مورد استفاده قرار می‌گیرد در حالی که نواحی با بافت ساده با روش‌های مرسوم فشرده‌سازی ویدئو کد می‌شود.

در مقاله [۲۴] دو معیار حساسیت مورد توجه قرار گرفته است که عبارتند از روشنایی و بافت. در این مقاله اشاره شده است که انسان‌ها به کنتراست روشنایی نسبت به مقدار روشنایی بسیار حساس‌ترند. لذا سطح آستانه‌ی این دو معیار محاسبه شده و از ترکیب این دو، نقشه‌ای تولید کرده‌اند. با توجه به ارائه این روش در ارتباطات تصویری نظیر کنفرانس‌ها این نقشه با افزودن رنگ پوست انسان تعدیل شده تا معیار حساسیت در نواحی با رنگ پوست افزایش پیدا کند و صورت کاربر بیشتر مورد توجه قرار بگیرد.

۲-۱-۴- گودی مرکزی شبکه

در چشم انسان، یک ناحیه دایره‌ای به قطر حدود ۱٫۵ میلی‌متر بر روی شبکه و در راستای محور دید قرار دارد. این ناحیه گودی مرکزی^۲ نام دارد. چگالی سلول‌های مخروطی در این بخش از شبکه بیشتر است. این چگالی با افزایش زاویه از محور دید^۳ به شدت کاهش می‌یابد. پنجاه درصد بخش بینایی مغز به گودی مرکزی اختصاص دارد. این ناحیه اگر چه تنها حدود ۲ درجه حول محور دید را پوشش می‌دهد، اما صحنه را با بیشترین دقت مکانی^۴ درک می‌کند. با افزایش فاصله از محور دید (دید جانبی^۵) دقت مکانی بصورت لگاریتمی کاهش می‌یابد. در شکل X قدرت بینایی چشم چپ نسبت به زاویه از محور دید نمایش داده شده است.

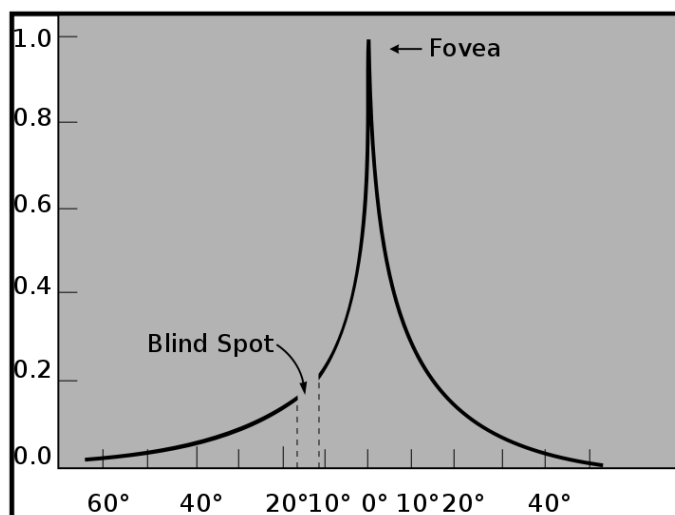
^۱ Complex texture

^۲ Fovea

^۳ Eccentricity

^۴ Resolution

^۵ Peripheral Vision



شکل ۱-۲- قدرت بینایی چشم چپ نسبت به زاویه از محور دید

بر اساس این مشاهدات، یک مدل حساسیت بصری که تابعی از فرکانس مکانی^۱ و زاویه از محور دید است، در [۱۸] ارائه گردید. در این تابع، با افزایش زاویه حساسیت کاهش می‌یابد. علاوه بر این، کاهش برای مولفه‌های با فرکانس بالاتر بیشتر است. این مدل برای پیدا کردن مولفه فرکانسی با بیشترین قابلیت دیده شدن نیز می‌تواند مورد استفاده قرار گیرد [۲۵].

۲-۱-۵- توجه بصری

ما هر روز از سامانه بینایی برای انجام کارها و ارتباط با دنیای خارج بهره می‌بریم. از جمله دیدن فیلم در تلویزیون یا دیدن مسابقه‌ی فوتبال در ورزشگاه که با استفاده از سامانه بینایی قادر به انجام‌شان هستیم. در واقع ۸۰ درصد اطلاعاتی که روزانه دریافت می‌کنیم از طریق سامانه بینایی است. قسمتی از چشم که وظیفه‌ی دریافت اطلاعات را برعهده دارد شبکیه^۲ نام دارد. شبکیه در هر ثانیه قادر است تا ۱۰ میلیارد بیت اطلاعات دریافت کند [۲۶]. همچنین ۱۰ تا ۲۰ میلیارد سلول عصبی در قشر مغز^۳ وجود دارد. این حقیقت نشان می‌دهد تا چه اندازه اطلاعات بصری به مغز ارسال می‌شود، بنابراین این غیرممکن است که مغز توانایی تحلیل تمام این اطلاعات را داشته باشد. لذا یکی از وظایف اصلی سامانه بینایی انسان، تشخیص زیرمجموعه‌های مهم بصری^۴ است که به آن زیرمجموعه‌ی برجستگی^۵ گویند. ایده‌ی یافتن نواحی برجستگی در سامانه‌های رایانه‌ای از سامانه برجستگی انسانی ناشی شده است. هرچند سامانه بینایی افراد تفاوت دارند، ولی در مجموع ارتباط قابل توجهی در تشخیص زیرمجموعه‌های برجستگی یافت شده است [۲۷].

^۱ Spatial Frequency

^۲ Retina

^۳ Cerebral cortex

^۴ Visual subnet

^۵ Salient subnet

از نقطه نظر روش محاسبه، مدل‌های برجستگی، به سه دسته قابل تقسیم می‌باشند؛ مدل‌های پایین-به-بالا که از ویژگی‌های پایه تصویر یا ویدئو، نظیر رنگ، لبه، بردار حرکت و غیره به نواحی برجسته می‌رسند. دسته‌ی دوم مدل‌های بالا-به-پایین هستند که با توجه به مفاهیم و اطلاعاتی که از موجودیت‌های معنادار اطراف‌مان قابل درک است، به نواحی مورد علاقه دست می‌یابند. در این مدل‌ها از طریق یافتن موجودیت‌هایی نظیر انسان و ماشین که مفهوم انتزاعی برای انسان دارند به مدل برجستگی مورد نظر دست پیدا می‌کنند. دسته‌ی سوم از ترکیب رویکردهای دو دسته‌ی قبل استفاده می‌کند.

توجه پایین-به-بالا در خود چشم و مسیر انتقال اطلاعات از چشم به مغز انجام می‌شود. در واقع چشم انسان قبل از ارسال تمام اطلاعات دریافت شده توسط حسگرهای موجود در چشم به مغز، قسمتی از آن اطلاعات را غربال می‌کند. به عنوان مثال، قسمت‌هایی از تصویر که تغییر روشنایی در آنها زیاد باشد، به عنوان اطلاعات مهم به مغز فرستاده می‌شود و توجه انسان را به خود جلب می‌کند. توجه بالا-به-پایین بر اساس دستوری که از مغز به چشم می‌رسد ایجاد می‌شود. در این حالت، مغز انسان بر اساس وظیفه‌ای که در دست اجرا دارد، چشم را وادار می‌کند روی قسمت مشخصی از صحنه تمرکز کرده و اطلاعات آن قسمت را برایش ارسال کند.

یکی از روش‌های کاهش نرخ بیت، استفاده از خروجی مدل‌های برجستگی است. از آنجایی که مدل‌های برجستگی، نواحی مورد علاقه‌ی کاربر را پیش‌بینی می‌کنند، لذا از خروجی این مدل‌ها می‌توان ROI تصویر را مشخص کرد. پس از تشخیص نواحی مورد علاقه با روش‌های فشرده‌سازی ویدئو براساس ROI نرخ بیت را می‌توان کاهش داد. در ادامه به بررسی روش‌های ارائه شده در این زمینه پرداخته خواهد شد.

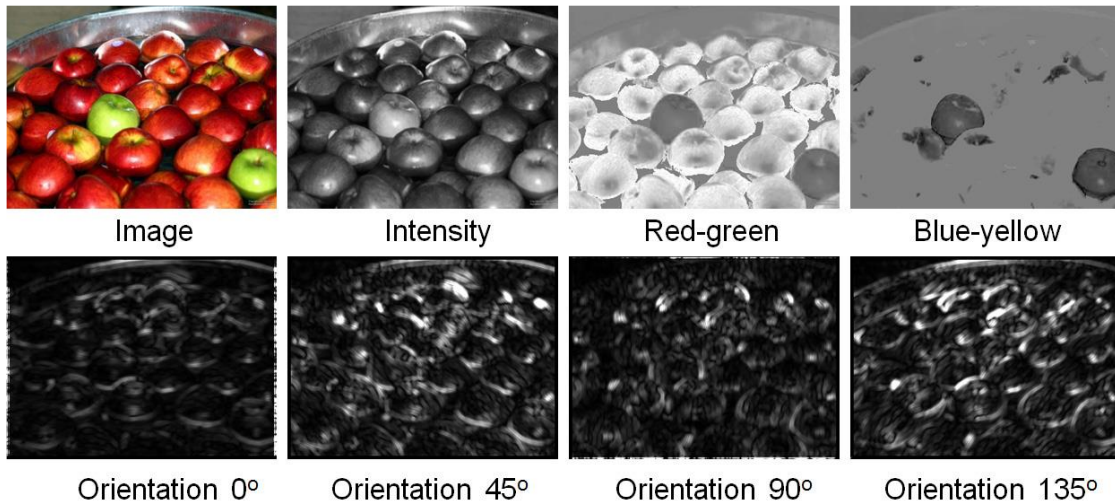
مدل کلاسیک برجستگی توسط ایتی^۱ در سال ۱۹۹۸ پیشنهاد شده است [۲۷]. در این رویکرد برجستگی بصری از تفاوت مکانی همسایه‌ها، در چند ویژگی بدست می‌آید. این رویکرد سه پیمانه اصلی دارد: استخراج ویژگی‌های پیش‌توجه^۲، ترکیبات بین‌مقیاسی نقشه‌ها و نرمال‌کردن، و ترکیب کنتراست نقشه‌ها.

در پیمانه‌ی اول، هفت ویژگی پیش‌توجه استخراج می‌شوند که شامل شدت اختلاف رنگ‌های قرمز-سبز و آبی-زرد و همچنین از چهار جهت^۳ (همانطور که در شکل ۲-۲ نشان داده شده است) استفاده کرده‌اند.

^۱ Itti

^۲ Preattentive

^۳ Orientation



شکل ۲-۲ - اعمال ویژگی‌های پیشنهادی ایتی

ایتی برای محاسبه‌ی ویژگی شدت روشنایی از رابطه زیر استفاده می‌کند:

$$I = \frac{r + g + b}{3} \quad (۱-۲)$$

که در آن r, g, b مقادیری بین صفر و یک هستند که نماینده‌ی کانال‌های آبی، سبز و قرمز تصویر ورودی می‌باشند. به این ترتیب با داشتن کانال‌های آبی، سبز و قرمز می‌توان رنگ‌های مخالف قرمز-سبز و آبی-زرد را استخراج کرد. دلیل انتخاب این رنگ‌ها تئوری هرینگ^۱ است. در این تئوری ادعا شده است که سامانه بینایی انسان برخی رنگ‌ها (از جمله قرمز و سبز، آبی و زرد) را نمی‌تواند به صورت همزمان درک کند. بنابراین رنگ‌های مخالف قرمز-سبز و آبی-زرد، رنگ‌های پردازش اطلاعات در سامانه بینایی انسان هستند. برای محاسبه‌ی رنگ‌های ضد، ابتدا r و g را با متغیر I ($I < 0.1$) نرمال می‌کنند تا رنگ‌ها از شدت جدا شوند. لذا از آنجا که بیننده به اختلاف رنگ‌ها در روشنایی کم حساس نیست نرمال شده‌ی کانال قرمز به صورت زیر قابل محاسبه است:

$$\hat{r} = \begin{cases} r/I, & \text{if } I \geq 0.1 \\ 0, & \text{otherwise} \end{cases} \quad (۲-۲)$$

همچنین نرمال شده‌ی رنگ‌های سبز و آبی به همین شیوه انجام می‌شود:

$$R = \max(0, \hat{r} - \frac{\hat{g} + \hat{b}}{2}) \quad (۳-۲)$$

$$G = \max(0, \hat{g} - \frac{\hat{r} + \hat{b}}{2}) \quad (۴-۲)$$

$$Y = \max(0, \frac{\hat{r} + \hat{g}}{2} - \frac{|\hat{r} - \hat{g}|}{2} - \hat{b}) \quad (۵-۲)$$

^۱ Hering

رنگ‌های ضد، آبی-زرد و قرمز-سبز، از اختلاف کانال‌های رنگ بدست آمده، قابل محاسبه است. برای اثر بخشی بیشتر، مقاله‌ی [28] معادلات زیر را برای بدست آمدن رنگ‌های ضد پیشنهاد داده است:

$$RG = \frac{r - g}{\max(r, g, b)} \quad (6-2)$$

$$BY = \frac{b - \min(r, g)}{\max(r, g, b)} \quad (7-2)$$

برای جلوگیری از نوسان در RG و BY در روشنایی کم، هرگاه $\max(r, g, b) < 0.1$ باشد مقادیر RG و BY را صفر قرار می‌دهند.

ویژگی جهت با محاسبه‌ی کانولوشن^۱ کانال شدت و فیلتر گابور^۲ در چهار زاویه به دست خواهد آمد. این محاسبات برای زوایای صفر، ۴۵، ۹۰ و ۱۳۵ درجه انجام می‌گردد:

$$O(\theta) = |I * G_0(\theta)| + |I * G_{\pi/2}(\theta)| \quad (8-2)$$

در (۸-۲)، θ زاویه‌ی اعمال فیلتر است [۲۶]:

$$G_{\Psi}(x, y, \theta) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\delta^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \Psi\right) \quad (9-2)$$

(۹-۲) فیلتر Gabor با فاز Ψ و نسبت تصویری^۳ γ و انحراف معیار δ و طول موج^۴ λ را محاسبه می‌کند. همچنین مختصات (x', y') انتقال یافته‌ی^۵ (x, y) با توجه به زاویه θ است که به صورت زیر بدست می‌آید:

$$x' = y \sin(\theta) + x \cos(\theta), \quad (10-2)$$

$$y' = y \cos(\theta) - x \sin(\theta). \quad (11-2)$$

در ادامه نقشه‌های ویژگی تصویر بدست می‌آید. این نقشه‌ها با اعمال عملیات محاسبه‌ی اختلافات مرکز-پیرامون بین یک مرکز که در مقیاس کوچک c تعریف شده و یک سری نقاط پیرامون که در مقیاس بزرگ تر s هستند، به دست می‌آیند. در اینجا برای بدست آوردن نقشه‌ی روشنایی، ۶ نقشه‌ی $I(c, s)$ خواهیم داشت. ($c \in \{2, 3, 4\}$ و $\delta \in \{2, 3\}$ و $s = c + \delta$) به این ترتیب نقشه‌های ویژگی مربوط به شدت روشنایی از رابطه‌ی زیر محاسبه می‌شوند:

^۱ Convolution

^۲ Gabor

^۳ Aspect ratio

^۴ Wavelength

^۵ Transform

$$I(c, s) = |I(c) \ominus I(s)| \quad (۱۲-۲)$$

مجموعه‌ی بعدی نقشه‌ها هم به طور مشابه برای رنگ‌ها به دست می‌آیند. همچنین علاوه بر چهار نقشه‌ی اولیه نقشه‌های دیگری هم برای تضادهای قرمز/سبز و سبز/قرمز، آبی/زرد و زرد/آبی با رابطه زیر ایجاد می‌شوند [۲۶]:

$$RG(c, s) = |(R(c) - G(c)) \ominus (G(s) - R(s))| \quad (۱۳-۲)$$

$$BY(c, s) = |(B(c) - Y(c)) \ominus (Y(s) - B(s))| \quad (۱۴-۲)$$

نقشه‌های جهت، $O(c, s, \theta)$ ، نیز تضاد بین جهت‌گیری‌های داخلی مقیاس‌های مرکز و پیرامون را به دست می‌دهند:

$$O(c, s, \theta) = |O(c, \theta) \ominus O(s, \theta)| \quad (۱۵-۲)$$

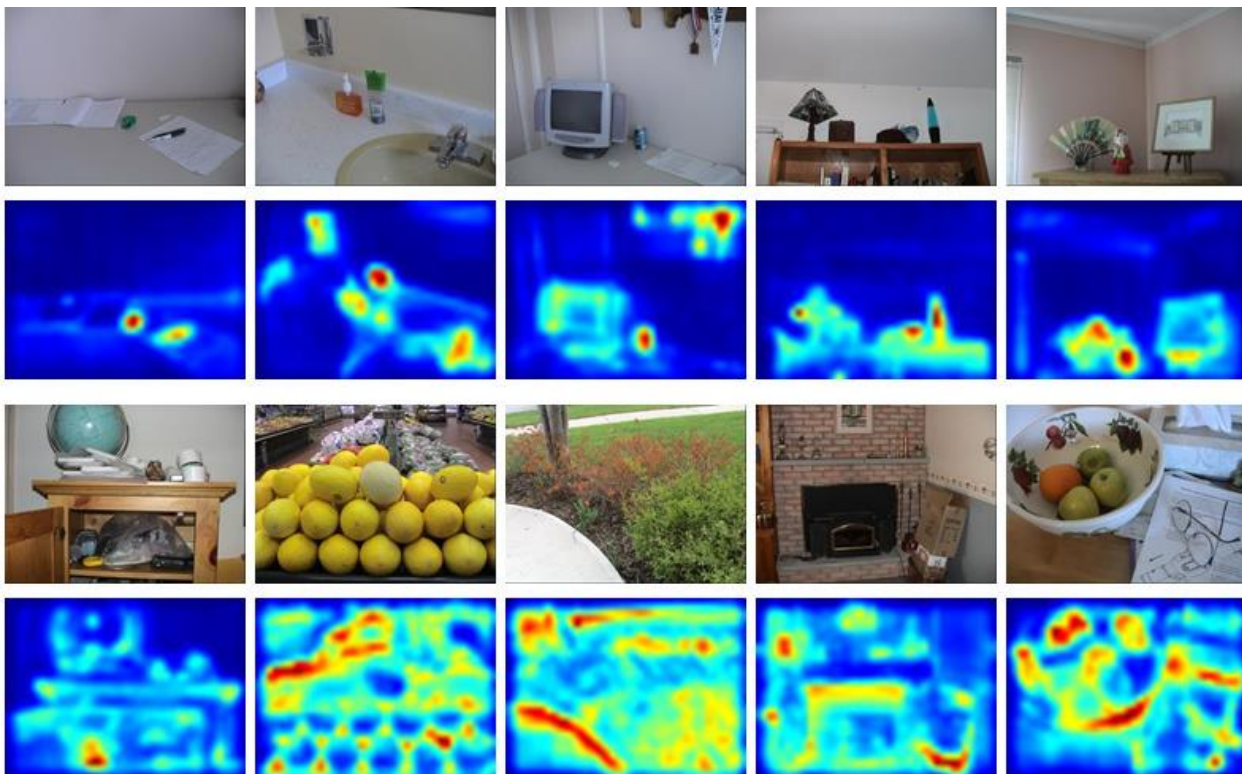
با توجه به توضیحات داده شده در نهایت ۴۲ نقشه‌ی تضاد^۱ بدست می‌آید (۶ نقشه برای شدت روشنایی، ۱۲ نقشه برای رنگ، و ۲۴ عدد برای جهت). در مرحله بعد باید این نقشه‌ها با هم ترکیب شده و نقشه‌ی نهایی تولید شود. از آنجایی که این نقشه‌ها نرمال نیستند، ابتدا نرمال شده و سپس به صورت خطی ترکیب می‌شوند. در شکل ۲-۳ مدل ایتی بر روی تعدادی تصویر اعمال شده است که نتایج قابل مشاهده است. همانطور که از این تصاویر مشخص است در تصاویر ردیف اول مدل عملکرد خوبی دارد و به خوبی نواحی برجستگی را مشخص کرده است ولی در تصاویر ردیف سوم نواحی برجستگی بسیار بد تعیین شده‌اند و مدل بسیار ضعیف عمل کرده است. مشخص است که این مدل تصاویر با اشیای کم و غیر شلوغ به خوبی عمل می‌کند ولی در تصاویر شلوغ و پر شی عملکرد مناسبی ندارد. دلیل این امر استفاده‌ی مدل ایتی از تضادهای محلی^۲ می‌باشد. بسیاری از مدل‌ها بر این اساس کار می‌کنند. این دسته مدل‌ها بر اساس یافتن بی‌نظمی‌های محلی^۳ عمل می‌کنند از جمله [۲۹، ۳۰] که نیز از این ویژگی بهره برده‌اند. برای مقابله با این پدیده روش‌های یافتن بی‌نظمی‌های کلی^۴ پیشنهاد شد.

^۱ Contrast

^۲ Local Contrast

^۳ Local Irregularity

^۴ Global Irregularity Detection



شکل ۲-۳ - اعمال مدل ایتی بر روی مجموعه داده‌ی [۳۱] Toronto-120. تصاویر ردیف دوم بهترین جواب‌های اعمال مدل ایتی بر روی این مجموعه داده و تصاویر ردیف چهارم بدترین جواب‌های مدل ایتی بر روی مجموعه داده‌ی ذکر شده است [۲۷].

برای یافتن بی‌نظمی‌های کلی می‌توان از روشی بسیار ساده استفاده کرد به این صورت که رنگ‌های استفاده شده در تصویر را به صورت آماری بدست آورد و نواحی که از رنگی‌های با احتمال کمتر در تصویر حضور دارند به عنوان نواحی مهم‌تر انتخاب شوند. این روش بسیار ساده، ایده پایه‌ی مقاله‌ی [۳۲] می‌باشد. البته این روش نیز به دلیل سادگی در تصاویر شلوغ بد عمل می‌کند. روشی دیگر در مقاله‌ی [۳۳] با استفاده از مدل گرافی، بی‌نظمی‌های کلی تصویر را می‌یابد. این روش تصویر را به صورت یک گراف کاملاً متصل^۱ در نظر می‌گیرد که گره‌ها نماینده‌ی زیر مجموعه‌های بصری^۲، نظیر ماکروبلوک در تصویر، و وزن یال‌ها با توجه به عدم شباهت^۳ و نزدیکی^۴ گره‌های اتصال به یکدیگر ایجاد می‌شود. بر اساس وزن‌های حاصله پیمایش اتفاقی^۵ در گراف ایجاد می‌شود که در نهایت گره‌هایی که کمتر در پیمایش دیده شده‌اند نواحی برجستگی را نمایش می‌دهند زیرا نواحی خاص و مهمتری هستند. برای یافتن شباهت و نزدیکی، از فاصله‌ی ویژگی‌های هر گره و فاصله‌ی مکانی گره‌ها از هم، استفاده می‌کند. این ویژگی‌ها می‌تواند شامل مواردی نظیر جهت، رنگ، شدت و ویژگی‌های

^۱ Fully connected

^۲ Visual subset

^۳ Dissimilarities

^۴ Closeness

^۵ Random Walk

زمانی نظیر بردارهای حرکت باشد. این روش زمان اجرای بالایی دارد که ناشی از پیمایش‌های این گراف استفاده می‌شود. این مدل نیز در صحنه‌هایی که نواحی برجستگی به هم نزدیک و متراکم هستند بسیار خوب عمل می‌کند زیرا عدم شباهتشان با دیگر نواحی بالا و نزدیکی به نواحی شبیه نیز بالا است. در صحنه‌هایی که اهداف برجستگی زیاد و تصویر کمی شلوغ است این مدل به خوبی عمل نمی‌کند.

با الهام گرفتن از روش ارائه شده توسط هارل مدل‌های دیگری نیز این مدل گرافی را توسعه دادند [۳۴, ۳۵]. [۳۶] مدل گرافی را در دو سطح یافتن بی‌نظمی‌های کلی و یافتن بی‌نظمی‌های محلی ارائه کرد که به دو گراف مجزا از این دو سطح منتج می‌شود و با استفاده از این دو گراف نواحی برجستگی را تعیین می‌کند. برخی دیگر از روش‌ها نیز با ترکیب بی‌نظمی‌های کلی و بی‌نظمی‌های محلی مدل برجستگی خود را تولید کرده‌اند [۳۷-۳۹].

روش‌های زیادی تاکنون معرفی شده است. ولی این مدل‌ها اغلب کارایی لازم را نداشته و معیارهای مقایسه توسط خود ارائه‌دهنده‌ها ارائه می‌شد و مقایسه‌ی عادلانه‌ای وجود نداشت. به همین دلیل Judd محکی^۱ برای سنجش مدل‌های مختلف ارائه کرده است. محک Judd [۴۰] بهترین روش‌های ارائه شده با توجه به معیارهای مقایسه‌ی روش‌های توجه بصری را یافته و رتبه‌بندی کرده است. در جدول ۱-۲ شش مدل برتر و سال آن‌ها از نظر معیار AUC، بر طبق محک Judd نشان داده شده است. در ادامه این شش روش توضیح داده خواهد شد:

جدول ۱-۲- شش مدل برتر در محک جاد

نام مدل	سال انتشار
eDN	۲۰۱۴
BMS	۲۰۱۳
CovSal	۲۰۱۳
FES	۲۰۱۱
JUDD	۲۰۰۹
GBVS	۲۰۰۶

در مقاله [۴۱] روش توجه پایین-به-بالا توضیح داده شده است که با حد آستانه گرفتن^۲ از هر کانال رنگی تعداد زیادی نقشه‌های دودویی^۳ ایجاد شده است، که بر اساس آن، مدل را آموزش و ارزیابی می‌کنند. این مقاله بر طبق اصل Gestalt's Figure Ground Segregation بنا نهاده شده است. بر اساس این اصل، ما به محیط اطراف به صورت کلی نگاه می‌کنیم و نه به صورت بخش‌های جدای از

^۱ Benchmark

^۲ Thresholding

^۳ Boolean

هم، همچنین در هر تصویر پیکرهایی^۱ وجود دارند که توجه نگاه ما به آن‌هاست. پیکر، اصطلاحی است که می‌تواند شامل رنگ، عمق حرکت و هر چیزی که باعث شود آن قسمت از تصویر از پشت زمینه متمایز شود، باشد. در این مقاله رنگ را به عنوان تنها ویژگی‌ای که باعث متمایز شدن پیکرها می‌شود، در نظر گرفته است. لذا با تفکیک‌سازی کانال‌های رنگ و اعمال آستانه بر روی هر کانال رنگی و استفاده از داده‌های ردیاب‌چشم، نواحی برجستگی را تعیین می‌کند.

Judd در مقاله‌ی [۴۲] روشی را ارائه کرد تا با استفاده از ویژگی سطح پایین و بالای تصویر، مدل را آموزش دهد. برای این کار در مجموع از ۳۳ ویژگی استفاده کرده است. برای آموزش مدل از SVM بهره برد. جهت پیاده‌سازی مدل، ابتدا همه‌ی تصاویر به اندازه‌ی ثابت 200×200 درآمده و سپس ۳۳ ویژگی تصویر محاسبه می‌گردد. بر روی نقطه‌ای که چشم، بر روی آن متمرکز بوده فیلتر گوسی زده و از ویژگی‌های حاصل از آن ناحیه جهت آموزش استفاده می‌کند.

در مقاله‌ی [۴۳] روشی سریع برای یافتن ویژگی‌های بهینه‌ی نواحی برجستگی در تصویر پیشنهاد شده است. برای این کار از مدل سلسله‌مراتبی شبکه‌ی عصبی برای آموزش بهره برده‌اند. در این کار از مدل لوله‌ای که Judd پیشنهاد داده است برای آموزش مدل استفاده کردند و اصلی‌ترین کار انجام شده بهبود انتخاب ویژگی‌ها و یافتن ویژگی‌های سطح بالا جهت بهبود تشخیص نواحی برجستگی است.

توکلی در مقاله‌ی [۴۴] نگاشت برجستگی پیشنهادی‌اش را براساس اصل فاصله از مرکز توسعه داده است. پس از آن که محک Judd نشان داد تنها در نظر گرفتن دایره‌ای در مرکز تصویر، به عنوان نواحی برجستگی کارا تر از بسیاری روش‌های پیچیده در تعیین نگاشت برجستگی است، اعمال تاثیر فاصله از مرکز در بسیاری از مدل‌ها استفاده شد. توکلی از این معیار در بهبود روش خود استفاده کرده است. همچنین در این روش با استفاده از ویژگی‌های تصویر و شبکه‌ی بی‌زین ساده، مدل خود را توسعه داده است. به علت سرعت بالای این مدل و عدم نیاز به آموزش این روش قابل استفاده در کاربردهای بلادرنگ می‌باشد.

در مقاله [۴۵] از کواریانس تصویر جهت جداسازی ناحیه‌ای از تصویر که با محیط اطراف متفاوت بوده، استفاده کرده است. برای این کار ابتدا تصویر را به نواحی غیر همپوشان می‌شکند؛ سپس این نواحی را با برخی ویژگی‌ها نظیر رنگ، شدت نور، جهت و ... توصیف می‌کند. با توجه به توصیفی که برای نواحی شکسته شده استفاده کرده، فاصله را از همسایه‌های محاسبه می‌کند. نواحی برجستگی را نواحی‌ای در نظر می‌گیرد که بیشترین تفاوت را از همسایه‌ها در تصویر داشته باشند. در پیاده‌سازی، همه‌ی تصاویر را به اندازه 512×512 تبدیل می‌کند و پنج بار الگوریتم را با شکستن تصویر به بلاک‌های مربعی با ضلع ۸، ۱۶، ۳۲، ۶۴ و ۱۲۸ اجرا می‌کند. در مقاله [۳۳] روش توجه بصری بر اساس نظریه گراف پیشنهاد شد که پیشتر این روش توضیح داده شد.

^۱ Figure

با توجه به تحقیقات ما، تنها روشی که برای محتوای بازی ارائه شده است مقاله‌ی [۴۶] می‌باشد. مدل پیشنهادی این مقاله از ترکیب دو نگاشت تشکیل می‌یابد. نگاشت اول با استفاده از مدل پایه‌یایی که پیشتر به آن اشاره شده است، محاسبه می‌شود. نگاشت دوم به صورت بالا-به-پایین به تصاویر نگاه می‌کند و با استفاده از محاسبه‌ی سیگنال فوری به ازای نگاشت‌های حاصل از ویژگی‌های مدل پایه‌یایی بدست می‌آید. ویژگی‌های حاصل از این دو نگاشت با توجه به خیره‌گی چشم به صورت یک روش دسته‌بندی نگاه شده و با استفاده از کمینه‌مربعات خطی^۱ آموزش داده شده‌اند.

۲-۲- کد کردن ادراکی

از میان تحقیقات انجام شده در حوزه کاربرد CG، سهم مقالاتی که به چالش پهنای‌بند پرداخته‌اند، بسیار کمتر از سایر زمینه‌ها، از جمله کاهش تاخیر، است. در [۴۷] پیشنهاد شده است برای تطبیق جویبار ویدئو با پهنای‌بند موجود و سایر ویژگی شبکه برخی فریم‌ها حذف شوند. در [۴۸] توزیع بار بین کارسازهای CG بر اساس یک رابطه‌ی استخراج شده میان متوسط امتیاز نظر بازیکن‌ها و نرخ بیت و نرخ فریم ویدئوی بازی انجام شده است. با این وجود، این کارها غیرادراکی هستند و ویژگی‌های سامانه بینایی چشم انسان را در نظر نگرفته‌اند.

در [۴۹] با برقراری ارتباط میان موتور بازی و کدکننده ویدئو و در اختیار قرار دادن اطلاعات اشیای بازی به کدکننده، پیچیدگی پیمانه جستجوی حرکت کدکننده ویدئو تا ۳۹٪ کاهش داده شده است بطوریکه نرخ بیت و PSNR ویدئوی کدشده‌ی خروجی تقریباً ثابت مانده است. در [۵۰] نیز از روش مشابهی برای کاهش زمان کدکردن استفاده شده است. آنها همچنین سعی کرده‌اند نقاط مهم بازی را با استفاده از اطلاعات عمق بدست آورند. با این وجود، استفاده‌ی تنها از اطلاعات عمق برای انواع بازی‌ها تعمیم‌پذیر نیست زیرا در برخی از بازی‌ها اشیایی در نزدیکی دوربین وجود دارند که در جریان بازی نیستند و بازیکن خیلی کم به آنها توجه می‌کند. در [۵۱] سعی شده است تا با استفاده از تشابهاتی که بین فریم‌های بازیکن‌ها در بازی‌های چندکاربره وجود دارد، و همچنین استفاده از ویژگی‌های شبکه‌های اقتضایی برای ارسال‌های نقطه به نقطه، پهنای‌بند مورد نیاز بازیکن‌ها کاهش داده شود. در [۵۲] تخصیص منابع در سامانه‌های CG در قالب یک مسئله بهینه‌سازی قیددار فرایندهای تصادفی فرمول‌بندی و حل شده است. یکی از قیدهای مسئله نرخ‌بیت کاربر است که نباید از حدی کمتر باشد. البته این حد بصورت دستی تنظیم می‌شود. در [۵۳] پیشنهاد شده است تا به جای ارسال کامل ویدئو از سمت کارساز به کارخواه، بخش اساسی فرایند به تصویرکشیدن در سمت کارخواه انجام شده و جزئیات در قالب ویدئو و به عنوان لایه افزایشی برای کارخواه ارسال شود. این روش اگرچه باعث کاهش خوبی در پهنای‌بند می‌شود، اما در آن امکان استفاده از کدکننده‌های

^۱ Linear Least-Squares

موجود در بازار وجود ندارد. همچنین قدرت پردازشی کارخواه‌ها باید بگونه‌ای باشد که از عهده به تصویر کشیدن لایه پایه برآیند.

روش‌های متفاوتی برای بکارگیری مدل‌های ادراکی در فرایند فشرده‌سازی ویدئو در کاربردهایی غیر از بازی ارائه شده است [۱۶]. در این بخش، این روش‌ها را در قالب سه دسته بررسی می‌کنیم.

۲-۲-۱- پیش پردازش ویدئو

این روش‌ها قبل از ارسال ویدئو به کدکننده، بر اساس مدل ادراکی ویدئوی را تغییر می‌دهند. در [۵۴] نواحی غیر مهم ویدئو مات^۱ شده و نواحی مهم با همان کیفیت اولیه رها می‌شوند. از آنجا که مولفه‌های با فرکانس بالا در طی فرایند مات‌سازی از بین می‌روند، افزونگی کاهش یافته و باعث افزایش نرخ فشرده‌سازی می‌شود. در [۵۵]، نویسندگان نشان دادند در کاربرد ویدئوهای غیرتعاملی، با اعمال برون‌خط فیلتر Foveation قبل از ارسال ویدئو به کدکننده H.264/AVC، می‌توان بصورت متوسط ۲۰ درصد نرخ بیت را کاهش داد. با این وجود، مات کردن باعث بروز اعوجاج‌های غیر قابل چشم‌پوشی است و کیفیت کیفی را کاهش می‌دهد [۵۶]. گروه دیگری از روش‌های این دسته، با استفاده از دستکاری صحنه^۲ نرخ بیت را کنترل می‌کنند. در [۵۷]، قرار گرفتن یا نگرفتن اشیاء در صحنه بر اساس میزان پهنای‌باند در دسترس فعلی و طبق یک الگوریتم اولویت‌دهی اشیاء می‌باشد. از آنجا که در این دسته از روش‌ها کنترل کیفیت و کد کردن جدای از هم انجام می‌شوند، نرخ فشرده‌سازی آنها در بین روش‌های کد کردن ادراکی کمترین است [۱۶].

۲-۲-۲- جفت کدکننده/واکدکننده

این روش‌ها برای اعمال مدل‌های ادراکی جفت کدکننده و واکدکننده مخصوص ارائه می‌کنند. در [۵۸] یک جفت کدکننده/واکدکننده انتها به انتها پیشنهاد شده است که از انواع الگوریتم‌های کد کردن مبتنی بر گودی مرکزی استفاده می‌کنند. کد کردن مبتنی بر سنتز^۳ نیز در این دسته قرار می‌گیرد. در این نوع کد کردن یک پیمانه استاندارد برای کد کردن نواحی بدون بافت و یک پیمانه دیگر برای کد کردن اطلاعاتی که بافت سایر نواحی را توصیف می‌کند، وجود دارد [۲۳]. با این وجود، از آنجا که این روش‌ها نیاز به جایگزینی واکدکننده‌های موجود با واکدکننده‌های جدید را دارند، از نظر اقتصادی مطلوب نیستند [۱۶].

^۱ Blur

^۲ Scene Composition

^۳ Synthesis-based Video Coding

۲-۳- تنظیم پارامترهای کدکننده

این دسته از روش‌ها از امکان تغییر پارامترهای کدکننده ویدئو برای نواحی مختلف، که ذاتاً در طراحی آن وجود دارد، استفاده می‌کنند. یک گروه از این روش‌ها، پارامتر پیمان‌بندی (QP) را طوری تنظیم می‌کنند که کیفیت نواحی مهم بالا و کیفیت سایر نواحی پایین باشد. در [۵۹]، به ناحیه با بیشترین اهمیت بصورت دستی یک QP اختصاص داده می‌شود. سپس به ترتیب اهمیت نواحی از بالا به پایین، هر ناحیه دیگر L واحد بزرگتر تنظیم می‌شود که L یک عدد صحیح مثبت است. آنها اهمیت هر ناحیه را بر اساس یک مدل توجه چندرسانه‌ای تعیین می‌کنند. در [۶۰]، از اختلاف فریم و اطلاعات رنگ پوست به عنوان یک ضریب برای تعیین QP هر ناحیه استفاده شده است. در چند مقاله نیز رابطه بین QP و اعوجاج مدل شده است [۶۱-۶۳]. در [۶۴]، از یک مدل JND و Foveation برای تنظیم QP استفاده شده است. آنها از مدل خود برای تنظیم ضریب لاگرانژ بهینه‌سازی نرخ اعوجاج استفاده کردند.

الگوریتم‌های اختصاص بیت ادراکی نیز در این دسته قرار می‌گیرند. در روش‌های معمول کنترل نرخ، مانند [۶۵]، اختصاص بیت با ترتیب پیمایش سطری^۱ انجام می‌پذیرد. وقتی تعداد بیت‌های هدف برای یک واحد کدشونده مشخص شد، آن واحد کد شده و بعد از مصرف واقعی بیت‌ها، بیت‌های باقیمانده بین واحدهای کد نشده بطور مساوی تقسیم می‌شود. اگر پیچیدگی کدشدن بین واحدهای کدشونده بطور قابل توجهی متفاوت باشد، این الگوریتم‌ها بیت‌ها را در واحدهای کدشونده اولیه هدر می‌دهند و واحدهای آخری با کمبود بیت و در نتیجه کاهش کیفیت مواجه می‌شوند. برای جلوگیری از این مشکل، الگوریتم‌های کنترل نرخ باید بر اساس محتوای ویدئو و ویژگی‌های ادراکی اختصاص بیت انجام دهند. در [۶۶]، نرخ بیت هدف هر ناحیه بر اساس پیچیدگی‌های ماکروبلوک‌های موجود در آن ناحیه تنظیم می‌شود. آنها برای تخمین پیچیدگی، از MSE استفاده کرده‌اند. در [۶۷]، نرخ بیت هدف هر ماکروبلوک بر اساس فعالیت بافت و حرکت آن تعیین می‌شود.

گروه دیگری از روش‌های این دسته، مسئله اختصاص بیت را به شکل یک مسئله قیددار بهینه‌سازی فرمول‌بندی می‌کنند. نویسندگان [۶۸] کیفیت عینی ویدئو در یک نرخ بیت مشخص را بیشینه می‌کنند. نویسندگان [۵۶] دوگان این مسئله را حل می‌کنند. آنها نرخ بیت ویدئو در یک کیفیت مشخص را کمینه می‌کنند. هر دوی این روش‌ها، مقدار اعوجاج MSE را بر اساس میزان توجه وزن می‌دهند. مدل نرخ-اعوجاج ارائه شده در این رساله نیز در این دسته قرار دارد. با این تفاوت که علاوه بر توجه مکانیزم گودی چشم را نیز در اعوجاج در نظر می‌گیرد.

^۱ Raster Order Scan

۲-۳- معیارهای کیفیت ادراکی ویدئو

معیارهای کیفیت فشرده‌سازی ویدئو به دو دسته تقسیم می‌شوند: معیارهای کمی^۱ و معیارهای کیفی^۲. معیارهای کمی، کیفیت ویدئوی کد شده را با استفاده از روش‌های ریاضی محاسبه می‌کنند. در حالی که معیارهای کیفی کیفیت ویدئو را از دیدگاه بیننده بررسی می‌کنند. ارزیابی کیفیت به صورت کیفی بسیار قابل اعتمادتر و دقیق‌تر از ارزیابی کیفیت به صورت کمی می‌باشد. مشکل معیارهای کیفی زمان‌گیر و هزینه‌بر بودنشان است. بنابراین معیارهایی کمی توسعه داده شده‌اند تا به صورت خودکار کیفیت ادراکی بیننده‌ها را تخمین زنند.

۲-۳-۱- معیارهای کیفی

برای محاسبه کیفیت ویدئو به صورت کیفی، ویدئوی کد شده به بیننده نمایش داده می‌شود و از او خواسته می‌شود به کیفیت ویدئو نمره دهد. معیارهای محیطی و محتوایی زیادی بر روی نتیجه کیفیت کیفی تاثیرگذار است. لذا استانداردهایی برای نحوه‌ی آزمون این نوع ارزیابی در نظر گرفته شده است تا از عوامل تاثیرگذار بیرونی جلوگیری کرده و نتایج قابل اعتمادی حاصل شود؛ و در صورت اجرای دوباره، نتایج جدید مشابه نتایج قبلی بدست آید. ارزیابی کیفی با روش استاندارد Double Stimulus Continuous Quality Scale، در توصیه نامه ITU-R Recommendation 1974-1997 بیان شده است. در این رساله نیز، برای ارزیابی کیفی از این استاندارد استفاده شده است.

۲-۳-۲- معیارهای کمی

معیارهای کمی کیفیت به سه دسته تقسیم می‌شوند. دسته اول منبع-دسترس‌پذیر^۳ هستند. این معیارها زمانی که سیگنال اصلی کاملاً در دسترس باشد مورد استفاده قرار می‌گیرد. دسته دوم منبع-نیمه‌دسترس‌پذیر^۴ هستند. این معیارها زمانی که سیگنال اصلی کاملاً در دسترس نباشد مورد استفاده قرار می‌گیرند. دسته آخر معیارهایی هستند که سیگنال اصلی به‌هیچ‌وجه در دسترس نباشد که به آن منبع-بدون‌دسترس^۵ گویند. در آزمایش‌های این رساله از دسته‌ی اول، منبع دسترس‌پذیر، استفاده شده است.

بسیاری از تکنیک‌های فشرده‌سازی، از PSNR برای اثبات کارایی‌شان استفاده می‌کنند. PSNR معیاری است که میزان کاهش کیفیت در کل صحنه را نسبت به صحنه‌ی اصلی مقایسه می‌کند؛ در

^۱ Objective

^۲ Subjective

^۳ Full-reference

^۴ Reduced-reference

^۵ No-reference

حالی که اساس تکنیک‌های فشرده‌سازی ادراکی بر حفظ کیفیت در نواحی مهم و مورد علاقه تصویر است. لذا معیاری که تمام صحنه را در نظر بگیرد معیار منصفانه‌ای برای مقایسه نیست. PSNR قابلیت جایگزینی با معیارهایی که ارتباط نزدیکتری به معیارهای کیفی کیفیت دارند را دارا می‌باشد. برخی از معیارهایی که برای محاسبه‌ی کیفیت کمی فشرده‌سازی ادراکی پیشنهاد شده‌اند، از تغییر فرم ساده‌ی PSNR استفاده کرده‌اند. برخی دیگر از MSE وزن دار استفاده کرده‌اند. آنها وزن‌ها را بر اساس اهمیت ادراکی نواحی تعیین می‌کنند. به این ترتیب معیارهای مختلفی از جمله 1 FPSNR، PPSNR و SPSNR ارائه شده‌اند. با توجه به مقاله‌ی [۱۶]، این معیارها منصفانه نیستند و با توجه به مدل فشرده‌سازی ادراکی، که ارائه داده‌اند معیار مقایسه را طوری پیشنهاد داده‌اند که بیشترین کارایی را، در مدل‌شان نشان دهند. بنابراین معیار EWPSNR که ویژگی منصفانه بودن و نزدیکی به معیارهای کیفی را داراست، بهتر است استفاده شود [۵۶].

از EWPSNR برای محاسبه‌ی کیفیت ویدئوی کد شده از داده‌ی ردیاب چشم استفاده می‌کنند. با این کار می‌توان تخمین کمی مناسبی از ارزیابی کیفی داشت. برای این کار EWPSNR پس از محاسبه‌ی EWMSE و با توجه به معادلات زیر بدست می‌آید [۵۶]:

$$EWMSE = \frac{1}{MN \sum_{x=1}^M \sum_{y=1}^N w_{x,y}} \sum_{x=1}^M \sum_{y=1}^N (w_{x,y} \cdot (I'_{x,y} - I_{x,y})^2) \quad \text{رابطه (۱۶-۲)}$$

$$EWPSNR = 10 * \log\left(\frac{(2^n - 1)^2}{EWMSE}\right) \quad \text{رابطه (۱۷-۲)}$$

$$w_{x,y} = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\left(\frac{(x-x_e)^2}{2\sigma_x^2} + \frac{(y-y_e)^2}{2\sigma_y^2}\right)} \quad \text{رابطه (۱۸-۲)}$$

در رابطه‌های بالا I و I' به ترتیب فریم‌های اصلی و فریم‌های کد شده هستند. همچنین M و N به ترتیب طول و عرض فریم در مقیاس پیکسل هست و n تعداد بیت لازم برای هر پیکسل تصویر است. $w_{x,y}$ وزن نقطه (x,y) است بطوریکه $\sum_{x=1}^M \sum_{y=1}^N w_{x,y} = MN$ ؛ وزن نرمال شده بر اساس نقطه تمرکز (x_e, y_e) بدست آمده از کاربر توسط دستگاه ردیاب چشم، و در نهایت σ_x و σ_y پارامترهای مربوط به فاصله و زاویه دید کاربر هستند.

۴-۲- جمع‌بندی

در این فصل انواع روش‌های جویبارسازی گرافیک رایانه‌ای، مدل‌های ادراکی، روش‌های اعمال آنها در کدکننده و معیارهای ارزیابی کیفیت ادراکی ویدئو مورد بررسی قرار گرفت. تاکنون مدل ادراکی

^۱ Foveal PSNR

مخصوص بازی‌های رایانه‌ای ارائه نشده است. البته در [۴۶] یک مدل پایین-به-بالای عمومی بر روی ویدئوهای بازی ارزیابی شده است. از سوی دیگر، مدل‌های ادراکی موجود تفاوت الگوهای توجه در بین بینندگان را در نظر نگرفته‌اند. در زمینه مدل نرخ-اعوجاج ادراکی، مدل‌های ارائه شده تنها یکی از دو سازوکار توجه یا گودی مرکزی شبکیه را در نظر گرفته‌اند. مدل‌های ادراکی و نرخ-اعوجاج ارائه شده در این رساله، کاستی‌های مطرح‌شده را در نظر می‌گیرند.

فصل سوم

مدل های ادراکی پیشنهادی

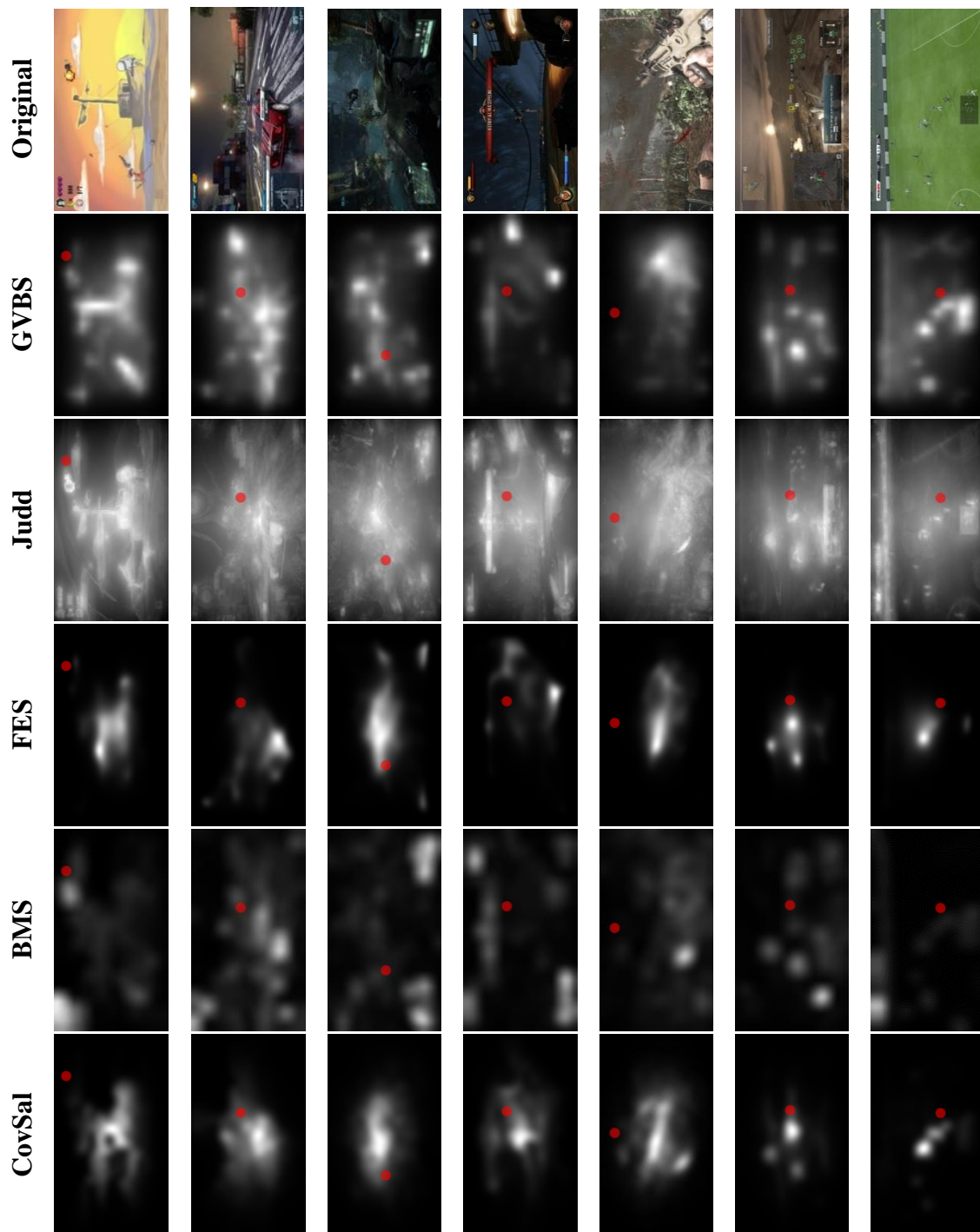
هدف این رساله تعیین اهمیت نواحی مختلف ویدئوی بازی و استفاده از آن در کاهش و کنترل نرخ بیت مورد نیاز بازی‌های مبتنی بر رایانش ابری است. این کار دو مزیت دارد. اول، تعداد افرادی که می‌توانند از این سرویس بهره‌مند شوند، افزایش می‌یابد. دوم، انتقال حجم کمتر اطلاعات در شبکه، هزینه کمتری نیز برای بازیکن‌ها خواهد داشت. بدین ترتیب ارائه این سرویس برای اپراتورهای مخابرات و تلویزیون عملی‌تر می‌شود.

فشرده‌سازی ادراکی ویدئو^۱ یکی از راه‌های موثر کاهش نرخ بیت می‌باشد. مدل‌های ادراکی تاکنون ارائه شده عملکرد مناسبی بر روی ویدئوهای بازی ندارند. به همین جهت، دو مدل ادراکی مخصوص کاربرد بازی در این رساله ارائه می‌شود. مدل اول یک مدل توجه بصری است که هر دو سازوکار پایین-به-بالا و بالا-به-پایین توجه را در نظر می‌گیرد. برای توجه پایین-به-بالا از ویژگی‌های سطح پایین سیگنال و برای توجه بالا-به-پایین از الگوریتم‌های اولویت‌دهی اشیا بازی استفاده می‌شود. مدل دوم نیز به منظور نشان دادن تاثیر مهارت بازیکن بر روی الگوی توجه وی ارائه شده است.

بازیکن هنگام بازی، بر حسب نوع بازی، مهارت خود و هدفی که در ذهن دارد، به قسمت‌های خاصی از تصویر توجه می‌کند. بنابراین در یک لحظه مشخص از بازی، بازیکن به بخش‌های از تصویر که در موفقیت وی اهمیت دارند، توجه بیشتری داشته و در مشاهده سایر بخش‌ها کم دقت است. این مشاهده بخصوص در مورد بازی‌هایی که سرعت بالایی دارند و موفقیت بازیکن در گرو عکس العمل سریع وی است، صحیح می‌باشد [۷]. علاوه بر این، چشم انسان بطور ذاتی به بخش‌هایی از تصویر که ویژگی‌های خاصی (مانند تغییر شدت نور بالا) داشته باشند، معطوف می‌شود. این قبیل امور به ما کمک می‌کنند تا اهمیت بخش‌های مختلف تصویر برای بازیکن را تخمین بزنیم و بخش‌های با اهمیت را با کیفیت بهتر (نرخ بیت بیشتر) و بخش‌های کم اهمیت را با کیفیت کمتر (نرخ بیت کمتر) ارسال کنیم. تخصیص نرخ بیت به بخش‌های مختلف تصویر بگونه‌ای انجام می‌شود که نرخ بیت کل ویدئو با پهنای باند در دسترس بازیکن مطابقت داشته باشد.

در شکل ۳-۱ دقت پایین بهترین مدل‌های توجه بصری تاکنون ارائه شده با توجه به محک Judd نشان داده شده است. در این شکل، هر نقطه قرمز، محل نگاه یک بازیکن نمونه به آن فریم در حال بازی را نشان می‌دهد. توجه داشته باشید که فریم‌های بازی بر اساس ورودی‌های دریافتی از کاربر به تصویر کشیده می‌شوند. بنابراین فریم‌ها با یکدیگر متفاوت هستند و امکان نمایش نقطه خیرگی بیش از یک بازیکن برای هر فریم وجود ندارد. به عبارت دیگر هر بازیکن فریم‌های یکتایی دارد. در بخش‌های بعدی همین فصل دقت پایین سایر مدل‌ها را بصورت عددی نشان خواهیم داد.

^۱ Perceptual video coding



شکل ۳-۱ - اعمال مدل‌های مختلف بر روی صحنه‌هایی از بازی‌ها و نمایش عدم تطبیق با خیرگی چشم

۳-۱- مدل توجه بازی

بمنظور تشخیص اهمیت نواحی مختلف تصویر برای بازیکن یک مدل توجه ترکیبی ارائه شده است. این مدل در واقع دو نوع مختلف توجه در انسان را شبیه‌سازی و با یکدیگر ترکیب می‌کند. ورودی این مدل یک فریم از بازی و خروجی آن یک نگاشت سطوح خاکستری^۱ می‌باشد. درجه سفیدی یک ناحیه در این نگاشت میزان اهمیت آن ناحیه از دید بازیکن را نشان می‌دهد. این نگاشت می‌تواند برای تعیین مناسب پارامترهای کدکننده به منظور کنترل نرخ بیت ویدئو بازی مورد استفاده قرار گیرد. به عنوان مثال می‌توان مقدار QP برای ماکرو بلاک‌هایی که در بخش‌های کم اهمیت قرار دارند را مقداری افزایش داد بگونه‌ای که نرخ بیت آنها کاهش یافته و در عین حال کاهش کیفیت آن نواحی برای بازیکن نامحسوس باشد.

برای طراحی این مدل عوامل مختلفی باید در نظر گرفته شوند. بی تردید، مهمترین عامل منطق بازی است. بعید است بدون در نظر گرفتن نوع بازی، ماموریت‌ها و شخصیت‌های آن بتوان یک مدل کامل ارائه داد. به عنوان مثال، در یک بازی شلیکگر اول شخص، دشمن‌هایی که بازیکن در جنگ با آنها درگیر شده است نسبت به سایر واحدهایی که در صحنه دیده می‌شوند، اهمیت بیشتری دارند. عامل دیگر در طراحی مدل، ویژگی‌های ذاتی چشم انسان و همچنین ساختار توجه در انسان است. در انسان دو نوع توجه وجود دارد: توجه پایین-به-بالا^۲ یا تحریک‌محور^۳ و توجه بالا-به-پایین^۴ و وظیفه‌محور^۵. توجه پایین-به-بالا در خود چشم و مسیر انتقال اطلاعات از چشم به مغز انجام می‌شود. در واقع چشم انسان قبل از ارسال تمام اطلاعات دریافت شده توسط سنسورهای موجود در چشم به مغز، قسمتی از آن اطلاعات را غربال می‌کند. به عنوان مثال، قسمت‌هایی از تصویر که تغییر روشنائی در آنها زیاد باشد، به عنوان اطلاعات مهم به مغز فرستاده می‌شود و توجه انسان را به خود جلب می‌کند. نوع دیگر توجه، توجه بالا-به-پایین است که بر اساس دستوری است که از مغز به چشم می‌رسد. در این حالت، مغز انسان بر اساس وظیفه‌ای که در دست اجرا دارد، از چشم تقاضا می‌کند روی قسمت مشخصی از صحنه تمرکز کرده و اطلاعات آن قسمت را برایش ارسال کند.

مدل‌های اولیه توجه فقط نوع پایین-به-بالا را در نظر می‌گرفتند. اما به تدریج با شکل‌گیری دیتاست‌های مبتنی بر داده‌های واقعی و مشاهده اختلاف نتایج، دانشمندان با اضافه کردن توجه نوع بالا-به-پایین دقت این مدل‌ها را افزایش دادند. به عنوان مثال، پیمان‌هایی مانند تشخیص صورت انسان، ماشین و متن به این مدل‌ها اضافه گردید. مدل ارائه شده در این رساله نیز هر دو مدل توجه را در نظر می‌گیرد. شکل ۲-۳ پیمان‌های تشکیل دهنده این مدل را نشان می‌دهد. توجه داشته باشید که

^۱ Grayscale map

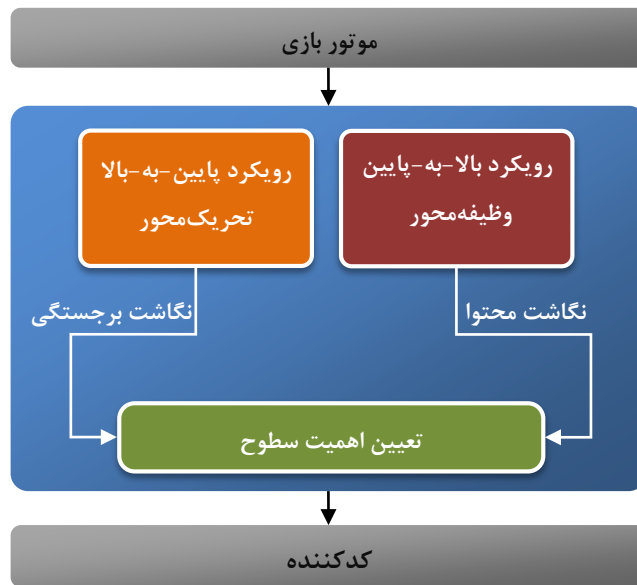
^۲ Bottom-up attention

^۳ Stimulus-driven attention

^۴ Top-down attention

^۵ Task-driven attention

خروجی این مدل‌ها لزوماً با هم هم‌پوشانی ندارند و تکیه بر فقط یکی از آنها باعث می‌شود بعضی نواحی مهم در تصویر شناسایی نشده و کیفیت تجربه بازیکن پایین بیاید.



شکل ۳-۲ - مدل توجه بازی ارائه شده

در شکل ۳-۲، سه پیمانانه^۱ دیده می‌شود. در پیمانانه پایین-به-بالا معمولاً ویژگی‌های سطح پایین بصری مانند شدت نور، رنگ، زاویه، بافت و حرکت، در مقیاس‌های مختلف، از تصویر استخراج می‌شود. سپس یک نگاشت اهمیت برای هر یک از این ویژگی‌ها محاسبه شده و نرمال‌سازی می‌گردد. در نهایت این نگاشت‌ها بصورت خطی یا غیرخطی با هم ترکیب می‌شوند تا یک نگاشت اهمیت نهایی بدست آید. این نگاشت اهمیت هر پیکسل را نشان می‌دهد. در پیمانانه بالا-به-پایین سعی می‌شود با در نظر گرفتن منطق بازی و حالت فعلی، اشیا مهم از دید بازیکن شناسایی شده و یک نگاشت اهمیت دیگر ساخته شود. به عنوان مثال، اگر سوخت ماشین بازیکن در حال تمام شدن باشد، این بازیکن با احتمال بالایی به علائم نصب شده توجه می‌کند تا پمپ بنزین را پیدا کند. با این منطق، در نگاشت خروجی پیمانانه توجه بالا-به-پایین پیکسل‌های متناظر با علائم پمپ بنزین با درجه سفیدتری نشان داده می‌شوند. پیمانانه سومی که در این شکل قابل مشاهده است، پیمانانه تعیین سطح اهمیت نام دارد. وظیفه این پیمانانه ترکیب نگاشت‌های اهمیت دو پیمانانه قبلی و ارائه یک نگاشت اهمیت نهایی می‌باشد. همانطور که مشاهده می‌شود، مدل ارائه شده پیمانانه‌ای می‌باشد و برای هر یک از پیمانانه‌های آن می‌توان الگوریتم مستقلی در نظر گرفت. به عنوان مثال، پیمانانه تعیین سطح اهمیت می‌تواند به سادگی یک عملگر بیشینه و یا به پیچیدگی یک مدل هوشمند یادگیری ماشین پیاده‌سازی شود.

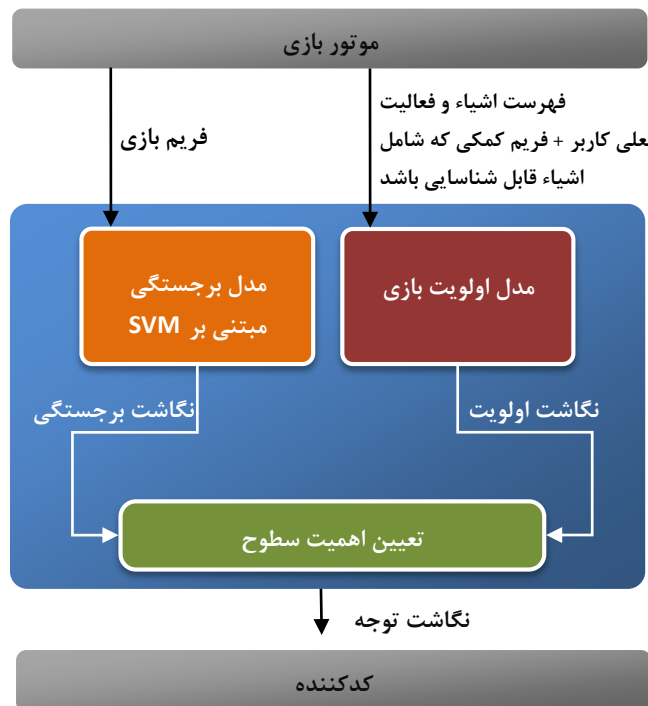
^۱ Module

وقتی نگاهت اهمیت نهایی توسط مدل توجه بازی ساخته شد، این نگاهت در اختیار کدکننده قرار داده می‌شود تا بر اساس درجه اهمیت پیکسل‌های موجود در تصویر پارامترهای خود را تنظیم کند. تنظیم پارامترها بگونه‌ای خواهد بود که نواحی با اهمیت با کیفیت بیشتر (نرخ بیت بالا) و نواحی کم اهمیت با کیفیت کمتری (نرخ بیت پایین) کد شوند. در این فرآیند کیفیت تجربه در نظر گرفته شده و اعوجاج‌ها بگونه‌ای اعمال می‌شوند که تا جای ممکن از دید بازیکن پنهان بمانند.

۳-۱-۱- پیاده‌سازی

هدف این رساله اثبات تاثیر مثبت مدل ارائه شده در کاهش و کنترل نرخ بیت ویدئو بازی می‌باشد. بنابراین در پیاده‌سازی یک نمونه از مدل، برای هر یک از پیمان‌های شکل ۳-۲ یک نمونه از الگوریتم‌های ارائه شده اخیر در نظر گرفته شده است. اگرچه این انتخاب ممکن است بهینه نباشد، اما برای رسیدن به اهداف این رساله کافی است.

در ادامه به تشریح پیاده‌سازی انجام شده برای مدل توجه بازی و استفاده از آن در کنار کدکننده H.264/AVC می‌پردازیم. انتخاب ما برای پیمان‌ها پایین-به-بالا، مدل مبتنی بر ماشین بردار پشتیبانی است که توسط خانم Judd ارائه شد. این مدل بر اساس یک دیتاست بزرگ ردیابی چشم توسعه داده شده است [۴۲]. برای پیمان‌ها بالا-به-پایین نیز مدل اولویت‌بندی اشیاء بازی ارائه شده در [۶۹] مورد استفاده قرار گرفته است. در مورد پیمان‌ها تعیین سطح اهمیت نیز از یک عملگر بیشینه استفاده شده است.

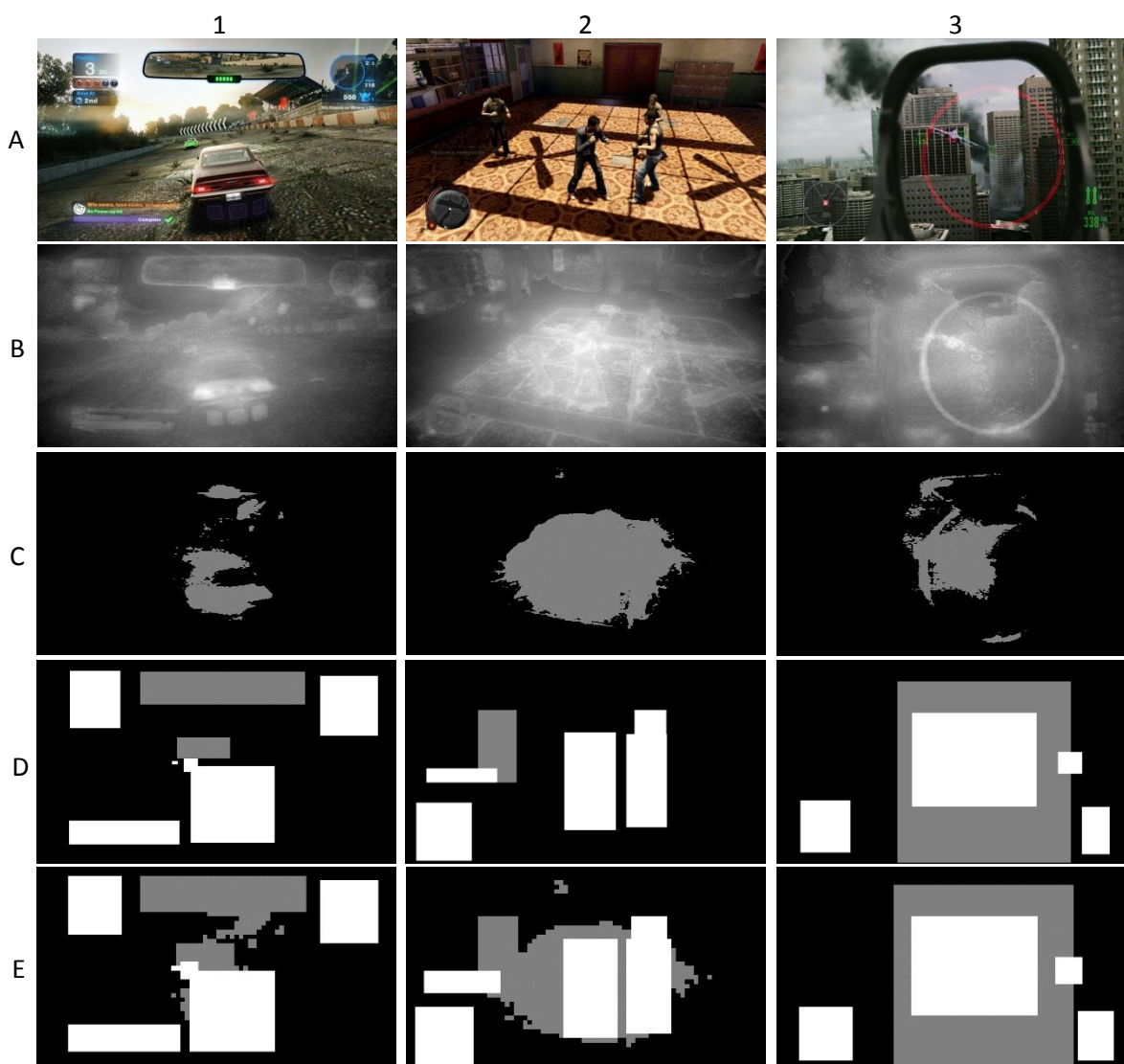


شکل ۳-۳ - یک نمونه پیاده‌سازی مدل توجه بازی

شکل ۳-۳ چگونگی ساخت یک نگاشت چند سطحی توسط مدل توجه بازی را نشان می‌دهد. هنگامی که موتور بازی^۱ مشغول به تصویر کشیدن فریم فعلی است، یک لیست از اشیا موجود در فریم فعلی و فعالیت فعلی بازیکن در اختیار مدل توجه بازی قرار می‌دهد. موتور بازی همچنین، به طور موازی، به تصویر کشیدن یک فریم اضافه را آغاز می‌کند. این فریم شامل دقیقاً همان اشیا موجود در فریم فعلی است، البته بدون هیچ جلوه ویژه‌ی گرافیکی. موتور بازی به هر شی‌ای در این فریم یک رنگ یکتا اختصاص می‌دهد تا تصویر دو بعدی آن شی بر روی فریم فعلی برای دستکاری‌های بعدی قابل تشخیص باشد. توجه داشته باشید که این به تصویر کشیدن اضافی شامل هیچ پردازش بافت، نور یا سایه‌ای نمی‌باشد و می‌تواند حتی قبل از به تصویر کشیدن فریم اصلی اتمام پذیرد. برای کاهش هزینه پردازشی می‌توان یک گام دیگر نیز پیش گذاشت و بجای خود اشیا از جعبه محیطی آنها استفاده کرد. پیدا کردن جعبه محیطی از امکانات پایه کتابخانه‌های گرافیکی محسوب می‌شود. بطور موازی، پیمانانه بالا-به-پایین بر اساس روش اولیت‌بندی ارائه شده در [۶۹] به هر یک از اشیا بر اساس فعالیت فعلی بازیکن یک ضریب اهمیت اختصاص می‌دهد. این ضریب جنبه‌های مختلف بازی از جمله غوطه‌وری، کیفیت بصری و وظایف بازیکن را در نظر می‌گیرد. هنگامیکه به تصویر کشیدن فریم اضافی به پایان رسید، این فریم در اختیار پیمانانه بالا-به-پایین قرار می‌گیرد. در پیمانانه بالا-به-پایین، به هر یک از اشیا، که اکنون با رنگ یکتای خود قابل تشخیص هستند، بر اساس اولویت‌شان یک درجه خاکستری نسبت داده می‌شود. در این رساله، ما تنها سه سطح اولویت پایین، متوسط و بالا را در نظر گرفتیم. بعد از آنکه موتور بازی فریم اصلی را به تصویر کشید، آن را در اختیار پیمانانه پایین-به-بالا قرار می‌دهد. این پیمانانه با استفاده از مدل Judd نگاشت اهمیت فریم اصلی را محاسبه می‌کند. سپس، این نگاشت با grayscale با استفاده از یک آستانه به یک نگاشت سیاه و سفید تبدیل می‌شود. اگر این آستانه خیلی کوچک در نظر گرفته شود، برخی نواحی مهم از دست می‌روند. اگر هم خیلی بزرگ انتخاب شود، نواحی کم اهمیت نیز در نگاشت گنجانده می‌شوند. در هر دو حالت عملکرد مدل را پایین می‌آید. ما در کارمان به صورت آزمون و خطا مقدار این آستانه را طوری تعیین کردیم که نواحی با اهمیت با درصد خوبی در نگاشت قرار گیرند. این مقدار آستانه می‌تواند برای کل ویدئو مورد استفاده قرار گیرد. در واقع لزومی به انتخاب آن برای تک‌تک فریم‌ها وجود ندارد. در نهایت نگاشت‌های خروجی دو پیمانانه توجه پایین-به-بالا و بالا-به-پایین به پیمانانه تعیین سطح اهمیت فرستاده می‌شوند و در آنجا توسط عملگر بیشینه با هم ترکیب می‌گردند. بطور دقیق، ابتدا یک عملیات بیشینه‌گیری پیکسل-به-پیکسل بر روی این دو نگاشت انجام می‌شود. سپس نتیجه به بلوک‌های ۱۶ در ۱۶ تقسیم‌بندی شده و به هر کدام از این بلوک‌ها یک مقدار معادل بیشینه ۲۵۶ پیکسل موجود در آن نسبت داده می‌شود. این کار از آن جهت لازم است که ما در این پایان‌نامه از کدکننده H.264/AVC استفاده کردیم و واحد پردازشی در این کدکننده ماکروبلوک می‌باشد. بنابراین پیمانانه تعیین سطح اهمیت را بگونه‌ای پیاده سازی کردیم که به ازای هر ماکروبلوک یک سطح اهمیت مشخص کند. توجه داشته باشید که در

^۱ Game engine

صورت استفاده از عملگر میانگین یک ماکرو بلاک با تعداد نسبتاً کمی پیکسل با اهمیت ممکن است به دلیل میانگین‌گیری به عنوان یک ماکرو بلاک کم اهمیت در نظر گرفته شود. برای تعمیم‌پذیری بیشتر مدل ارائه شده، دو وزن در نظر گرفتیم تا با استفاده از آنها بتوان میزان تاثیر هر یک از پیمانه‌های توجه بالا-به-پایین و پایین-به-بالا در نگاشت نهایی را کنترل کرد. ما با آزمون و خطا دریافتیم که اگر وزن نگاشت بالا-به-پایین دو برابر وزن نگاشت پایین-به-بالا انتخاب شود، نتایج مدل به واقعیت نزدیک‌تر خواهند بود.



شکل ۳-۴ - نتایج پیمانه‌های مدل توجه بازی برای سه نمونه فریم. (A فریم، B نگاشت پایین-به-بالا، C نگاشت پایین-به-بالا بعد از اعمال آستانه، D نگاشت بالا-به-پایین و E نگاشت نهایی).

وقتی درجه اهمیت هر ماکروبلوک مشخص شد، می‌توان از آن برای انتخاب مناسب پارامترهای کدکننده استفاده کرد بگونه‌ای که اختصاص بیت به مقدار موردنظر نزدیک گردد. در این رساله، بدون وارد شدن خدشه به تعمیم‌پذیری، ما پارامتر QP را در نظر گرفتیم. بنابراین، مقدار QP هر ماکروبلوک بر اساس درجه اهمیت آن ماکروبلوک تنظیم می‌شود. با توجه به آنکه ما در آزمایش خود از سه سطح اولویت استفاده کردیم، سه مقدار QP نیز در نظر گرفتیم. به ماکروبلوک‌های با اهمیت بیشتر، مقدار QP کوچکتری اختصاص داده می‌شود تا این ماکروبلوک‌ها با کیفیت بالاتری کد شوند. شکل ۳-۴ نگاشت خروجی هر یک از پیمان‌ها و همچنین نگاشت نهایی برای سه فریم نمونه را نشان می‌دهد.

۳-۱-۲- ارزیابی مدل توجه بازی

افزایش مقدار QP در بعضی ماکروبلوک‌ها بطور مشخص مقدار PSNR و سایر معیارهای کمی مانند SSIM را کاهش می‌دهد. اما همانطور که قبلاً گفته شد، انتظار داریم با استفاده از مدل توجه بازی و تخمین صحیح درجه اهمیت ماکروبلوک‌ها، این تغییرات از دید بازیکن پنهان بماند. برای تایید این فرضیه از ارزیابی کیفی^۱ استفاده می‌کنیم. عملکرد مدل توجه بازی در دو گام بررسی می‌شود. در یک گام مدل توجه بازی را بر روی تک فریم‌هایی (فریم-I) از بازی‌های مختلف و در گام بعدی آن را بر روی ویدئوهای (فریم-I و فریم-P) ضبط شده از تعدادی بازی اعمال می‌کنیم.

۳-۱-۲-۱- ارزیابی مدل بر تک فریم

برای مشخص شدن تاثیر مدل توجه بازی، ارزیابی اولیه بر روی تعدادی فریم از بازی‌های مختلف انجام شد. بطور دقیقتر، آزمایش اولیه بر روی هجده فریم انجام شد. سیزده فریم از این مجموعه از بازی‌های Call of Duty و Far Cry که از نوع شلیکگر اول شخص هستند، انتخاب شد. پنج فریم دیگر از بازی‌های Need For Speed، Moto Racer و Chicken Invaders انتخاب گردید. از آنجا که شرایط این آزمایش با آزمایش ویدئو یکسان است و نتایج آزمایش ویدئو به شرایط واقعی نزدیک‌ترند، از توضیح بیشتر در این زمینه خودداری می‌کنیم و در بخش بعد وارد جزئیات می‌شویم.

۳-۲-۱-۲- ارزیابی مدل بر روی ویدئو

در این مرحله نه ویدئو از هشت بازی معروف اخیر ضبط کردیم. در ضبط این ویدئوها دقت شد تا آنها به مقدار کافی شامل جزئیاتی مانند اشیاء محیطی، عوامل دشمن و نیروهای متحد باشند. این بازی‌ها از لحاظ معیارهای گوناگون از جمله نوع بازی، سرعت حرکت و نور، متنوع انتخاب شده‌اند. جدول ۳-۱ اطلاعات مربوط به این بازی‌ها را نشان می‌دهد.

^۱ Subjective test

جدول ۳-۱- اطلاعات بازی‌های انتخاب شده برای ارزیابی مدل توجه بازی

انتشار	سبک	مخفف	نام بازی
۲۰۱۲	شلیکگر اول شخص	COD	CALL OF DUTY BLACK OPS II
۲۰۱۳	شلیکگر جنگ فنی	SGW	SNIPER GHOST WARRIOR 2
۲۰۱۲	ماجراجویی	SLD	SLEEPING DOGS
۲۰۱۰	مسابقات خودرو	BLR	BLUR
۲۰۱۱	مسابقات خودرو	NFS	NEED FOR SPEED THE RUN
۲۰۱۲	وزرشی	PES	PRO EVOLUTION SOCCER 2013
۲۰۱۲	شلیکگر جنگ فنی، مخفیانه	SNE1/SNE2	SNIPER ELITE V2
۲۰۱۱	شبیه‌ساز جنگ پروازی	ACA	ACE COMBAT: ASSAULT HORIZON

در ویدئوی بازی COD، شخصیت اصلی در یک اتاق با یک دشمن در حال نبرد است. در مقابل او سه درب وجود دارد که ممکن است دشمن از آنها وارد اتاق شود. در داخل اتاق چند میز و جعبه چوبی نیز دیده می‌شود. این ویدئو کم سرعت، کم محتوا و کم نور است.

در ویدئوی بازی SGW، در یک جنگل پردرخت قرار دارد و مشغول کشتن نگهبانان یک پل می‌باشد. درخت‌ها، بوته‌ها، سنگ‌ها و علف از اشیایی هستند که در صحنه دیده می‌شوند. این ویدئو نسبتاً پر سرعت، متوسط محتوا و کم نور است.

در ویدئوی بازی SLD، شخصیت اصلی در یک اتاق قرار دارد. دکوراسیون اتاق به سبک چینی است. در اتاق دو دشمن وجود دارد و بازیکن مشغول مبارزه با آنها می‌باشد. یک دوست از بازیکن نیز در صحنه دیده می‌شود. این ویدئو پر سرعت، پر محتوا و پر نور است.

در ویدئوی بازی BLR، بازیکن در یک مسابقه ماشین شرکت کرده است. غیر از ماشین خود بازیکن، دو ماشین دیگر در صحنه دیده می‌شود. تعدادی آیتم در پیست مسابقه وجود دارد که بازیکن می‌تواند با رد شدن از رویشان، آنها را بدست آورد و از طریق آنها به سایر ماشین‌ها حمله و یا خود را در برابر حمله آنها محافظت کند. این ویدئو پر سرعت، پر محتوا و نسبتاً پر نور است.

در ویدئوی بازی NFS، شخصیت اصلی سوار بر ماشین خود در حال فرار است و دو ماشین در تعقیب وی هستند. این تعقیب و گریز در خیابان‌های داخل یک شهر انجام می‌گیرد. علائم چشمک زنی در کنار خیابانها قرار دارند تا بازیکن بتواند برای عبور از دوره‌های تند در زمان مناسب عکس العمل نشان دهد. قسمتی از نقشه شهر در گوشه تصویر دیده می‌شود. این ویدئو پر سرعت، متوسط محتوا و کم نور است.

در ویدئوی بازی PES، بازیکن پشت محوطه جریمه حریف قرار دارد؛ از آنجا به فرورارد پاس می‌دهد و فرورارد توپ را به سمت دروازه شوت می‌کند. در این حین، مهره‌های حریف سعی می‌کنند توپ را باز پس گیرند. این ویدئو نسبتاً پر سرعت، پر محتوا و روشن است.

از بازی SNE دو ویدئو ضبط کردیم. در ویدئوی SNE1 شخصیت اصلی در نقش یک تک تیرانداز عمل می‌کند. وی در ارتفاع قرار دارد و اسلحه خود را یکی از دو دشمنی در دوردست مشغول صحبت با یکدیگر هستند، زوم کرده است. نواحی مرکزی این ویدئو کم سرعت، پر محتوا و روشن است، در حالیکه نواحی کناری آن کم سرعت، کم محتوا و تیره می‌باشد.

در ویدئوی SNE2، در پشت یک مجسمه پناه گرفته است و سعی می‌کند سربازی که در مقابلش قرار گرفته است را از پای درآورد. آنها در حیاط یک ساختمان هستند و یک ماشین در کنار آنها در حال سوختن است. این ویدئو متوسط سرعت، پر محتوا و روشن است.

در نهایت، در ویدئوی بازی ACA، بازیکن سعی می‌کند هواپیمای دشمن را منهدم کند. آنها بر فراز یک شهر با ساختمان‌های بلند پرواز می‌کنند. اطلاعات پروازی نیز بر روی صفحه دیده می‌شود. این ویدئو پر سرعت، پر محتوا و روشن است.

ویدئوهای انتخاب شده بازه وسیعی از انواع بازی‌ها را می‌پوشانند و از لحاظ سرعت، محتوا و نور متنوع هستند. بنابراین نتایج ارزیابی، پتانسیل مدل توجه بازی برای بازه وسیعی از بازی‌ها را نشان می‌دهد. شکل ۳-۵ اولین فریم این ویدئوها را نشان می‌دهد. همه این ویدئوها با سرعت ۳۰ فریم بر ثانیه ضبط و با یک کدکننده بدون اتلاف^۱ ذخیره شدند. هر ویدئو ۴ ثانیه طول دارد و شامل ۱۲۰ فریم می‌باشد.

برای کدکردن ویدئوها، از نرم افزار JM V18.4 استفاده کردیم. این نرم افزار یک پیاده‌سازی از استاندارد H.264/AVC است. اکثر شرکت‌های فعال در زمینه CG از همین استاندارد برای کدکردن ویدئوی بازی استفاده می‌کنند. برای آنکه ماکرو بلاک‌هایی که درجه اهمیت برابری با یکدیگر دارند، در یک گروه قرار گیرند از ابزار Flexible Macro-block Ordering (FMO) استفاده کردیم. مشابه شرکت Onlive و سایر شرکت‌های عرضه‌کننده CG که با کیفیت HD بازی ارائه می‌کنند، ما دقت مکانی 720p را برای انجام آزمایش‌ها انتخاب کردیم.

^۱ Lossless

COD



SGW



SLD



BLR



NFS



PES



SNE1



SNE2



ACA



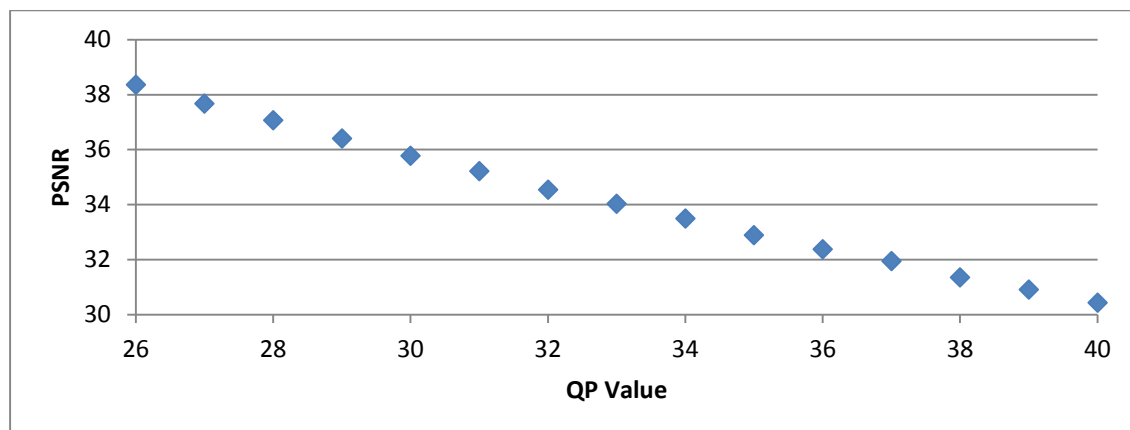
شکل ۳-۵ - اولین فریم هر یک از نه ویدئوی بازی‌های استفاده شده در ارزیابی

در این آزمایش هر یک از ویدئوها را یکبار با استفاده از مدل توجه بازی و یکبار بدون استفاده از آن کد کردیم و نرخ بیت آنها در این دو حالت را با یکدیگر مقایسه کردیم. سپس ویدئوهای کدشده را واکنش کردیم و کیفیت آنها را بر اساس معیارهای کمی^۱ و کیفی مقایسه کردیم. در همه موارد اندازه GOP را ۱۵ در نظر گرفتیم (یک فریم I در ابتدا به همراه ۱۴ فریم P پشت سر آن). هنگام استفاده از مدل توجه بازی، برای تشخیص ROI تنها از فریم ابتدایی هر GOP استفاده گردید و آن ROI برای کل فریم‌های GOP ثابت در نظر گرفته شد. جدول ۲-۳ پارامترهایی که در این آزمایش برای کدکننده در نظر گرفتیم را نشان می‌دهد.

جدول ۲-۳- نام و مقدار پارامترهای کدکننده H.264/AVC استفاده شده در ارزیابی

مقدار	نام
Baseline	پروفایل
۳,۱	سطح
۱	تعداد فریم‌های مرجع
EPZS	شمای تخمین حرکت
۳۲	دامنه جستجو
روشن	تشخیص مد نرخ-اعوجاج
خاموش	کنترل نرخ

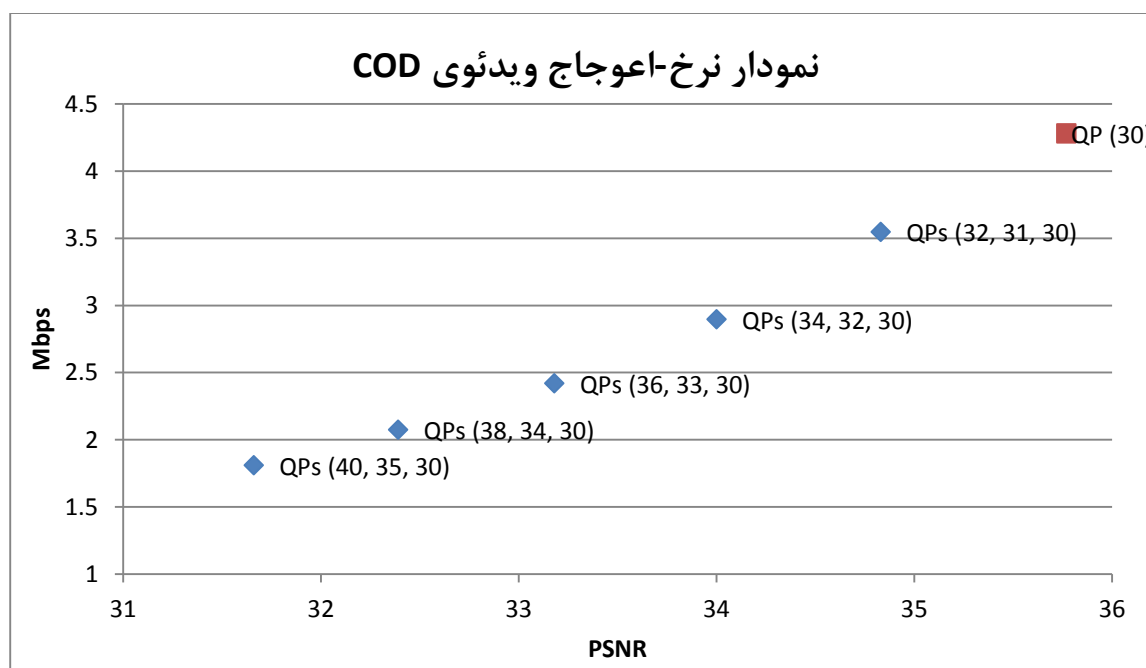
ویدئوهای مرجع در این آزمایش بدون استفاده از مدل توجه بازی کد شدند. بنابراین برای آنها تنها یک مقدار QP در نظر گرفته شد. برای رعایت انصاف در مقایسه، بیشترین مقدار QP که به ازای آن، کسی قادر به تشخیص اعوجاج در ویدئوها نبود، انتخاب گردید. این مقدار بیشینه برای QP، ۳۰ بود و به ازای آن متوسط مقدار PSNR برای کل ویدئوها، ۳۶ dB بدست آمد. شکل ۳-۶ نشان می‌دهد چگونه با افزایش مقدار QP مقدار PSNR برای ویدئوی COD کاهش می‌یابد.



^۱ Objective

شکل ۳-۶ - مقدار PSNR ویدئو COD به ازای مقادیر مختلف QP

در گام بعدی ویدئوهای مرجع را با استفاده از مدل توجه بازی کد کردیم. در این حالت با توجه به اینکه سه سطح اهمیت در نظر گرفته بودیم، سه مقدار QP تعیین کردیم. برای تعیین مقادیر مناسب این QP ها تمام سه‌تایی‌های به فرم $(30 + 2X, 30 + X, 30)$ را بررسی کردیم که در آن $1 \leq X \leq 5$. شکل ۳-۷ نمودار نرخ-اعوجاج این مقادیر برای ویدئوی COD را نشان می‌دهد. در این نمودار، با افزایش مقدار X ، هم نرخ بیت و هم PSNR کاهش می‌یابند. با این وجود، ارزیابی کیفی ما نشان می‌دهد مقداری کاهش در PSNR قابل تحمل است و کیفیت تجربی بازیکنان را کاهش نمی‌دهد. دلیل این امر آن است که مقدار QP برای نواحی با اهمیت با مقدار QP در سناریوی تک-QP یکسان نگاه داشته شده است.



شکل ۳-۷ - نمودار نرخ-اعوجاج ویدئوی COD به ازای سه‌تایی‌های مختلف QP

جدول ۳-۳ مقادیر PSNR، SSIM و نرخ بیت هر نُه ویدئو را برای سناریوهای مختلف آزمایش شده نشان می‌دهد. جدول ۳-۴ درصد کاهش نرخ بیتی که در استفاده از سناریوهای چند-QP نسبت به سناریوی تک-QP بدست می‌آید را نشان می‌دهد. با توجه به این جدول اگرچه با انتخاب سه‌تایی $(40, 35, 30)$ برای مقادیر QP ها، امکان کاهش نرخ بیت تا ۵۰ درصد وجود دارد، ما این سناریو را برای ارزیابی کیفی انتخاب نکردیم. زیرا در این سناریو مقدار اعوجاج در نواحی کم اهمیت به قدری بود که کیفیت تجربه بازیکن را تحت تاثیر قرار دهد. در عوض سناریوی را انتخاب کردیم که اعوجاج

نواحی کم اهمیت دقیقاً در یک سناریو بعد از آن قابل مشاهده باشد^۱. در این سناریو مقدار QP برای نواحی کم اهمیت، اهمیت متوسط و با اهمیت به ترتیب ۳۴، ۳۲ و ۳۰ بود.

جدول ۳-۳- مقادیر PSNR (dB)، SSIM و نرخ بیت (Mbps) برای QP های مختلف

Scenario	COD			SGW			SLD		
	PSNR	SSIM	Bitrate	PSNR	SSIM	Bitrate	PSNR	SSIM	Bitrate
QP (30)	35.77	0.89	4.28	36.4	0.91	3.47	35.02	0.91	6.15
QPs (32, 31, 30)	34.83	0.87	3.54	35.44	0.89	2.67	34.13	0.89	5.28
QPs (34, 32, 30)	34	0.85	2.89	34.58	0.88	2.08	33.33	0.87	4.65
QPs (36, 33, 30)	33.18	0.83	2.41	33.71	0.86	1.66	32.50	0.85	4.21
QPs (38, 34, 30)	32.39	0.82	2.07	32.88	0.85	1.40	31.67	0.83	3.92
QPs (40, 35, 30)	31.66	0.8	1.80	32.11	0.83	1.22	30.91	0.81	3.71
Scenario	BLR			NFS			PES		
	PSNR	SSIM	Bitrate	PSNR	SSIM	Bitrate	PSNR	SSIM	Bitrate
QP (30)	36.76	0.93	12.81	39.89	0.97	4.52	39.44	0.94	1.05
QPs (32, 31, 30)	35.85	0.92	11.16	39.21	0.96	4.04	38.8	0.93	0.99
QPs (34, 32, 30)	34.96	0.91	9.77	38.56	0.96	3.64	38.16	0.93	0.87
QPs (36, 33, 30)	33.93	0.9	8.53	37.87	0.95	3.32	37.47	0.93	0.77
QPs (38, 34, 30)	32.91	0.89	7.57	37.07	0.95	3.07	36.68	0.92	0.70
QPs (40, 35, 30)	31.92	0.88	6.80	36.37	0.94	2.87	35.86	0.92	0.64
Scenario	SNE1			SNE2			ACA		
	PSNR	SSIM	Bitrate	PSNR	SSIM	Bitrate	PSNR	SSIM	Bitrate
QP (30)	39.85	0.97	1.77	36.8	0.94	5.37	39.02	0.97	6.23
QPs (32, 31, 30)	39.05	0.97	1.51	35.91	0.93	4.58	38.26	0.97	5.72
QPs (34, 32, 30)	38.27	0.96	1.25	35.02	0.92	3.77	37.47	0.96	5.17
QPs (36, 33, 30)	37.52	0.96	1.06	34.14	0.91	3.19	36.65	0.96	4.74
QPs (38, 34, 30)	36.69	0.95	0.91	33.24	0.90	2.73	35.72	0.95	4.37
QPs (40, 35, 30)	35.91	0.95	0.77	32.34	0.88	2.32	34.77	0.94	4.04

جدول ۳-۴- درصد کاهش نرخ بیت برای سناریوهای چند-QP

	COD	SGW	SLD	BLR	NFS	PES	SNE1	SNE2	ACA	متوسط
QPs (32, 31, 30)	17.10	22.77	14.07	12.89	10.52	6.11	14.45	14.70	8.14	13.42%
QPs (34, 32, 30)	32.34	39.83	24.34	23.71	19.32	16.93	29.35	29.66	17.08	25.84%
QPs (36, 33, 30)	43.46	51.9	31.60	33.44	26.41	26.3	40.10	40.56	23.94	35.30%
QPs (38, 34, 30)	51.54	59.49	36.24	40.87	31.91	33.31	48.54	49.15	29.89	42.33%
QPs (40, 35, 30)	57.72	64.59	39.65	46.92	36.48	38.91	56.21	56.67	35.11	48.03%

^۱ Just-Noticeable Difference (JND)

برای بررسی تاثیر هر یک از پیمان‌های مدل به تنهایی، ویدئوها را یکبار فقط با استفاده از نگاشت توجه پایین-بالا و یکبار فقط با استفاده از نگاشت بالا-پایین-کد کردیم. در این آزمایش‌ها از همان سه مقدار QP مذکور استفاده گردید. البته از آنجا که مقدار اعوجاج در سناریوی فقط پایین-بالا قابل چشم پوشی نبود و همچنین محدودیت زمانی که در استاندارد استفاده شده وجود داشت، این سناریو در ارزیابی کیفی گنجانده نشد.

برای ارزیابی کیفی از روش Double Stimulus Continuous Quality Scale تشریح شده در توصیه‌نامه ITU-R Recommendation 500 (ITU-R 1974-1997) استفاده گردید. سی فرد در این ارزیابی شرکت کردند. همه آنها در زمینه پردازش ویدئو و ارزیابی کیفیت تصویر غیر حرفه‌ای بودند. جدول ۳-۵ اطلاعات دموگرافیک شرکت‌کنندگان را نشان می‌دهد.

جدول ۳-۵ - اطلاعات دموگرافیک شرکت‌کنندگان

مهارت در بازی‌های رایانه‌ای از نظر کاربر				
عالی	خوب	متوسط	ضعیف	خیلی ضعیف
٪۱۰	٪۱۵	٪۳۵	٪۱۵	٪۲۵
میزان ساعت بازی در ماه				
بیش از ۳۰	۲۱-۳۰	۱۱-۲۰	۶-۱۰	کمتر از ۵
٪۵	٪۱۰	٪۳۰	٪۲۰	٪۳۵
سکویی که تاکنون کاربر بر روی آن بازی کرده است				
گوشی همراه	تبلت	کنسول	کامپیوتر	
٪۸۵	٪۵	٪۳۵	٪۸۰	
سبک بازی که کاربر تا کنون بازی کرده است				
ورزشی	ماجراجویی	شلیکگر اول شخص		
٪۸۰	٪۳۵	٪۸۵		
شبیه‌ساز پرواز	جنگی	مسابقه‌ای		
٪۱۰	٪۲۰	٪۹۵		

هر یک از ویدئوهای موجود در ارزیابی توسط هر یک از شرکت‌کنندگان دیده شد. این روال حدود بیست دقیقه از وقت آنها را گرفت. ویدئوها با دقت مکانی اصلی خود نشان داده شدند تا هیچ اعوجاجی ناشی از تغییر مقیاس وجود نداشته باشد. فاصله شرکت‌کنندگان تا صفحه نمایشگر طبق توصیه Rec ITU-R 812، چهار برابر ارتفاع نمایشگر تنظیم شد. توجه داشته باشید که مدل توجه بازی به این فاصله وابسته نیست. این فاصله تنها یکی از شرایط انجام آزمون کیفی است. ویدئوهای تک-QP و چند-QP با ترتیب تصادفی و سه ثانیه صفحه خاکستری میانشان به شرکت‌کنندگان نشان داده شد و آنها کیفیت این ویدئوها را با مشخص کردن عددی بین ۱ تا ۵ تعیین کردند. قبل از شروع ارزیابی به آنها گفته شد که ارزیابی آنها می‌تواند یک عدد حقیقی در این بازه باشد. در شروع هر جلسه

ارزیابی، پنج نمایش خارج از برنامه قرار داده شد. اولین نمایش برای آشنایی شرکت‌کننده با روال ارزیابی و چهار نمایش بعدی برای پایداری‌سازی نظر وی. از آنجا که استاندارد به ما اجازه می‌دهد سطوح کیفیت را برای کاربران مشخص کنیم، از تعاریف بکار رفته در [۶] استفاده کردیم. این تعاریف در جدول ۳-۶ گزارش شده‌اند.

جدول ۳-۶- اطلاعات دموگرافیک شرکت‌کنندگان

امتیاز	تعریف
۵-۴/۵	هیچ خرابی در بازی احساس نشد.
۴/۵-۴	مقداری خرابی احساس شد ولی تاثیری در بازی نداشت.
۴-۳	خرابی قابل تشخیص بود که احتمالاً در بازی تاثیر داشت.
۳-۲	خرابی کاملاً احساس می‌شود و بازی را مختل کرده است.
۲-۱	محیط بازی اذیت‌کننده است و بازی قابل انجام نیست.

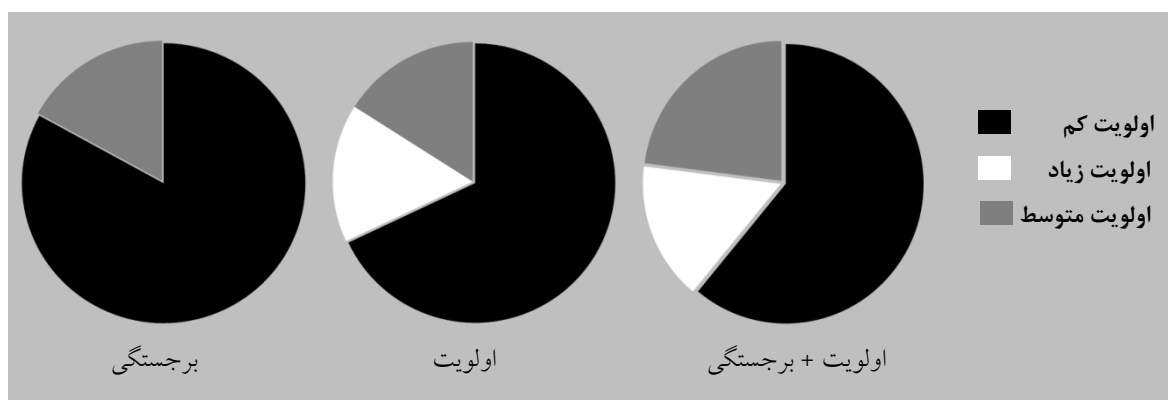
جدول ۳-۷- نتایج آزمایش برای دو سناریو

	متوسط امتیاز نظر									
	COD	SGW	SLD	BLR	NFS	PES	SNE1	SNE2	ACA	Change
تک-QP	5	4.87	5	5	5	4.82	5	4.92	4.97	-
چند-QP (اولویت)	4.47	4.57	4.92	4.8	4.62	3.6	4.45	4.6	4.55	-9.02%
Multi QP (اولویت + برجستگی)	4.44	4.65	5	4.95	4.72	3.8	4.55	4.6	4.7	-7.15%
	نرخ بیت (Mbps)									
	COD	SGW	SLD	BLR	NFS	PES	SNE1	SNE2	ACA	Change
تک-QP	4.28	3.47	6.15	12.81	4.52	1.05	1.77	5.37	6.23	-
چند-QP (اولویت)	2.78	2.00	4.5	9.63	3.53	0.82	1.01	3.57	5.15	-29.62%
Multi QP (اولویت + برجستگی)	2.89	2.08	4.65	9.77	3.64	0.87	1.25	3.77	5.17	-25.84%
	PSNR (dB)									
	COD	SGW	SLD	BLR	NFS	PES	SNE1	SNE2	ACA	Change
تک-QP	35.77	36.4	35.02	36.76	39.89	39.44	39.85	36.8	39.02	-
چند-QP (اولویت)	33.9	34.48	33.19	34.85	38.36	38.13	37.68	34.84	37.46	-4.76%
Multi QP (اولویت + برجستگی)	34	34.58	33.33	34.96	38.56	38.16	38.27	35.02	37.47	-4.33%
	SSIM									
	COD	SGW	SLD	BLR	NFS	PES	SNE1	SNE2	ACA	Change
تک-QP	0.89	0.91	0.91	0.93	0.97	0.94	0.97	0.94	0.97	-
چند-QP (اولویت)	0.85	0.88	0.87	0.91	0.96	0.93	0.96	0.92	0.96	-2.29%
Multi QP (اولویت + برجستگی)	0.85	0.88	0.87	0.91	0.96	0.93	0.96	0.92	0.96	-2.29%

جدول ۳-۷ نتایج آزمایش‌های ما برای دو سناریو را نشان می‌دهد. همانطور که قبلاً نیز بیان شد، در یک سناریو تنها از نگاشت توجه بالا-به-پایین و در سناریوی دیگر از هر دو نگاشت توجه بالا-به-پایین و پایین-به-بالا استفاده شد. در این جدول کیفیت بصری در قالب متوسط امتیاز نظر^۱ بیان شده است. در این جدول همچنین نرخ بیت با واحد KB، PSNR با واحد dB و شاخص SSIM گنجانده شده است. آخرین ستون این جدول درصد کاهش معیارهای کمی و کیفی سناریوهای چند-QP نسبت به سناریوی تک-QP را نشان می‌دهد.

همانطور که از نتایج برمی‌آید، مدل توجه بازی توانسته است بطور میانگین ۲۵٪ نرخ بیت را کاهش دهد. مشخص است که چنین کاهش مقادیر PSNR و SSIM را نیز کم خواهد کرد. اما همانطور که مشاهده می‌شود از دید شرکت‌کنندگان، کیفیت بصری تنها ۷,۱۵٪ کاهش یافته است. توجه داشته باشید که هفت ویدئو از نه ویدئو امتیازی بین ۴,۵ تا ۵ دریافت کرده‌اند. بنابراین قسمت زیادی از عوجاج‌های ایجاد شده از دید آنها پنهان مانده و در کیفیت تجربه‌ی بازی آنها تأثیر نداشته است. توجه داشته باشید که با توجه به تعاریف متناظر با امتیازدهی‌ها، شرکت‌کنندگان در هفت ویدئو از نه ویدئو متوجه خرابی نشده‌اند.

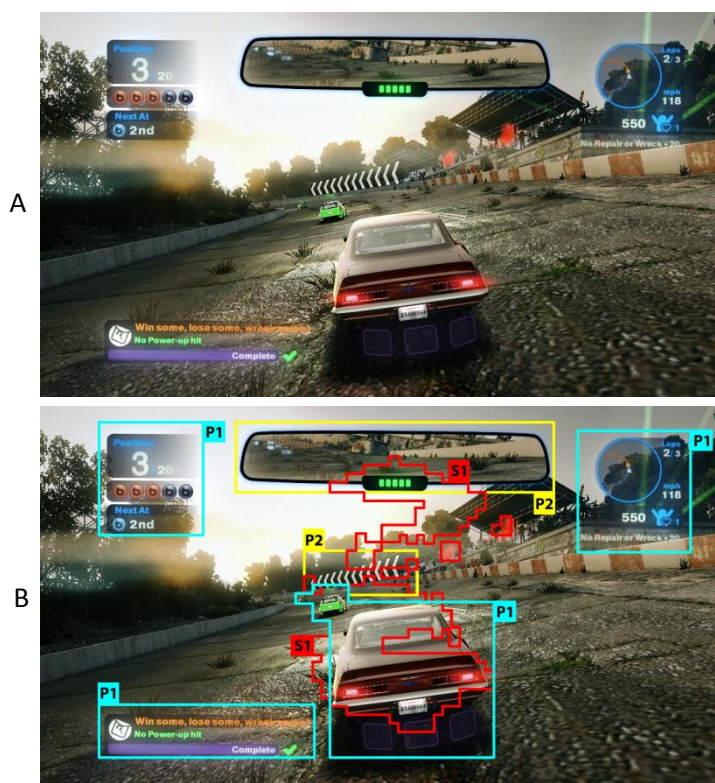
طبق جدول ۳-۷، کیفیت بصری، PSNR و SSIM در سناریوی «هر دو نوع توجه» نسبت به سناریوی «فقط توجه بالا-به-پایین» بهتر است. اگرچه کاهش نرخ بیت در سناریوی «فقط توجه بالا-به-پایین» $25,84 - 29,62 = 3,78\%$ بیشتر بوده است، اما کیفیت سناریوی «هر دو نوع توجه» $1,26 = 9,02/7,15$ برابر آن است. بنابراین نتیجه می‌گیریم که استفاده از هر دو نگاشت توجه در مدل توجه بازی تعادل بهتری بین نرخ بیت و کیفیت تجربه برقرار می‌کند. شکل ۳-۸ درصد متوسط نواحی کم اهمیت، اهمیت متوسط و با اهمیت در نگاشت‌های «فقط توجه بالا-به-پایین» و «هر دو نوع توجه» را نشان می‌دهد.



شکل ۳-۸ - درصد متوسط نواحی کم اهمیت، اهمیت متوسط و با اهمیت در نگاشت‌های مختلف

^۱ Mean Opinion Score (MOS)

شکل ۳-۹ دو نسخه از اولین فریم ویدئوی بازی BLR را نشان می‌دهد. تصویر بالایی در این شکل، فریم مرجع کد شده بدون استفاده از مدل توجه بازی و با مقدار ۳۰ برای پارامتر QP را نشان می‌دهد. تصویر پایینی، همین فریم را که اینبار با استفاده از مدل توجه بازی و سه مقدار QP برای نواحی کم اهمیت (۳۴)، اهمیت متوسط (۳۲) و با اهمیت (۳۰) کد شده است، نشان می‌دهد. در این تصویر، مرزهای قرمز (S1) نواحی با اهمیت در نگاشت توجه پایین-به-بالا را مشخص می‌کنند. مرزهای آبی (P1) و زرد (P2) نیز به ترتیب نواحی با اهمیت و اهمیت متوسط در نگاشت توجه بالا-به-پایین را نشان می‌دهند. نرخ فشرده‌سازی در سایر نواحی بیشتر و در نتیجه کیفیت آنها کمتر می‌باشد.



شکل ۳-۹ - اولین فریم کد شده ویدئوی BLR در سناریوی (A) تک-QP و (B) چند-QP

۳-۱-۳- جمع‌بندی

در این بخش از رساله، یک مدل توجه بصری مخصوص کاربرد بازی ارائه گردید. این مدل با ترکیب نتایج الگوریتم‌های محاسباتی توجه پایین-به-بالا و بالا-به-پایین، اهمیت نواحی مختلف ویدئوی بازی از دید بازیکن را تخمین می‌زند. پیمانانه بالا-به-پایین این مدل نقش اصلی در تعیین اهمیت اشیا بازی ایفا کرده و در نتیجه اطلاعات مناسبی از محتوای آن بدست می‌دهد. این اطلاعات در مرحله بعدی می‌تواند باعث بهبود خروجی کدکننده گردد. برای نمایش عملکرد مثبت مدل در کنترل نرخ بیت ویدئوها، یک نمونه از آن، با استفاده از دو الگوریتم مشخص، پیاده‌سازی شد. البته امکان استفاده از

سایر الگوریتم‌ها در این مدل نیز وجود دارد. نتایج آزمایش‌ها نشان می‌دهد با تنظیم QP در کدکننده ویدئو بر اساس مدل پیاده‌سازی شده، می‌توان بطور متوسط تا ۲۵ درصد نرخ بیت ویدئوها را کاهش داد.

۳-۲- مدل توجه مبتنی بر مهارت

مدل‌های توجه بصری معمول فرض می‌کنند الگوهای توجه بصری میان بینندگان مختلف تقریباً یکسان است. این فرض در کاربردهای تماشای آزادانه^۱ و جستجوی بصری^۲ و همچنین تماشای ویدئوهای غیرتعاملی صادق است. به عنوان مثال، در تماشای ویدئو کارگردان با استفاده از فن‌های خاص و جلوه‌های ویژه توجه مخاطب را هدایت می‌کند. با این وجود، این فرض در مورد کاربرد بازی صادق نیست. در کاربرد بازی با چالش به نام «مهارت کاربر» مواجه هستیم. «مهارت کاربر» در بازی‌ها باعث متفاوت شدن رفتار بازیکنان از یکدیگر می‌شود. در واقع، با افزایش مهارت، بازیکن به صورت هدفمندتری به صحنه‌های بازی نگاه می‌کند. به عنوان مثال، تجربه‌ی بازیکن این امکان را به وی می‌دهد تا رویدادهای آینده در بازی را پیش‌بینی کرده و برای افزایش سرعت عکس‌العمل خود در رسیدن به اهداف بازی، نگاهش را ناحیه خاصی از ویدئو معطوف کند. الگوهای توجه بازیکن‌های همسطح ممکن است به دلایلی، مانند عادت‌های خاص هر بازیکن، کاملاً یکسان نباشند، اما بخاطر مشترک بودن اهداف بازی تا حدود زیادی شبیه یکدیگر هستند.

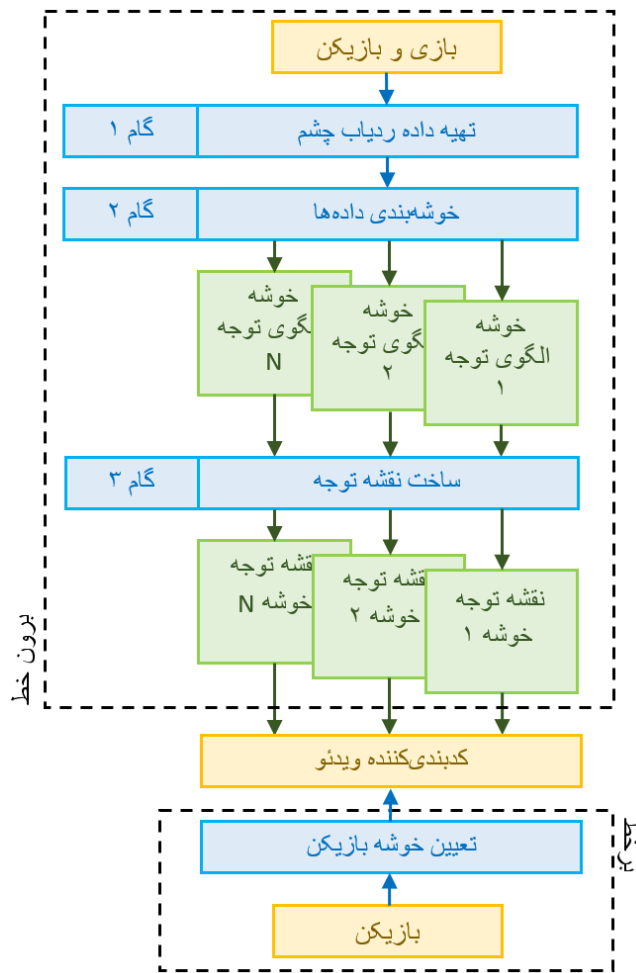
در این بخش از رساله یک مدل توجه مبتنی بر مهارت ارائه می‌کنیم. مزیت این مدل نسبت به مدل قبلی، عدم نیاز آن به منبع کد بازی و در نظر گرفتن مهارت بازیکن‌ها در ساخت نقشه‌های توجه است. این مدل بر مبنای داده‌های دستگاه ردیاب چشم که در حین بازی کردن بازیکن‌ها جمع‌آوری شده، توسعه می‌یابد. بعد از ساخت مدل، از آنجا که در عمل دستگاه ردیاب چشم وجود ندارد، مهارت و متعاقباً الگوی توجه بازیکن‌ها از روی توانایی آنها در کسب امتیاز در بازی بدست می‌آید.

۳-۲-۱- توسعه مدل

در این بخش، روش‌شناسی که برای توسعه مدل توجه مبتنی بر مهارت استفاده کردیم را تشریح می‌کنیم. اگرچه ما این روش‌شناسی را بر روی بازی توسعه‌یافته در دانشگاه تهران اجرا کردیم، اما امکان اجرای آن بر روی سایر بازی‌ها نیز وجود دارد. شکل ۳-۱۰ سه گام این روش‌شناسی را نشان می‌دهد. این شکل همچنین بخش‌های برخط و برون خط را مشخص می‌کند.

^۱ Free Viewing

^۲ Visual Search



شکل ۳-۱۰- مراحل توسعه مدل توجه مبتنی بر مهارت

همانطور که در شکل ۳-۱۰ مشخص است، تمام مراحل ساخت مدل توجه بصورت برون خط انجام می‌شود. بنابراین، هر چقدر هم که ساخت مدل محاسبات پیچیده‌ای داشته باشد، در زمان اجرا نه سربار تاخیرداری بر روی کارساز و نه سربار توان‌خواهی بر روی کارخواه، نخواهد داشت. اولین گام، انجام آزمایش‌های ردیاب چشم و جمع‌آوری داده‌های خیرگی از تعداد زیادی بازیکن و در حالت‌های مختلف می‌باشد. مرحله بعدی خوشه‌بندی داده‌های خیرگی جمع‌آوری شده و استخراج الگوهای توجه با استفاده از الگوریتم‌هایی مانند الگوریتم خوشه‌بندی Fuzzy C-means می‌باشد. از آنجا که مدل برای هر بازی یکبار ساخته می‌شود و بعد از آن به دفعات در سامانه CG مورد استفاده قرار می‌گیرد، در این مرحله دقت الگوریتم‌های مورد استفاده مهم‌تر از پیچیدگی آنها است. مهمترین نکته در گام خوشه‌بندی استفاده از یک متر مناسب است که در مرحله اجرا بشود آن را تخمین زد. به عنوان مثال، از آنجا که نمی‌توان به ازای هر بازیکن CG اطلاعات ردیاب چشم جمع‌آوری کرد، این متر نباید به فرد وابسته باشد. در عوض، باید بر پایه معیاری باشد که بتوان آن را برای هر فرد تخمین زد. به همین دلیل، ما استفاده از مهارت را پیشنهاد می‌کنیم زیرا می‌توان آن را از روی

جلسه‌های بازی قبلی بازیکن، امتیازهایش و همچنین الگوی ارسال کنترل‌های ورودی‌اش تخمین زد. در این رساله، از امتیاز برای تخمین مهارت و متعاقباً الگوی توجه بازیکن‌ها استفاده شده است. گام آخر تولید نقشه توجه برای هر یک از خوشه‌ها می‌باشد. وقتی نقشه‌ها ساخته شدند، در اختیار کدکننده قرار داده می‌شوند. کدکننده می‌تواند از این نقشه‌ها برای تنظیم پارامترهای کدکردن مانند مقادیر QP استفاده نماید و کیفیت نواحی مختلف از فریم‌های بازی را با توجه به اهمیت ادراکی آن نواحی کنترل کند بطوریکه کیفیت در نواحی پراهیمت، بالا و در سایر نواحی پایین باشد تا بیت‌های در دسترس به درستی در بین بخش‌های مختلف تقسیم گردند. در ادامه جزئیات هر یک از گام‌های روش‌شناسی ارائه خواهد شد.

۳-۲-۱-۱- جمع‌آوری داده‌های ردیاب چشم

تا زمان انجام این رساله سه دیتاست ردیاب چشم مخصوص کاربرد بازی یافت شد [۴۶، ۷۰، ۷۱] که با توجه به ایراداتی که داشتند که ما را به توسعه‌ی یک دیتاست جدید واداشت. اولین ایراد آن است که داده‌های ردیاب چشم تنها در مقاله‌ی [۴۶] در حین انجام بازی جمع‌آوری شده است. دو مجموعه داده‌ی دیگر تنها با نمایش ویدئو بازی به آزمایش‌دهندگان، داده‌های ردیاب چشم را جمع‌آوری کرده‌اند. بنابراین این مجموعه داده‌ها قابل اعتماد نیستند. همچنین مجموعه داده‌ی [۴۶] نیز دارای مشکلات عمده‌ای از جمله به‌روز نبودن بازی‌ها، دقت مکانی پایین ویدئو بازی و تعداد کم شرکت‌کنندگان آزمایش است. اما مهم‌ترین ایراد این مجموعه داده استفاده از ردیاب چشم چانه-ثابت^۱ است. با استفاده از این ردیاب کاربر نمی‌تواند آزادانه حرکت داشته باشد و این در کاربرد بازی، بر روی توجه بازیکن‌ها تاثیرگذار است. در جدول ۳-۸ این دیتاست‌ها با یکدیگر مقایسه شده‌اند. در تهیه دیتاست مخصوص بازی در این رساله به موارد زیر توجه شده است:

- از آنجا که در حال حاضر شرکت‌های CG بازی‌ها را دقت مکانی HD جویبارسازی می‌کنند، از دقت مکانی 720p استفاده گردید.
- بازی انتخاب شده توسط تیمی از توسعه‌دهندگان در دانشگاه تهران ساخته شده است. با همکاری این تیم، بازی بگونه‌ای تغییر داده شد تا اشیا و رویدادهای بازی در هر جلسه ضبط گردند تا در پژوهش‌های آتی قابل استفاده باشند.
- تعداد شرکت‌کنندگان مناسب مدل پیشنهادی است. اگر تعداد شرکت‌کنندگان کم باشد، دسته‌بندی افراد معنی‌دار نخواهد بود.
- از دستگاه ردیاب چشم راه‌دور^۲ استفاده شد. این دستگاه بر خلاف دستگاه‌های چانه-ثابت و سر-بند^۱ حرکت بازیکن را محدود نمی‌کنند. در نتیجه واکنش‌های فیزیکی بازیکن در حین بازی آزادانه انجام شده و الگوی توجه وی به واقعیت نزدیک‌تر است.

^۱ Chin rest

^۲ Remote

جدول ۳-۸- ویژگی دیتاست تهیه شده و دیگر دیتاست‌های مرتبط با کاربرد بازی

مجموعه داده‌ی پیشنهادی	مقاله‌ی [۴۶]	CRCNS[71]	DIEM[70]	
●	●	○	○	داده‌ها در حین بازی جمع‌آوری شده
○	○	●	●	داده‌ها از ویدئو بازی جمع‌آوری شده
●	●	●	○	ویدئو بازی
○	○	○	●	ویدئو تبلیغ بازی
۸۴	۵	۸	-	تعداد شرکت‌کننده
۱۳۵	۲۴	-	۴	تعداد ویدئو
۱۲۸۰*۷۲۰	۶۸۰*۴۸۰	۶۸۰*۴۸۰	متنوع	کیفیت
خام	خام	MPEG-1	-	فرمت ویدئو
از راه دور	چانه ثابت	چانه ثابت	-	نوع ردیاب چشم

۳-۲-۱-۱-۱- بازی «عروسک قشنگ من، سُمی»

بازی «عروسک قشنگ من، سُمی» توسط تیمی چهار نفره، متشکل از دو برنامه‌نویس و دو طراح گرافیک در مسابقات دانشجویی نمایشگاه بین‌المللی تهران در تابستان ۱۳۹۳ تولید شده است. در نوشتن این بازی از موتور بازی GameMaker Studio استفاده شده است. سبک این بازی شلیک‌گر نَوردش افقی^۲ است. این بازی شامل هفت مرحله می‌باشد و هر مرحله دو قسمت دارد. در قسمت اول هر مرحله، شخصیت اصلی می‌تواند به دشمنان شلیک کرده و موارد جمع‌شدنی^۳ موجود در صحنه، مانند قلب و صدف، را جمع کند. هر چند وقت یکبار موانعی هم در سر راهش قرار می‌گیرد که بایستی با پریدن، از آن‌ها اجتناب کند. در قسمت دوم شخصیت اصلی در حال سقوط است. هدف کاربر در این قسمت، تنها جمع‌آوری موارد جمع‌شدنی است.

^۱ Head Mounted^۲ Side Scrolling Shooter^۳ Collectable Items



شکل ۳-۱۱ - صحنه‌هایی از بازی «عروسک قشنگ من، سُمی»

در شکل ۳-۱۰ تصاویری از بازی نمایش داده شده است. در این بازی سه نوع دشمن وجود دارد. شخصیت بازی باید با یکی از سه نوع اسلحه خود (شاتگان، یوزی و بازوکا) این دشمنان را از بین ببرد. اسلحه‌های قوی‌تر امتیاز بیشتری به همراه دارند. البته هزینه استفاده از آنها هم بیشتر است. بنابراین بازیکن برای بدست آوردن امتیاز بیشتر باید در انتخاب آنها دقت داشته باشد. جهت پیاده‌سازی مدل توجه مبتنی بر مهارت پیشنهادی، داده‌های ردیاب چشم برای مرحله اول این بازی جمع‌آوری گردید. جزئیات این کار در بخش بعدی توضیح داده شده است.

۳-۲-۱-۱-۲-۳ - روال جمع‌آوری داده

از آنجا که ثابت نگه داشتن سر برای بازیکن‌ها در حین بازی در عمل غیرممکن است، از دستگاه ردیاب چشم راه‌دور Tobii X2-30 استفاده کردیم [۷۲]. این دستگاه یک جعبه حرکت سر مجازی در جلوی خود دارد و تا وقتی سر کاربر از این جعبه خارج نشده است می‌تواند چشم وی را ردیابی کند. دقت دستگاه در فاصله ۶۰ سانتیمتری با واسنجی^۱ ۹-نقطه حدود ۰,۴ درجه گزارش شده است [۷۳]. جدول ۳-۹ اطلاعات دموگرافیک شرکت‌کنندگان را نشان می‌دهد. هشتاد نفر در بازه سنی ۱۹ تا ۳۰ سال در این آزمایش شرکت کردند و در مجموع ۱۳۵ جلسه جمع‌آوری داده برگزار شد. حدود ۲۳٪ از این افراد خانم بودند. توجه داشته باشید که تجربه افراد در این جدول توسط خود آنها گزارش شده

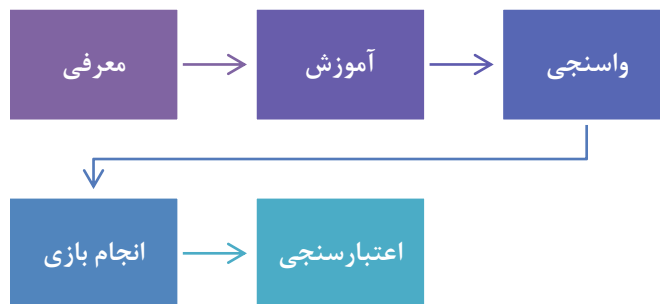
^۱ Calibration

است و مربوط به مهارت کلی آنها در انجام بازی‌های رایانه‌ای می‌باشد. اگرچه مهارت یک فرد بطور کلی بی‌ارتباط با مهارت وی در یک بازی خاص نیست، اما همه افراد با بازی سُمی ناآشنا بودند.

جدول ۳-۹- اطلاعات دموگرافیک شرکت‌کنندگان در آزمایش

تجربه در بازی‌های رایانه‌ای از نظر کاربران				
عالی	خوب	متوسط	ضعیف	خیلی ضعیف
۷,۶۹٪	۱۶,۶۷٪	۲۸,۲۱٪	۲۴,۳۶٪	۲۳,۰۸٪
میزان ساعت بازی در ماه				
بیش از ۳۰	۳۰-۲۱	۲۰-۱۱	۱۰-۶	کمتر از ۵
۳,۸۵٪	۱۲,۸۲٪	۲۵,۶۴٪	۴۸,۷۲٪	۸,۹۷٪
سکویی که تاکنون کاربر بر روی آن بازی کرده است				
گوشی همراه	تبلت	کنسول	کامپیوتر	
۷۳,۰۸٪	۳۵,۹٪	۲۳,۰۸٪	۸۰,۷۷٪	
سبک بازی که کاربر تا کنون بازی کرده است				
ورزشی	نوردش افقی	دویدن		
۷۱,۷۹٪	۴۲,۳۱٪	۵۶,۴۱٪		
شلیکگر به بالا	شلیکگر	بی‌انتها		
۳۲,۰۵٪	۶۷,۹۵٪	۴۲,۳۱٪		

هر جلسه از پنج گام تشکیل می‌شود: معرفی، آموزش، واسنجی، بازی و اعتبارسنجی. شکل ۳-۱۲ ترتیب این گام‌ها را نشان می‌دهد.



شکل ۳-۱۲- گام‌های آزمایش ردیاب چشم

برای هر جلسه مرحله واسنجی و بازی به طور متوسط ۷ دقیقه زمان برد، در حالی که سه مرحله دیگر ۱۰ دقیقه طول کشید. افرادی که بیش از یک بار بازی کردند، مرحله معرفی و آموزش را تنها یک بار انجام دادند.

در گام اول، توضیح مختصری از بازی به شرکت‌کنندگان داده شد. نحوه پریدن، شلیک‌کردن و تغییر اسلحه به آن‌ها آموزش داده شد. همچنین تصاویری از بازی که در آن‌ها دشمنان، موانع و اشیا بازی مشخص شده بود، نشان داده شد. در گام بعدی، از شرکت‌کنندگان خواسته شد تا برای مدت کوتاهی

بازی کنند تا بیشتر با آن آشنا شوند. در گام سوم، واسنجی ۹-نقطه انجام شد. این روال توسط سازنده برای افزایش دقت جمع‌آوری داده توصیه شده است. بعد از انجام واسنجی، بازی با کلیک بر روی دکمه شروع آغاز می‌شد. در حین بازی اطلاعات موشواره و صفحه‌کلید و ویدئوی بازی به همراه داده‌های ردیاب چشم ضبط شدند. برای ذخیره‌سازی فریم‌های بازی از یک نرم افزار تجاری استفاده شد که ویدئو را بدون از دست رفتن داده ذخیره می‌کند. در آخرین مرحله، اطلاعات ردیاب چشم جمع‌آوری بررسی شد تا از معتبر بودن آن‌ها اطمینان حاصل شود. اطلاعات یک جلسه معتبر در نظر گرفته شد، اگر بیش از نود درصد رکوردهای جمع‌آوری شده معتبر بودند. یک رکورد توسط دستگاه معتبر گزارش می‌شود، اگر دستگاه توانسته باشد در آن لحظه هر دو چشم فرد را شناسایی کرده باشد.

۲-۱-۲-۳- خوشه‌بندی

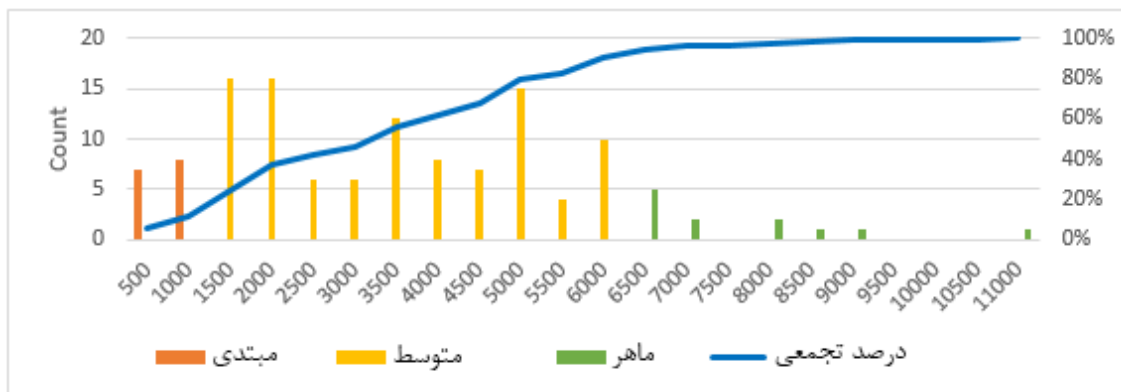
در این بخش، ابتدا با تحلیل داده‌های جمع‌آوری شده تاثیر مهارت بازیکن‌ها بر روی الگوهای توجه‌شان را نشان می‌دهیم. سپس یک روش برای بدست آوردن نقشه توجه ارائه می‌کنیم. در این رساله، شرکت‌کنندگان از نظر مهارت به سه دسته تقسیم شدند. اما تعداد سطوح مهارت می‌تواند برای هر بازی متفاوت باشد. یکی از عوامل موثر در این زمینه اهداف بازی است. به عنوان مثال، در بازی‌های سبک مسابقه ماشین، از یک طرف فعالیت‌هایی که منجر به بدست آوردن امتیازهای بالا می‌شوند، طوری هستند که بازیکن باید همواره به جاده نگاه کند و سریع واکنش نشان دهد. از طرف دیگر، جهت دوربین همواره طوری تنظیم می‌شود که جاده در وسط تصویر قرار گیرد. بنابراین، در چنین بازی‌هایی الگوهای توجه بین بازیکن‌ها نسبت به بازی‌هایی که در آنها اهداف بازی در صحنه پخش شده اند، تمایز کمتری دارند. اگرچه، حتی در بازی‌های سبک مسابقه ماشین نیز بازیکن‌های با مهارت بالا تمایل بیشتری به استفاده از نقشه و آینه عقب دارند که باعث می‌شود الگوی توجه آن‌ها نسبت به سایرین کمی متفاوت باشد.

جدول ۳-۱۰ بازه امتیاز هر سطح از مهارت را نشان می‌دهد. جزئیات بدست آوردن مرز امتیاز مهارت‌های مبتدی، متوسط و حرفه‌ای در بخش بعد توضیح داده می‌شود. شکل ۳-۱۳ نیز بافت‌نگاشت^۱ امتیازها را نشان می‌دهد. خط آبی در این نمودار، درصد انباشته شرکت‌کنندگان با امتیاز خاص را نشان می‌دهد.

جدول ۳-۱۰- تقسیم‌بندی شرکت‌کنندگان به سه گروه

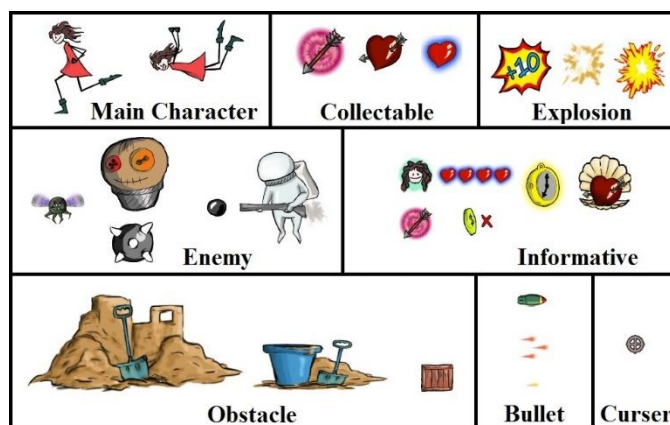
گروه	بازه امتیاز
مبتدی	زیر ۱۰۰۰
متوسط	بین ۱۰۰۰ تا ۶۰۰۰
حرفه‌ای	بالای ۶۰۰۰

^۱ Histogram



شکل ۳-۱۳ - بافت‌نگاشت امتیازهای شرکت‌کنندگان

توجه داشته باشید که مهارت‌های جدول ۳-۱۰ مربوط به بازی سُمی هستند و نباید با تجربه کلی بازیکن‌ها که توسط خودشان گزارش شده و در جدول ۳-۹ آمده‌اند، اشتباه گرفته شوند. به عنوان مثال، فردی که تجربه عالی گزارش کرده است، ممکن است اکثر اوقات بازی فوتبال بازی کرده باشد. چنین فردی، اگرچه نسبت به کسی که با نحوه استفاده از صفحه کلید برای بازی آشنا نیست، مهارت بیشتری دارد، اما عملکرد وی در بازی سُمی می‌تواند با بازی فوتبال متفاوت باشد. بدون در نظر پس‌زمینه، این بازی بیش از بیست نوع شی دارد. البته توجه داشته باشید که تعداد انواع اشیا بازی‌های CG ممکن خیلی بیشتر از این باشد. اما هدف ما نشان دادن رابطه مهارت با الگوهای توجه است. بنابراین استفاده از این بازی برای این هدف مناسب است. برای افزایش تعمیم‌پذیری به جای تحلیل الگوها در سطح شی، ابتدا اشیا بازی را دسته‌بندی و سپس الگوهای هر دسته را تحلیل می‌کنیم. به همین منظور، اشیا بازی را به هشت دسته تقسیم کردیم. شکل ۳-۱۴ دسته‌بندی ما را نشان می‌دهد.

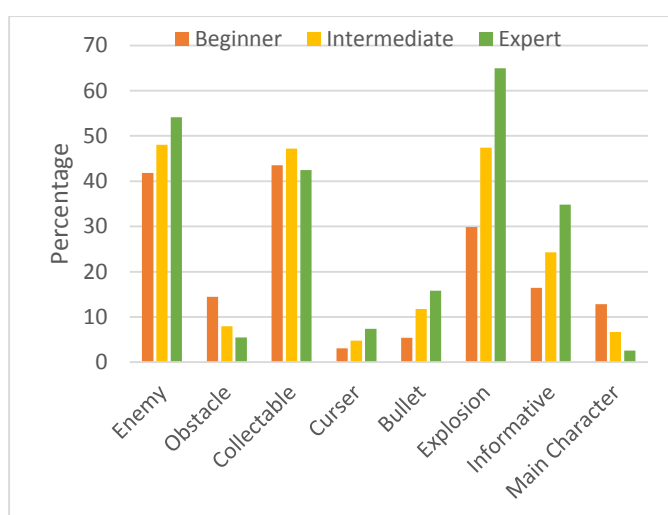


شکل ۳-۱۴ - دسته‌بندی اشیا بازی سُمی

دسته‌بندی‌های متفاوت، الگوهای توجه متفاوتی به همراه خواهند داشت. به عبارت دیگر، مقدار توجه در هر دسته به اشیا موجود در آن دسته وابسته است. با این وجود، نحوه دسته‌بندی اشیا موضوع این

رساله نیست. این دسته‌بندی فقط به منظور انتقال بهتر مطلب انجام شده است و در ساخت مدل از آن استفاده نمی‌شود. با این حال، اطلاعاتی که در این جا ارائه می‌شود، می‌تواند در ساخت مدل‌های توجه مبتنی بر اشیا بسیار مفید باشد.

برای تحلیل داده‌ها، درصد نقاط توجهی که در جعبه محیطی اشیا موجود در هر دسته قرار می‌گرفتند را محاسبه کردیم. همچنین جعبه محیطی را در هر چهار جهت ۱۶ پیکسل بزرگتر گرفتیم تا دقت دستگاه ردیاب چشم لحاظ شود. متوسط توجه در هر دسته برای مهارت‌های مختلف را نشان می‌دهد. توجه داشته باشید که از آنجا که تعداد ظهور اشیا از دسته‌های مختلف در طول بازی متفاوت بوده است، مقایسه توجه میان دسته‌های مختلف منصفانه نیست. در واقع، تفاوت میزان توجه بین بازیکن‌ها با مهارت‌های مختلف در هر دسته باید بصورت جداگانه بررسی شود.

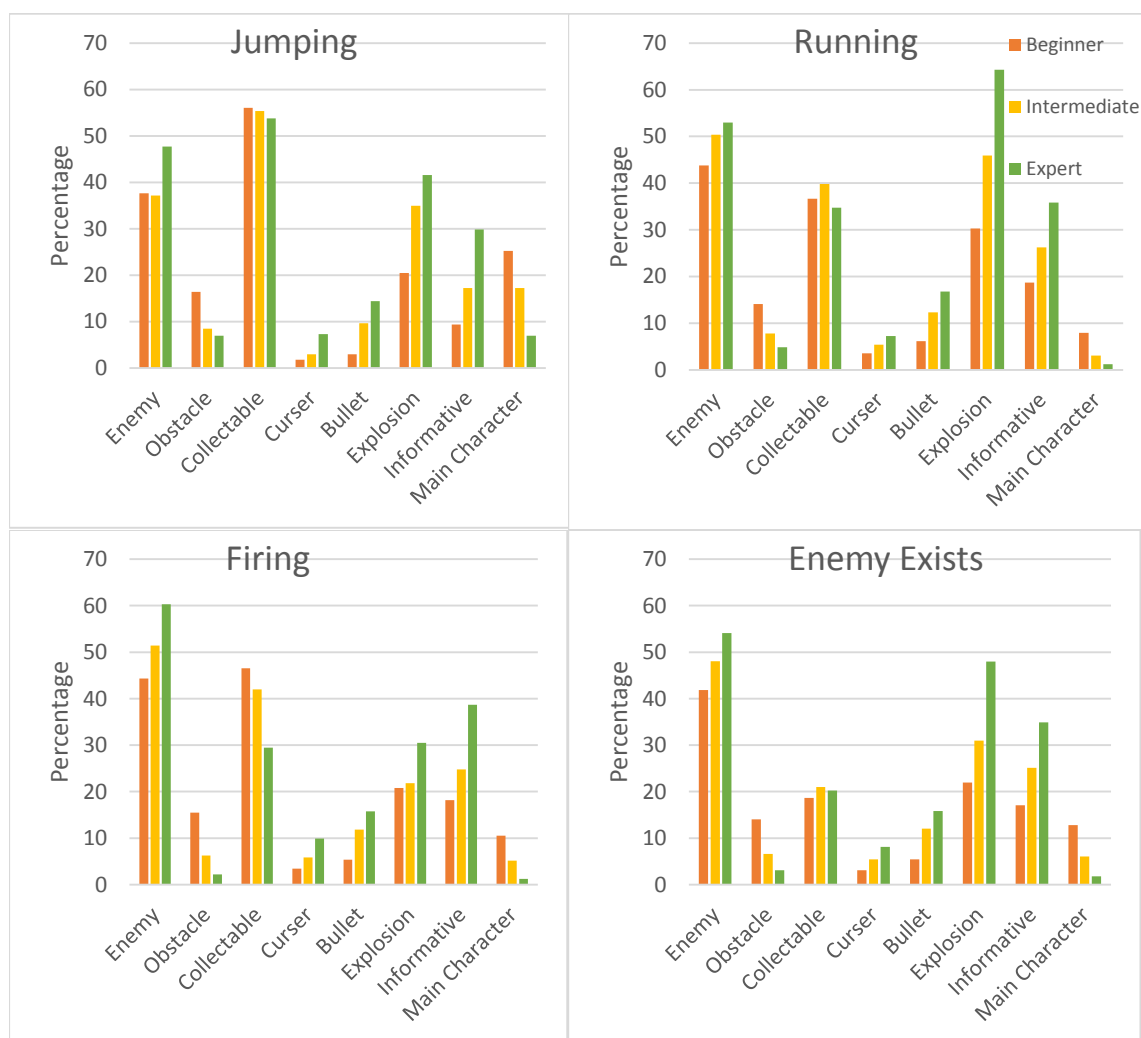


شکل ۳-۱۵ - متوسط توجه در هر دسته برای مهارت‌های مختلف

این شکل تاثیر مهارت بازیکن بر روی الگوی توجه وی را نشان می‌دهد. به عنوان مثال، افراد حرفه‌ای دو برابر افراد مبتدی به موارد حاوی اطلاعات در بازی توجه داشته‌اند. بصورت شهودی می‌توان این رفتار را اینگونه توجیه کرد که بازیکن‌های حرفه‌ای تمایل دارند امتیاز خود با خواندن اطلاعات مفید بازی هرچه بیشتر کنند. این شکل نه تنها، مقدار توجه برای هر سطح از مهارت را نشان می‌دهد، بلکه شیفت در توجه با افزایش مهارت بازیکن را نیز نشان می‌دهد. به عنوان مثال، هنگامیکه بازیکن با محیط بازی اخت می‌شود، دیگر بدون نگاه کردن مستقیم به آنها، از رویشان می‌پرد. به این ترتیب، می‌تواند دشمنان بیشتری بکشد و موارد بیشتری جمع کند و امتیاز بالاتری کسب نماید. در حقیقت، هر چه بیشتر بازی کند، توانایی تخمین اینکه موانع چه هنگام به وی نزدیک می‌شوند را بهتر کسب می‌کند. در این حالت، زمان را در ذهن خود نگه می‌دارد و وقتی آن زمان گذشت دکمه پرش را فشار می‌دهد، در حالیکه ممکن است به جای دیگری از صحنه توجه داشته باشد. این رفتار در دسته موانع

در شکل ۳-۱۵ مشخص است. متوسط وردایی در میزان توجه میان بازیکن‌های مبتدی، متوسط و حرفه‌ای به ترتیب ۱۲,۲۲، ۱۴,۳ و ۱۴,۱۶ درصد می‌باشد.

بخش دیگر تحلیل ما به بررسی الگوهای توجه در حالت‌های مختلف بازی اختصاص دارد. به همین منظور چهار حالت پرش، دویدن، شلیک و وجود دشمن را نظر می‌گیریم. دو حالت اول همزمان اتفاق نمی‌افتند ولی دو حالت بعدی ممکن است با هم اشتراک داشته باشند. توجه داشته باشید که این حالت‌ها، همانند دسته‌بندی‌ها، برای بهتر به تصویر کشیدن تاثیر مهارت در الگوهای توجه هستند و در ساخت مدل از آنها استفاده نمی‌شود. شکل ۳-۱۶ متوسط توجه در هر دسته در حالت‌های مختلف را نشان می‌دهد.



شکل ۳-۱۶ - متوسط توجه در هر دسته برای مهارت‌ها و حالت‌های مختلف

همانطور که مشخص است مقدار توجه در حالت‌های مختلف کاملاً قابل تمیز است. به عنوان مثال، بازیکن هنگام پرش توجه بیشتری به کاراکتر خود می‌کنند. این رفتار در بازیکن‌های مبتدی بیشتر مشخص است. به عنوان مثال دیگر، توجه به موارد جمع‌شونده در حالت پرش نسبت به حالت‌های

دیگر بیشتر است. علاوه بر منطق بازی، طراحی بازی نیز در این مسئله موثر است. زیرا موارد جمع‌شونده در بازی در ارتفاعات بالا قرار دارند. به همین دلیل، طراحان بازی می‌توانند با تحلیل الگوهای توجه، یاد بگیرند که چگونه باید توجه بازیکن را به یک هدف مشخص در بازی جلب کرد. به عنوان مثال، یک بازی ماجراجویی را در نظر بگیرید که در آن شخصیت بازی در قالب یک داستان به حل مسئله و کشف ندانسته‌ها می‌پردازد. در چنین بازی‌ای، اگر بازیکن نتواند در زمان قابل تحملی پیام‌های کدگشایی را پیدا کند، از بازی دلسرد خواهد شد و آن را ترک خواهد کرد. این موضوع به هیچ وجه مطلوب یک شرکت بازی ساز نیست. آنها می‌توانند با استفاده از مدل‌های توجه مبتنی بر مهارت با توجه به مهارت بازیکن پیام‌های کدگشا را در دایره توجه بازیکن قرار دهند. مهارت در اینجا مهم است زیرا اگر محاسب این پیام‌ها به آسانی قابل کشف باشد، بازی برای افراد ماهر کسل‌کننده خواهد شد و مشکلی یکسان البته از جهت مخالف بوجود خواهد آمد.

۳-۲-۱-۲-۱-۲-۳ روال خوشه‌بندی

بعد از تحلیل الگوها و مشخص شدن تاثیر مهارت در توجه، در اینجا به خوشه‌بندی داده‌های خیرگی می‌پردازیم. برای این کار ابتدا نقشه اهمیت هر یک از بازیکن‌ها را بدست می‌آوریم. پس، هر یک از فریم‌های بازی را به بلاک‌های ۱۶ در ۱۶ تقسیم می‌کنیم. این انتخاب بر اساس اندازه کوچک‌ترین واحد کدشونده در کدکننده H.264/AVC می‌باشد. بدین ترتیب، یک شبکه ۴۵ در ۸۰ از بلاک‌ها بدست می‌آید. مقدار متناظر هر بلاک یک گذاشته می‌شود، اگر نقطه خیرگی در آن بلاک افتاده باشد. در غیراینصورت مقدار آن بلاک صفر در نظر گرفته می‌شود. سپس مقادیر بلاک‌های متناظر در کل فریم‌های مربوط به یک بازیکن جمع و تقسیم بر تعداد آن فریم‌ها می‌شود. این روال باعث کاهش بُعد داده‌هایمان به ۳۶۰۰ می‌شود. از آنجا که این تعداد هنوز زیاد است، با استفاده از روش تحلیل مولفه‌های اصلی^۱ (PCA) بُعد داده‌ها را باز هم کاهش می‌دهیم.

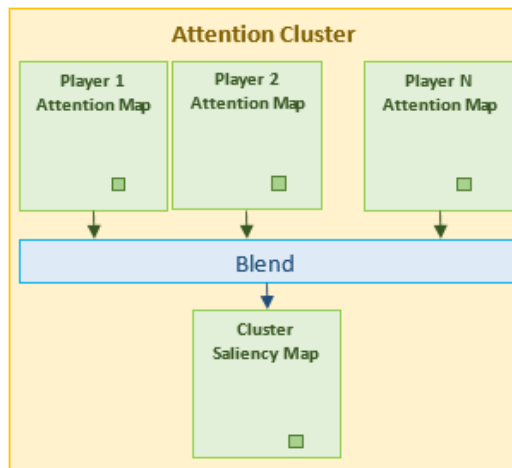
هر چه تعداد خوشه‌های متمایز بیشتر باشد، نقشه‌های توجه خاص‌تری بدست می‌آید و دقت مدل بیشتر می‌شود. با این حال، با افزایش تعداد خوشه‌ها، اندازه آنها کوچک و کوچک‌تر می‌شود و نقشه‌های ساخته شده از روی خوشه‌هایی که تعداد بازیکن‌ها در آنها کم است، تعمیم‌پذیر نخواهند بود. در این رساله، ما تعداد سه خوشه را انتخاب کردیم. زیرا انتخاب تعداد بیشتر باعث می‌شد بعضی خوشه‌ها برای انواع تعداد مختلف مولفه‌های PCA تنها دو یا سه عضو بیشتر نداشته باشند.

روش PCA را با تعداد مختلفی مولفه اعمال کردیم و نتیجه را با استفاده از الگوریتم Fuzzy C-means خوشه‌بندی کردیم. سپس میانگین و انحراف معیار اعضای هر خوشه با مرکز آن را بدست آوردیم. بهترین تعداد مولفه PCA پنج بود که میانگین و انحراف معیار آن به ترتیب ۰,۸۶ و ۰,۴۳۵ شد.

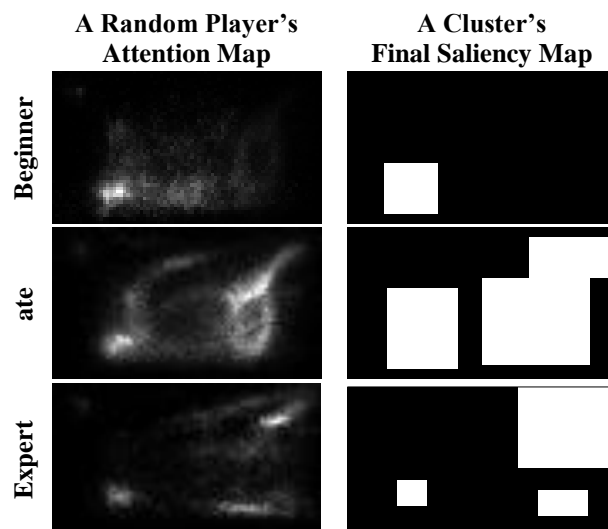
^۱ Principle Component Analysis (PCA)

۳-۲-۱-۳ - ساخت نقشه‌های توجه

به ازای هر خوشه، یک نقشه توجه ساخته می‌شود. به همین منظور، نقشه توجه بازیکن‌های موجود در یک خوشه را با یکدیگر ترکیب کرده و نقشه نهایی را تولید می‌کنیم. نقشه توجه هر بازیکن، در واقع احتمال توجه کردن وی به هر بلاک از ویدئو را مشخص می‌کند. تعداد بلاک‌های نقشه نهایی خوشه نیز برابر تعداد بلاک‌های نقشه هر یک از بازیکن‌های موجود در آن خوشه است. برای محاسبه مقدار هر بلاک در نقشه توجه خوشه، ابتدا میانگین بلاک‌های متناظر در نقشه توجه بازیکن‌های موجود در آن خوشه را بدست می‌آوریم. سپس با قطع ۰,۳ مقدار نهایی آن را محاسبه می‌کنیم. شمای بلاک‌های متناظر در نقشه‌های توجه بازیکن‌های یک خوشه و نقشه نهایی آن خوشه را نشان می‌دهد. شکل ۳-۱۷ نقشه نهایی تولید شده برای هر یک از سه خوشه‌ی مبتندی، متوسط و حرفه‌ای را نشان می‌دهد. شکل ۳-۱۸ در این شکل همچنین نقشه یک بازیکن نمونه با همان سطح از مهارت نمایش داده شده است.



شکل ۳-۱۷ - شمای بلاک‌های متناظر در نقشه‌های توجه



شکل ۳-۱۸ - نقشه توجه هر سطح از مهارت به همراه نقشه یک بازیکن نمونه از همان سطح

۳-۲-۲- پیش بینی الگوی توجه بازیکن

نقشه توجه هر خوشه بصورت برون خط و بر اساس داده‌های ردیاب چشم بدست می‌آید. مرحله بعد، تشخیص خوشه (الگوی توجه) یک بازیکن در عمل هنگام استفاده از سامانه CG است. از آنجا که دستگاه‌های ردیاب چشم، ارتباط انسان و رایانه^۱ را راحت‌تر می‌کنند و قیمت آنها به مرور کاهش می‌یابد، دور از انتظار نیست که در آینده آنها را بر روی دستگاه‌های همراه نصب کنند. در این صورت، شناسایی الگوی توجه هر بازیکن آسان‌تر از هر زمانی خواهد شد. به هر حال، در حال حاضر این دستگاه‌ها در دسترس نیستند و تشخیص الگوی بازیکن از روی داده خیرگی امکان‌پذیر نیست. بنابراین باید خوشه بازیکن‌ها در عمل را پیش‌بینی کنیم. از جمله داده‌هایی که برای این منظور در دسترس هستند، می‌توان به تاریخچه فعالیت‌ها و بازی‌های کاربر و همچنین فرامین کنترلی که وی به سامانه می‌فرستد، نام برد.

تعیین خوشه توجه بازیکن بصورت برخط انجام می‌شود. در این رساله، این کار بر اساس امتیاز کاربر است و می‌تواند در ابتدای هر جلسه بازی انجام پذیرد. این روش بار محاسباتی در سمت کارخواه نخواهد داشت، اگرچه بار سبکی، شامل تعیین خوشه و انتقال نقشه توجه به کدکننده، بر روی سمت کارساز بوجود می‌آید. اگر هیچ اطلاعاتی از بازیکن در دسترس نباشد، وی را مبتدی فرض می‌کنیم (امتیاز صفر). البته روش محتاطانه‌تر، آن است که وی را در خوشه بی‌اندازیم که مساحت نقشه بیشتری دارد و در نتیجه نواحی بیشتری از ویدئو با کیفیت بالاتر کد می‌شوند. به مرور به شکل‌گیری تاریخچه بازیکن، دقت تعیین خوشه افزایش می‌یابد.

برای اینکه مشخص شود که آیا تفاوت بین الگوهای توجه مربوط به عوامل تصادفی است و یا امتیاز در تعیین آنها دخیل است، از تحلیل ANOVA^۲ با ضریب اطمینان ۹۵ درصد استفاده کردیم. از آنجا که نسبت میانگین انحراف معیار بین دسته‌ها به میانگین انحراف معیار داخل دسته‌ها (۴۰,۱۴) که از توزیع F پیروی می‌کند از مقدار حیاتی توزیع F برای ضریب اطمینان ۹۵ درصد (۳,۱۱) بیشتر است، نتیجه می‌گیریم که از نظر آماری تاثیر امتیاز در دسته‌ها متفاوت است. جدول ۳-۱۱ نتایج این آزمایش را نشان می‌دهد.

جدول ۳-۱۱- نتایج تست ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2.21E+08	2	1.11E+08	40.14571	1.14E-12	3.115366
Within Groups	2.12E+08	77	2753106			
Total	4.33E+08	79				

^۱ Human-Computer Interaction (HCI)

^۲ Analysis of Variance (ANOVA)

برای بررسی نقش امتیاز بین هر دو خوشه، از آزمون t-Test استفاده کردیم. از آنجا که مقدار t-Test یا از مقدار حیاتی بزرگ‌تر و یا از منفی مقدار حیاتی کوچک‌تر است، امتیاز عامل متمایزکننده‌ای در بین گروه‌ها محسوب می‌شود. **Error! Reference source not found.** و جدول ۳-۱۳ و جدول ۳-۱۴ به ترتیب نتایج آزمون t-Test بین گروه‌های مبتدی-متوسط، متوسط-حرفه‌ای و مبتدی-حرفه‌ای را نشان می‌دهد.

جدول ۳-۱۲- نتایج تست t-Test برای مبتدی-متوسط

Hypothesized Mean Difference	0
df	63
t Stat	4.900292055
P(T<=t) one-tail	3.5104E-06
t Critical one-tail	1.669402222
P(T<=t) two-tail	7.0208E-06
t Critical two-tail	1.998340543

جدول ۳-۱۳- نتایج تست t-Test برای متوسط-حرفه‌ای

Hypothesized Mean Difference	0
df	19
t Stat	-4.64676
P(T<=t) one-tail	8.79E-05
t Critical one-tail	1.729133
P(T<=t) two-tail	0.000176
t Critical two-tail	2.093024

جدول ۳-۱۴- نتایج تست t-Test برای مبتدی-حرفه‌ای

Hypothesized Mean Difference	0
df	20
t Stat	7.462847078
P(T<=t) one-tail	1.67661E-07
t Critical one-tail	1.724718243
P(T<=t) two-tail	3.35321E-07
t Critical two-tail	2.085963447

در این بخش، بازیکن‌ها بر اساس امتیازشان به سه دسته تقسیم می‌شوند. برای این کار، ما دو مرز پایین و بالا را بدست آوردیم. مرز پایین مبتدی‌ها را از متوسط‌ها و مرز بالا متوسط‌ها را از حرفه‌ای‌ها جدا می‌کند. انتخاب این مرزها، دقت پیش‌بینی خوشه توجه بازیکن و متعاقباً کیفیت و نرخ فشرده‌سازی را تحت تاثیر قرار می‌دهد. انتخاب یک نقشه اشتباه دو عاقبت دارد. از یک طرف، نواحی

که مورد توجه کاربر هستند ممکن است کم اهمیت در نظر گرفته شده و کیفیت در آن نواحی کاهش داده شود. از طرف دیگر، نواحی که مورد توجه کاربر نیستند ممکن است پر اهمیت در نظر گرفته شده و بدون آنکه نیاز باشد، با نرخ بیت بیشتری کد شوند.

برای پیدا کردن مرزهای بالا و پایین، بیشترین و کمترین N درصد بافت‌نگاشت امتیاز را در نظر گرفتیم بطوریکه $N = i \cdot \Delta N, i \in \{1, 2, \dots, 6\}$ و $\Delta N = 5$ دقت‌های بدست آمده از این طریق در جدول ۳-۱۵ گزارش شده است. همانطور که مشخص است، انتخاب ۲۰ درصد (امتیاز ۱۰۰۰) به عنوان مرز پایین و انتخاب ۱۵ درصد (امتیاز ۶۰۰۰) به عنوان مرز بالا به ترتیب ۸۱٫۲۵ و ۹۲٫۰۶ درصد دقت دارند.

جدول ۳-۱۵- دقت مرزهای انتخاب شده

مرز پایین	امتیاز	دقت	مرز بالا	امتیاز	دقت
Least 5%	130	58.33%	Top 5%	7920	66.66%
Least 10%	550	66.66%	Top 10%	6240	78.57%
Least 15%	810	75.00%	Top 15%	6000	92.06%
Least 20%	930	81.25%	Top 20%	5780	90.47%
Least 25%	1300	80.50%	Top 25%	4820	87.30%
Least 30%	1550	79.75%	Top 30%	4510	84.12%

جدول ۳-۱۶- ماتریس درهم‌ریختگی پیش‌بینی خوشه‌ها

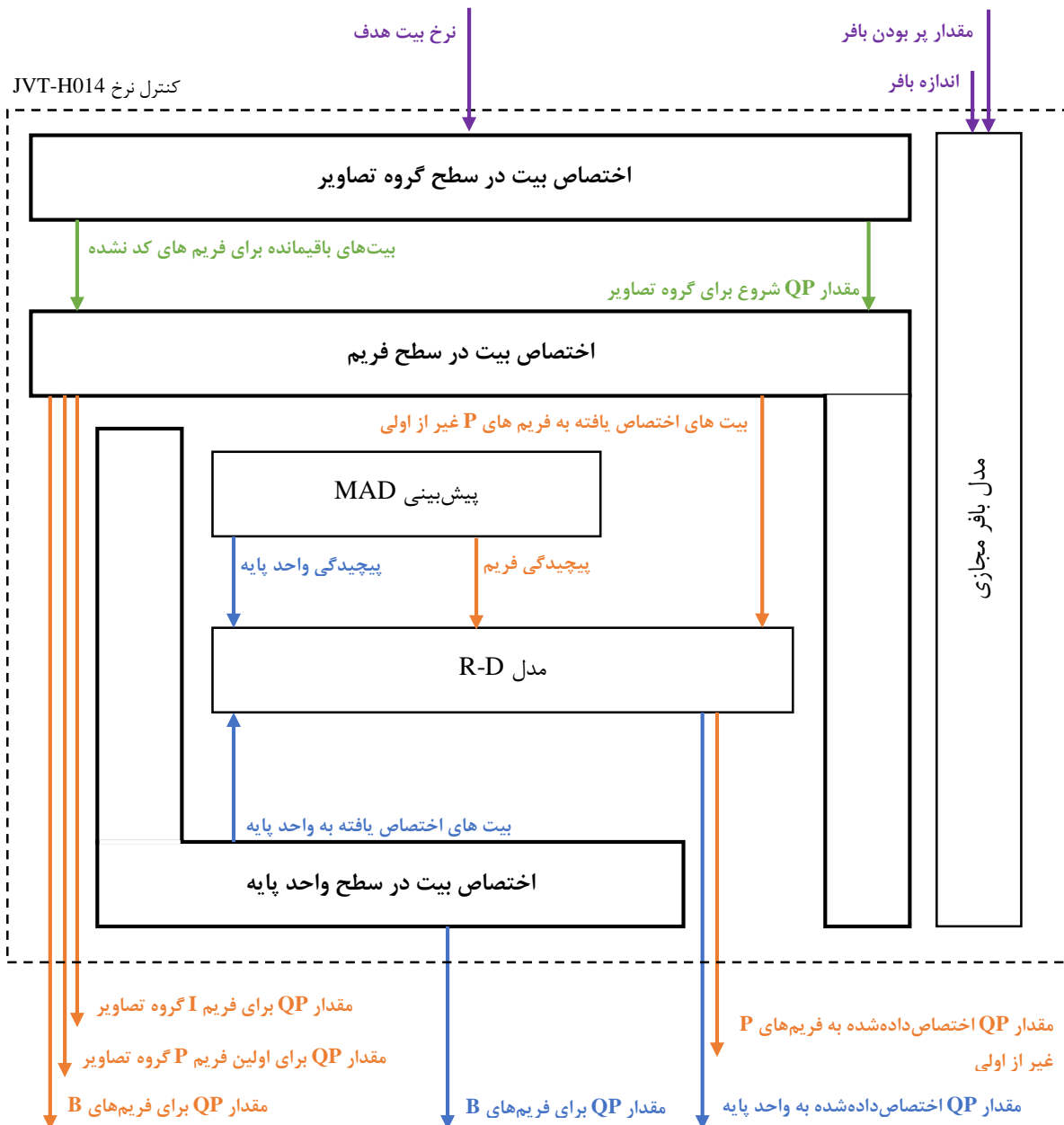
		پیش‌بینی شده		
		مبتدی	متوسط	حرفه‌ای
خوشه اصلی	مبتدی	62.5%	37.5%	0
	متوسط	0	97.43%	2.56%
	حرفه‌ای	0	14.28%	85.71%

از آنجا که سه خوشه وجود دارد، عملکرد نهایی در قالب ماتریس درهم‌ریختگی جدول ۳-۱۶ نشان داده شده است. دیتاست تست ما از ۲۰ درصد رکوردهایی که در مرحله آموزش استفاده نشده بودند، تشکیل شده است. با توجه به قطر این ماتریس، مرزهای انتخاب شده تا ۸۱٫۸۸ درصد در تشخیص خوشه توجه بازیکن‌ها، دقت دارند. ما معتقدیم این دقت با در نظر گرفتن ورودی‌های کاربر بیشتر نیز خواهد شد.

۳-۲-۳- اعمال مدل در کدکننده ویدئو

بعد از ساخت نقشه‌های اهمیت باید آنها را برای تنظیم پارامترهای کدکننده ویدئو استفاده کنیم، بطوریکه نرخ بیت بصورت موثر بین نواحی مختلف تقسیم گردد. در این خصوص روش‌های مختلفی پیشنهاد شده است. به هر حال، منطق پشت همه آنها کد کردن نواحی پراهمیت با کیفیت (نرخ بیت) بالا و سایر نواحی با کیفیت (نرخ بیت) پایین می‌باشد.

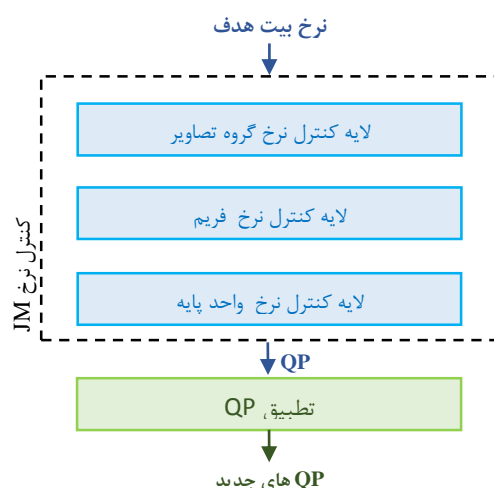
در حال حاضر، اکثر شرکت های CG از استاندارد H.264/AVC استفاده می کنند. بنابراین ما نیز از نرم افزار مرجع این استاندارد JM v18.4 [۷۴] استفاده کردیم. همانطور که در شکل ۳-۱۹ نشان داده شده است، سازوکار کنترل نرخ در JM سه لایه GOP، فریم و واحد پایه^۱ (BU) است [۷۵]. این سازوکار به یک مدل نرخ-اعوجاج درجه دو مجهز است که با استفاده از یک پیش بینی کننده خطی مقدار MAD را تخمین زده و در نهایت مقدار QP واحد کدشونده را محاسبه می کند.



شکل ۳-۱۹ - جزئیات کنترل کننده نرخ در JM

^۱ Basic Unit

به هر حال، محاسبه پیچیدگی ماکرو بلاک‌ها بدون در نظر گرفتن ویژگی‌های چشم انسان، رسیدن به کیفیت ادراکی مطلوب را تضمین نمی‌کند. بنابراین، برای بهره‌بردن از نقشه‌های توجه و در عین حال مزایای کنترل‌کننده نرخ اصلی، پیشنهاد می‌شود درست بعد از پیمانانه کنترل نرخ یک پیمانانه تطبیق QP قرار داده شود (شکل ۳-۲۰). وقتی QP واحد کدشونده توسط کنترل‌کننده نرخ مشخص شد، پیمانانه تطبیق مقدار QP ماکرو بلاک‌های درون BU را متناسب با مقدار متناظرشان در نقشه توجه تغییر می‌دهد.



شکل ۳-۲۰ - لایه های کنترل کننده نرخ در JM و محل پیمانانه تطبیق QP

تطبیق QP بر اساس نقشه توجه را می‌توان به شکل یک مسئله بهینه‌سازی فرمول‌بندی کرد [۵۶]. بر اساس این تحلیل، مقدار گام پیمانانه‌بندی باید متناسب با معکوس مقدار توجه انتخاب شود. نتایج آزمایش‌های ما نشان می‌دهد، استفاده از این روش باعث افزایش کیفیت ادراکی در نرخ بیت برابر می‌شود. با توجه به دردسترس بودن نقاط خیرگی، منصفانه آن است که برای اندازه‌گیری کیفیت از معیار EWPSNR استفاده کنیم.

جدول ۳-۱۷ - پارامترهای کدکننده استفاده شده در شبیه سازی‌ها

مقدار	نام
Baseline	پروفایل
۳,۱	سطح
۱	تعداد فریم‌های مرجع
EPZS	شمای تخمین حرکت
۳۲	دامنه جستجو
روشن	انتخاب مد بهینه نرخ-عوجاج

برای اندازه‌گیری تاثیر نقشه‌های توجه در نرخ فشرده‌سازی دو سری آزمون انجام دادیم. اول، کنترل‌کننده نرخ بیت را خاموش کردیم و مقادیر QP را بصورت تجربی تغییر دادیم. دوم، عملکرد کنترل‌کننده JM را نسبت به پیمانانه تطبیق QP سنجیدیم. در هر دو سری از نرم افزار مرجع JM استفاده کردیم. برای دسته‌بندی ماکرو بلاک‌ها در اسلایس‌های مجزا از FMO استفاده کردیم. از آنجا که دقت مکانی ویدئوهای ضبط شده 720p است، از درجه ۳,۱ استفاده کردیم. جدول ۳-۱۷ پارامترهای کدکننده در شبیه‌سازی‌های ما را خلاصه می‌کند.

۳-۲-۱- کنترل‌کننده نرخ خاموش

در این بخش، ویدئوها را به دو روش کد کردیم: تک-QP و چند-QP. ویدئوها را یک بار بدون استفاده از نقشه‌های توجه و در نتیجه تنها با یک QP کد کردیم. برای منصفانه بودن مقایسه، از چند فرد خبره درخواست کردیم این ویدئوها را ببینند و بیشترین مقدار QP را که برای آن در ویدئوها اعوجاجی نمی‌دیدند گزارش کنند. این مقدار برابر ۲۶ بود.

جدول ۳-۱۸- مقدار EWPSNR و نرخ بیت برای مقادیر مختلف QP

Scenario	B1		B2		B3		B4		B5		B6	
	EW PSNR	Bit Rate	EW PSNR	Bit Rate	EW PSNR	Bit Rate	EW PSNR	Bit Rate	EW PSNR	Bit Rate	EW PSNR	Bit Rate
QP (26)	43.33	1.94	43.95	1.90	44.19	2.13	45.17	2.07	44.47	1.77	43.85	2.09
QPs (28 , 26)	41.93	1.64	43.54	1.61	43.34	1.78	44.64	1.74	43.83	1.49	42.75	1.77
QPs (30 , 26)	40.41	1.37	43.11	1.35	42.31	1.49	44.05	1.45	43.12	1.24	41.50	1.48
QPs (32 , 26)	38.93	1.16	42.60	1.14	41.28	1.24	43.35	1.23	42.32	1.05	40.21	1.25

Scenario	I1		I2		I3		I4		I5		I6	
	EW PSNR	Bit Rate	EW PSNR	Bit Rate	EW PSNR	Bit Rate	EW PSNR	Bit Rate	EW PSNR	Bit Rate	EW PSNR	Bit Rate
QP (26)	44.22	2.37	45.35	4.01	42.81	1.80	43.23	1.90	43.76	2.12	44.00	2.09
QPs (28 , 26)	44.12	2.27	45.11	3.68	42.51	1.72	43.07	1.81	43.70	2.00	43.8	2.00
QPs (30 , 26)	43.97	2.17	44.82	3.35	42.14	1.63	42.87	1.71	43.60	1.87	43.58	1.90
QPs (32 , 26)	43.80	2.08	44.51	3.09	41.71	1.56	42.62	1.63	43.48	1.77	43.33	1.81

Scenario	E1		E2		E3		E4		E5		E6	
	EW PSNR	Bit Rate	EW PSNR	Bit Rate	EW PSNR	Bit Rate	EW PSNR	Bit Rate	EW PSNR	Bit Rate	EW PSNR	Bit Rate
QP (26)	45.16	2.79	46.31	2.21	47.16	2.00	44.02	1.89	44.79	1.95	46.30	2.76
QPs (28 , 26)	44.11	2.34	45.32	1.93	46.47	1.72	43.80	1.65	44.51	1.68	45.68	2.36
QPs (30 , 26)	42.98	1.95	44.11	1.69	45.59	1.46	43.48	1.44	44.16	1.46	44.93	1.99
QPs (32 , 26)	41.87	1.64	42.75	1.49	44.70	1.26	43.16	1.27	43.77	1.27	44.22	1.70

جدول ۳-۱۹- درصد تغییر EWPSNR و نرخ بیت برای مقادیر مختلف QP

	B1	B2	B3	B4	B5	B6	Average (%)
QPs (28, 26)	15.46	15.26	16.43	15.94	15.82	15.31	15.71
QPs (30, 26)	29.38	28.95	30.05	29.95	29.94	29.19	29.58
QPs (32, 26)	40.21	40	41.78	40.58	40.68	40.19	40.57

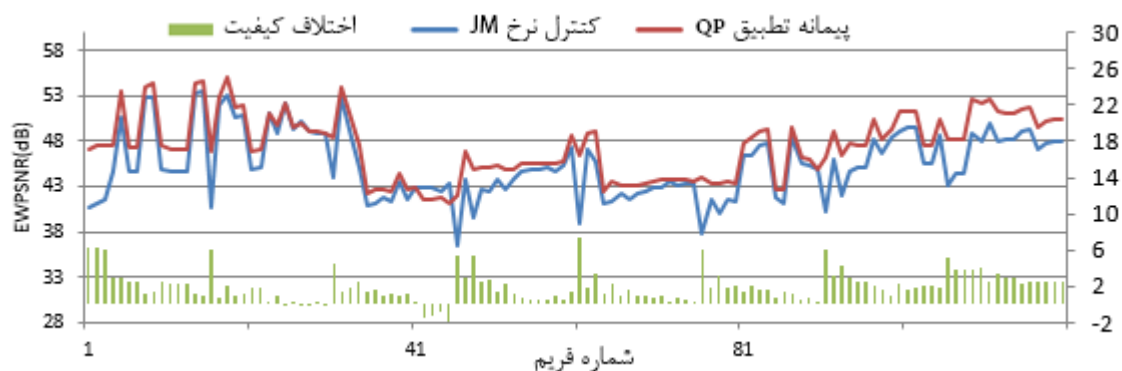
	I1	I2	I3	I4	I5	I6	Average (%)
QPs (28, 26)	4.22	8.23	4.44	4.74	5.66	4.31	5.27
QPs (30, 26)	8.44	16.46	9.44	10	11.79	9.09	10.87
QPs (32, 26)	12.24	22.94	13.33	14.21	16.51	13.4	15.44

	E1	E2	E3	E4	E5	E6	Average (%)
QPs (28, 26)	16.13	12.67	14	12.7	13.85	14.49	13.97
QPs (30, 26)	30.11	23.53	27	23.81	25.13	27.9	26.25
QPs (32, 26)	41.22	32.58	37	32.8	34.87	38.41	36.15

ویدئوهای مرجع یک بار دیگر، این بار با استفاده از نقشه‌های توجه کد شدند. در این مرحله، مقادیر QP متناسب با اهمیت ناحیه‌ها تنظیم شدند. تمام جفت‌های $(26, 26 + 2X)$ که در آن $1 \leq X \leq 3$ آزمایش شدند. جدول ۳-۱۸ مقدار EWPSNR و نرخ بیت این ویدئوها را در هر دو روش تک-QP و چند-QP نشان می‌دهد. جدول ۳-۱۹ درصد کاهش نرخ بیت در حالت چند-QP نسبت به حالت تک-QP را نشان می‌دهد.

۳-۲-۲-۳- کنترل‌کننده نرخ روشن

برای مقایسه عملکرد کنترل‌کننده نرخ JM با پیمانانه تطبیق QP ویدئوها را یک بار بدون استفاده از پیمانانه و بار دیگر با استفاده از پیمانانه کد کردیم و مقدار EWPSNR را اندازه گرفتیم. شکل ۳-۲۱ یک مثال از مقدار EWPSNR بر روی ۱۲۰ فریم متوالی از ویدئوی یک بازیکن متوسط است. همانطور که دیده می‌شود، کیفیت ویدئو هنگام استفاده از پیمانانه تطبیق QP بهتر است. البته توجه در چند فریم، کیفیت پایین آمده است که مربوط به خطای مدل توجه می‌شود.



شکل ۳-۲۱ - مقدار EWPSNR برای ۱۲۰ فریم متوالی از ویدئوی یک نمونه بازیکن متوسط

توجه داشته باشید که بازه EWPSNR و PSNR با یکدیگر متفاوت است. در واقع، در 1dB در PSNR با 1dB در EWPSNR متفاوت است و بازه حدود 40dB در EWPSNR منطقی است. جدول ۳-۲۰ متوسط EWPSNR در دو حالت با و بدون پیمانه تطبیق QP را نشان می‌دهد.

جدول ۳-۲۰- مقدار تغییر EWPSNR و نرخ بیت با و بدون استفاده از پیمانه تطبیق QP

	کنترل نرخ JM		تطبیق QP		اختلاف	
	EWPSNR	Bit Rate	EWPSNR	Bit Rate	EWPSNR	Bit Rate
مبتدی	40.19	1.91	40.84	1.92	+0.65	-0.01
متوسط	42.74	2.37	44.01	2.32	+1.27	+0.05
حرفه‌ای	41.63	1.71	43.6	1.71	+1.97	0.00

۳-۲-۴- ارزیابی دقت مدل توجه

برای اندازه‌گیری دقت مدل مبتنی بر مهارت پیشنهادی تعداد دفعاتی که نقاط خیرگی در نواحی پراهمیت قرار گرفته اند، را شمارش کردیم. البته، برای مقایسه منصفانه این مدل با سایر مدل‌های توجه بصری، مقدار قطع آن مدل‌ها را بگونه‌ای تنظیم کردیم که مساحت نقشه‌های تولید شده توسط همه مدل‌ها تقریباً یکسان شود.

جدول ۳-۲۱- دقت مدل‌های بصری و مدل پیشنهادی

مبتدی			
مدل	آستانه	مساحت مهم (%)	دقت (%)
Proposed	0.3	6.67	50.05
GVBS	0.5	8.56	13.34
ITTI	0.4	7.32	14.64
BMS	0.8	7.68	10.74
FES	0.7	8.36	5.91
JUDD	0.3	13.3	29.49
CovSal	0.6	7.60	3.86
متوسط			
مدل	آستانه	مساحت مهم (%)	دقت (%)
Proposed	0.3	45.11	96.68
GVBS	0.9	58.94	96.11
ITTI	0.9	63.76	86.63
BMS	1.0	94.33	96.67
FES	1.0	49.99	88.53
JUDD	0.5	38.20	72.36
CovSal	0.98	50.59	93.98

مدل	حرفه‌ای		
	دقت (%)	مساحت مهم (%)	آستانه
Proposed	73.36	25.47	0.3
GVBS	42.15	40.07	0.8
ITTI	40.45	29.81	0.7
BMS	55.03	35.65	0.9
FES	54.21	49.81	1.0
JUDD	37.68	36.76	0.5
CovSal	15.47	25.92	0.9

جدول ۳-۲۱ مقدار دقت انواع مدل‌های بصری برای هر یک از سطوح مهارت را بیان می‌کند. برای توضیح بیشتر، نحوه محاسبه دقت مدل GBVS برای بازیکن‌های مبتدی را شرح می‌دهیم. برای سایر مدل‌ها و سطح مهارت‌ها محاسبه به روش مشابهی انجام می‌شود. برای هر بازیکن مبتدی، GBVS را بر روی تمام فریم‌های ویدئو وی اعمال می‌کنیم. سپس نقشه‌های بدست آمده را با اعمال یک مقدار قطع به نقشه‌های دودویی تبدیل می‌کنیم. سپس متوسط مساحت نواحی پر اهمیت در بین این نقشه‌ها را حساب می‌کنیم. سپس متوسط این مقادیر بین تمام بازیکن‌های مبتدی را محاسبه کرده و آن را درصد مساحت پراهمیت^۱ (SAP) می‌نامیم. این روال را برای مقادیر قطع ۰,۱، ۰,۲، ۰,۳، ...، ۰,۵ و ۱,۰ تکرار کردیم تا مقدار SAP مساوی یا بیشتر از ۶۷,۶٪ شود. برای GBVS مقدار قطع مناسب ۰,۵ بود. بعد از مشخص شدن مقدار قطع و نقشه‌های دودویی درصد نقاط خیرگی که در نقاط پراهمیت نقشه‌ها می‌افتند را محاسبه می‌کنیم. طبق جدول ۳-۲۱، با مقدار قطع ۰,۵ خروجی GBVS منجر به SAP معادل ۸,۵۶٪ خواهد شد. با اینکه این مقدار از ۶۷,۶٪ بیشتر است، دقت آن از مدل پیشنهادی کمتر است.

برای مقایسه بیشتر مدل پیشنهادی، با استفاده از معیار NSS^۲ نیز عملکرد آن را با سایر مدل‌ها مقایسه کردیم. بطور کلی، این معیار ابتدا امتیازهای بصری (مقادیر نقشه‌های بصری) را نرمال می‌کند بطوریکه میانگین اعداد صفر و انحراف معیار آنها یک شود. سپس میانگین امتیاز مربوط به نقاط خیرگی را در نقشه نرمال شده به عنوان مقدار NSS معرفی می‌کند [۷۶]. در واقع NSS انحراف از میانگین در نقاط خیرگی است. اعداد بزرگتر احتمال کمتری برای نمایش در نتایج تجربی دارند. در واقع زمانی اعداد بزرگ دیده خواهند شد که نقشه بصری، نقطه خیرگی را به خوبی پیش‌بینی کند. با توجه به توضیحات داده شده NSS به صورت زیر بدست می‌آید [۷۶]:

$$\mu = \frac{1}{N} \sum_{n=1}^N s_n \quad (1-3)$$

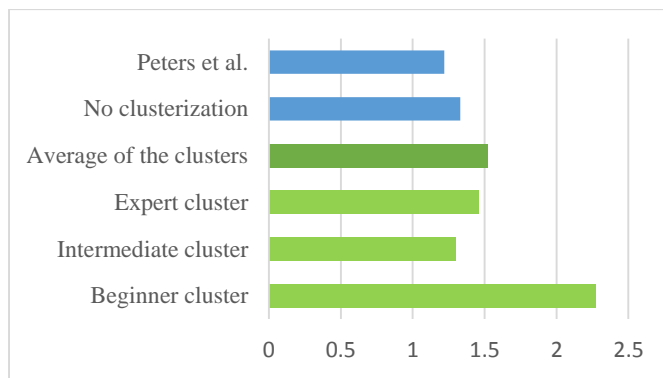
^۱ Salient Area Percentage (SAP)

^۲ Normalized Scanpath Saliency (NSS)

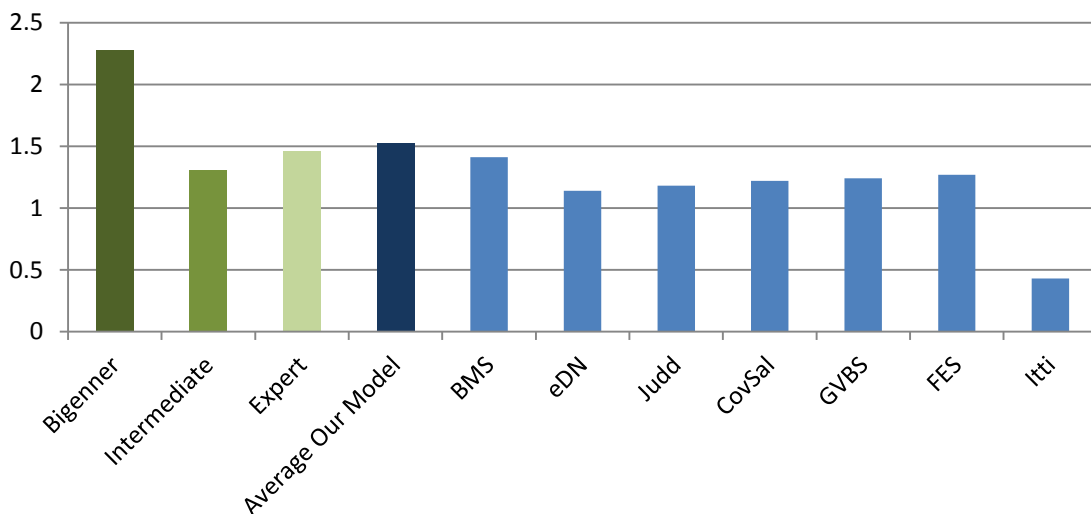
$$\mu = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (s_n + \mu)^2} \quad (2-3)$$

$$NSS = \frac{s_e - \mu}{\sigma} \quad (3-3)$$

در این رابطه‌ها s_n و s_e به ترتیب امتیاز بصری و نقطه‌ی خیرگی در مکان n ام هستند. N تعداد مکان‌های کاندید نقطه خیرگی در صحنه است. $NSS=1$ به این معنی است که نقطه خیرگی کاربر، در ناحیه‌ای از نقشه‌ی بصری قرار گرفته است که به اندازه یک انحراف معیار بالاتر از میانگین امتیازات نقشه است.



شکل ۲۲-۳ - مقایسه NSS با و بدون در نظر گرفتن مهارت



شکل ۲۳-۳ - مقدار NSS برای مدل‌های توجه بصری و مدل پیشنهادی

شکل ۳-۲۲ مقدار NSS برای سطوح مختلف مهارت را نشان می‌دهد. طبق این شکل، مقدار NSS در حالتی که مهارت را در نظر می‌گیریم، نسبت به حالتی که از خوشه‌بندی استفاده نمی‌کنیم، ۱۴٫۲۸٪ بیشتر است. مقدار کمتر NSS برای دسته متوسط نسبت به دو دسته دیگر را می‌توان به تنوع بیشتر الگوها در میان آنها نسبت داد. بر خلاف حرفه‌ای‌ها که بر روی قسمت‌های پر امتیاز صحنه تمرکز می‌کنند، و مبتدی‌ها که وقت زیادی روی بخش‌های محدودی از صحنه می‌گذارند، بازیکن‌های با مهارت متوسط مهارت و کنجکاوای لازم برای سرک کشیدن به نقاط مختلف صفحه را دارند. بنابراین ناحیه‌های بیشتر و وسیعتری را نگاه می‌کنند. این باعث می‌شود مساحت نقشه توجه آنها بزرگ‌تر شود. مساحت بزرگ‌تر خود باعث افزایش مثبت-غلط^۱ها می‌شود که توسط NSS جریمه می‌شوند و باعث کاهش مقدار آن می‌گردند. در شکل ۳-۲۲ همچنین مقدار NSS روش ارائه شده در [۴۶] که با توجه به جدول ۳-۸ نزدیکترین روش به کار ما است، گزارش شده است. مدل پیشنهادی نسبت به این مدل ۲۴٫۵۹٪ مقدار بیشتری NSS کسب کرده است.

شکل ۳-۲۳ مقدار NSS مدل پیشنهادی با سایر مدل‌ها را مقایسه می‌کند. از آنجایی که مدل‌های پیشنهادی در حوزه بازی ارائه نشده‌اند، لذا برای مقایسه عادلانه‌تر، نتایج حاصله از آنها را در محتوایی که مدل برای آن محتوا ارائه شده است، مقایسه کردیم. در صورت مقایسه این مدل‌ها در محتوای بازی اعداد به شدت کاهش می‌یابند و حتی در بعضی از آنها مقدار NSS منفی خواهد شد. مقادیر NSS روش‌های ذکر شده از محک Judd استخراج شده است.

۳-۲-۵- بحث

روش‌شناسی استفاده شده در این رساله برای توسعه مدل توجه مبتنی بر مهارت، برای سایر بازی‌ها نیز تعمیم‌پذیر است. البته هر کدام از گام‌های آن ممکن است بازی به بازی متفاوت باشد. بنابراین در استفاده از آن باید چند نکته را در نظر گرفت.

جمع‌آوری داده‌های ردیاب چشم پرهزینه و وقت‌گیر است. بنابراین، استفاده از مدل باید توجیه اقتصادی داشته باشد. از یک طرف، هر چه یک بازی صحنه‌ها و موقعیت‌های بیشتری داشته باشد، داده‌های بیشتری باید جمع گردد و بنابراین هزینه‌ها افزایش می‌یابد. از طرف دیگر، کاهش نرخ بیت برای آن ویدئو می‌تواند به صورت نمایی باعث افزایش مشتریان و درآمد شود. بنابراین، ارائه دهندگان سرویس باید یک تحلیل هزینه-فایده انجام دهند. به عنوان مثال، بطور متوسط هفت ساعت طول می‌کشد تا داستان اصلی بازی Call of Duty: Black Ops جدیدترین نسخه از سری Call of Duty، یکی از محبوب‌ترین بازی‌های سبک شلیک‌گر اول [۷۷]، را به سرانجام رساند [۷۸]. به این ترتیب، با فرض داشتن ده دستگاه ردیاب چشم و اینکه بر روی هر دستگاه در هر روز یک جلسه اجرا می‌شود، جمع‌آوری دویست جلسه برای این بازی بیست روز به طول خواهد انجامید.

^۱ False positive

دیتاست باید تا جای ممکن موقعیت‌های مختلف بازی را پوشش دهد. منظور از موقعیت بازی، ترکیبی از اشیا بازی، حالت‌ها و صحنه‌ها است. هر چه موقعیت بیشتری پوشش داده شود، دقت مدل بالاتر می‌رود. اگر موقعیت‌های بازی به اندازه کافی از هم قابل تمیز باشند، گام دوم روش‌شناسی (خوشه‌بندی) اجازه می‌دهد الگوی آنها را جداگانه در بیاوریم. در غیر اینصورت تاثیر آنها در گام سوم مدیریت خواهد شد جایی که احتمال نگاه کردن بازیکن‌ها به نواحی مختلف محاسبه می‌شود. اگر یک موقعیت طولانی مدت قابل تمیز در دیتاست پوشش داده نشود، دقت مدل پایین خواهد آمد. به هر حال، این مسئله برای الگوریتم‌های یادگیر صادق است.

برای بدست آوردن NSS بالا، مدل پیشنهادی باید برای هر بازی بطور جدا ساخته شود. زیرا هر بازی منطق و طراحی یکتای خودش را دارد. این دو عامل اهداف بازی را تعیین می‌کنند و اهداف بازی توجه بالا-به-پایین بازیکن‌ها را تحت تاثیر قرار می‌دهد. بنابراین دیتاستی که برای یک بازی تهیه شده است، لزوماً برای بازی دیگر مناسب نیست. با این وجود، ما معتقدیم ممکن است به ما کمک کند برای بازی‌های هم‌سبک، مانند سبک ماشین‌های مسابقه، الگوهای توجه مشابهی بدست آوریم. بطوریکه دیتاست یک بازی با کمی تغییرات برای بازی دیگری در همان سبک قابل استفاده باشد.

مدل پیشنهادی فرض می‌کند الگوی توجه افرادی که سطح مهارت یکسانی دارند، تقریباً مشابه هم هست. با این وجود، عوامل دیگری مانند سبک و عادت بازیکن‌ها نیز بر روی توجه آنها تاثیرگذار است. اما از آنجا که رسیدن به اهداف بازی حرکات و نکات خاصی از بازیکن‌ها می‌طلبد، درجه آزادی توجه آنها محدود می‌شود. این مورد به این معنی است که اگرچه بازیکن‌های هم‌سطح الگوهای دقیقاً یکسانی ندارند، اما الگوهایشان به اندازه کافی به هم شبیه است که در یک گروه خوشه‌بندی شوند. گام دوم روش‌شناسی دقیقاً همین کار را انجام می‌دهد و همانطور که نتایج آزمایش‌ها نشان می‌دهد دقت بهتری بدست می‌آید. برای ادامه تحقیقات در این زمینه، باید داده‌های ردیاب چشم از بازی‌ها و موقعیت‌های بیشتری جمع‌آوری شود. اگرچه این کار بسیار طاقت‌فرسا است. داده‌های جمع‌آوری شده در این رساله در دسترس عمومی قرار می‌گیرد تا اهرمی برای تحقیقات آینده در این زمینه باشد و یک پایه مناسب برای مقایسه الگوریتم‌های مشابه بوجود آید.

نکته دیگر نحوه پیش‌بینی سطح مهارت یک بازیکن یا به عبارت دقیق‌تر خوشه متناظر وی می‌باشد. در این رساله، ما از امتیاز استفاده کردیم زیرا امتیاز و مهارت همبستگی بالایی دارند. تحقیقات نشان داده است که امتیاز از مهارت، تاخیر شبکه و لغزش^۱ شبکه تاثیر می‌پذیرد [۷۹]. این تحقیق نشان می‌دهد بیشترین تاثیر امتیاز در بین اکثر سبک‌های بازی از مهارت است. از آنجا که در آزمایش‌های ما نه تنها تاخیر و لغزش شبکه بلکه سخت افزار و نرم افزار (قدرت پردازش، گرافیک، دستگاه‌های ورودی و ...) برای همه بازیکن‌ها یکسان بوده است، با اطمینان امتیاز به عنوان عاملی مناسب برای تشخیص مهارت استفاده کنیم. البته هر بازی محتوا و نکات خاص خود را دارد. در آینده قصد داریم، تاثیر سایر عوامل از جمله ورودی‌های کاربر در دقت پیش‌بینی مهارت را بسنجیم.

^۱ Jitter

۳-۳- جمع‌بندی

در این فصل نشان داده شد که مدل‌های توجه بصری فعلی با توجه به اینکه برای بازی‌های رایانه‌ای طراحی نشده‌اند، دقت پایینی در پیش‌بینی توجه بازیکن‌ها، به ویژه توجه بالا-به-پایین آنها، دارند. به همین دلیل، در وهله اول، یک مدل توجه بصری مخصوص بازی ارائه گردید. سپس نشان داده شد که مهارت بازیکن عامل مهمی در الگوی توجه وی می‌باشد. این مشاهده با فرض مدل‌های توجه بصری معمول که الگوی توجه میان بینندگان را تقریباً یکسان می‌دانند، در تضاد است. بنابراین، با توجه به اینکه دیتاست‌های مرتبط با بازی نمونه‌های کمی داشتند و تفاوت الگوهای توجه در آنها مشخص نمی‌شد، یک دیتاست بزرگ که مهارت بازیکن‌های مختلف در آن هویدا باشد، جمع‌آوری شد و تاثیر مهارت در الگوی بازیکن‌ها با سطوح مختلف مهارت بررسی گردید. نتایج آزمایش‌ها نشان می‌دهد در نظر گرفتن مهارت در مدل توجه بصری دقت آن را افزایش می‌دهد. در نهایت با قرار دادن هر یک از مدل‌های پیشنهادی در کدکننده H.264/AVC نشان داده شد که می‌توان با حفظ کیفیت ادراکی کاربر، نرخ بیت ویدئوهای بازی را کاهش داد. برای به‌کارگیری عامل مهارت در مدل توجه بازی، می‌توان مدل تاثیر مهارت در توجه بازیکن به اشیای مختلف بازی را بدست آورد و جدول ضرایب اهمیت در پیمان‌ه بالا-به-پایین مدل توجه بازی را گسترش داد.

فصل چهارم

مدل نرخ-اعوجاج پیشنهادی

مدل‌های ادراکی پیشنهادی، اهمیت نواحی مختلف ویدئو از دید بازیکن را مشخص می‌کنند. اگرچه از خروجی این مدل‌ها می‌توان در بهبود عملکرد کدکردن ویدئو، جوینبارسازی، اولویت‌بندی بسته‌های شبکه و سایر کاربردها استفاده کرد، با این وجود، تمرکز این رساله بر روی اعمال این مدل‌ها در کدکننده ویدئو می‌باشد. در این فصل، ابتدا بر اساس آزمایش‌های کیفی یک مدل نرخ-اعوجاج مبتنی بر توجه و گودی مرکزی شبکه استخراج می‌کنیم. سپس، اختصاص بیت را با استفاده از این مدل و در قالب یک مسئله بهینه‌سازی لاگرانژ حل می‌کنیم. همانطور که در فصل پیشینه تحقیق مطرح شد، از بین روش‌هایی که مدل‌های ادراکی را در کدکننده اعمال می‌کنند، آن‌هایی که پارامترهای کدکننده را تنظیم می‌کنند ارجحیت بیشتری دارند، زیرا استفاده از آن‌ها در کاربردهای عملی راحت‌تر و کیفیت ویدئوی تولید شده توسط آن‌ها بیشتر است. از بین همین روش‌ها نیز آن‌هایی که فرمول‌های بسته ارائه می‌کنند، به دلیل تعمیم‌پذیری بالاتر، از محبوبیت بیشتری برخوردارند. مدل پیشنهادی ارائه شده نیز در همین دسته قرار می‌گیرد. در ادامه این فصل، نحوه رسیدن به مدل و استفاده از آن در مسئله اختصاص بیت ارائه می‌گردد.

۴-۱- ساخت مدل

تابع نرخ-اعوجاج شناخته شده برای یک فرایند تصادفی گسسته مانای گوسی $\{x(n)\}$ با اعوجاج میانگین مربع خطا $[.80, .81]$ برابر است با:

$$R_\theta = \frac{1}{4\pi} \int_{A_1} \log_2 \frac{\Phi(\omega)}{\theta} d\omega \quad (۱-۴)$$

$$D(R_\theta) = \frac{1}{2\pi} \int_{A_1} \theta d\omega + \frac{1}{2\pi} \int_{B_1} \Phi(\omega) d\omega \quad (۲-۴)$$

که در آن $\theta \geq 0$ و $\Phi(\omega)$ تابع چگالی طیف توان $\{x(n)\}$ است و

$$\begin{cases} \text{Region } A_1: \{\omega \in (-\pi, \pi] \text{ and } \Phi(\omega) \geq \theta\} \\ \text{Region } B_1: (-\pi, \pi] - A_1 \end{cases} \quad (۳-۴)$$

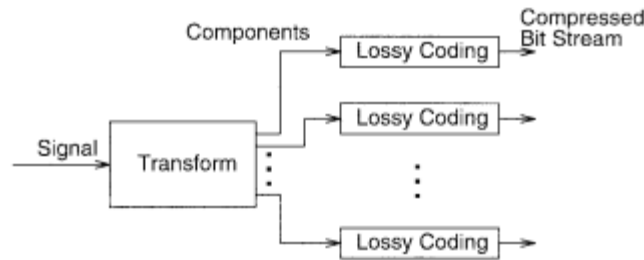
یک تفسیر از فرمول بالا آن است که R_θ کمینه بیت لازم برای رسیدن به متوسط اعوجاج D توسط یک کدکننده، بدون قید روی پیچیدگی و زمان پردازش، است. به طور مشابه، $D(R_\theta)$ کمینه اعوجاجی است که می‌توان در نرخ بیت R به آن رسید. در این روابط θ یک پارامتر اضافه است. طبق این روابط بهترین عملکرد کدکردن وقتی حاصل می‌شود که مولفه‌های فرکانسی که توان آن‌ها کمتر از θ است، حذف و بر روی بقیه مولفه‌ها اعوجاجی به اندازه θ اعمال شود.

در واقعیت، امکان استفاده از تبدیل‌های نامتناهی برای جداسازی یک سیگنال وجود ندارد. یک راه معمول، تقسیم‌بندی سیگنال به بخش‌های غیرهمپوشان و اجرای تبدیل بر روی هر بخش بصورت جداگانه است. در بکارگیری این روش دو نکته وجود دارد. اول، همبستگی میان بخش‌های همسایه است. این همبستگی در فرکانس‌های پایین بالاتر نیز می‌رود. مقداری از این همبستگی‌ها با اعمال فن‌های کاهش همبستگی روی بخش‌های همسایه قابل کاهش است. دوم، طیف توان موجود در روابط است که در عمل باید از روی نمونه داده‌ها تخمین زده شود. یک روش ساده و محبوب برای تخمین طیف توان سیگنال روش peridogram است که طیف را بر اساس میانگین وزن‌دار تبدیل‌های فوریه بخش‌های غیرهمپوشان محاسبه می‌کند. اگر مولفه‌های تبدیل بخش‌ها را تقریبی گسسته از طیف توان پیوسته ایده‌آل در نظر بگیریم، این روش با تبدیل متناهی که در روش‌های فشرده‌سازی استفاده می‌شود، سازگار است. با فرض نمونه‌گیری یکنواخت از دامنه فرکانسی، روابط گسسته عبارتند از:

$$R_\theta = \frac{1}{2L} \sum_{\omega_i \in A_1} \log_2 \frac{\Phi(\omega_i)}{\theta} \quad (۴-۴)$$

$$D(R_\theta) = \frac{1}{L} \sum_{\omega_i \in A_1} \theta + \frac{1}{L} \sum_{\omega_i \in B_1} \Phi(\omega_i) \quad (۵-۴)$$

که در آن L تعداد نمونه‌ها در یک بخش از داده و $w_i = i \cdot \frac{2\pi}{L}, i = 0, 1, \dots, L-1$ است. این سیستم در شکل نشان داده شده است.



شکل ۱-۴ - جداسازی سیگنال توسط تبدیل

رابطه نرخ بیت را اینگونه نیز می‌توان نوشت:

$$e^{L \cdot \alpha \cdot R \theta} = \prod_{\omega_i \in A_1} \frac{\Phi(\omega_i)}{\theta} \quad (۶-۴)$$

که در آن مقدار α برابر ۱,۳۸۶ است. در واقع، تابع نمایی نرخ بیت با ضرب واریانس مولفه‌های سیگنال تناسب دارد. در اعوجاج‌های پایین، ناحیه B_1 خالی می‌شود و در نتیجه $D = \theta$. بنابراین:

$$e^{L \cdot \alpha \cdot R \theta} = \prod_{i=0}^{L-1} \frac{\Phi(\omega_i)}{D} \quad (۷-۴)$$

یا به عبارت دیگر

$$D = E \cdot e^{-\alpha \cdot R(D)} \quad (۸-۴)$$

که در آن

$$E = \left(\prod_{i=0}^{L-1} \Phi(\omega_i) \right)^{1/L} \quad (۹-۴)$$

در واقع، یک منبع گوسی توام با ضرب چند منبع گوسی i.i.d تخمین زده می‌شود. E ضرب واریانس تمام مولفه‌های سیگنال است و بی‌نظمی و ردایی^۱ آن سیگنال نامیده می‌شود. بی‌نظمی سیگنال نماینده پیچیدگی کد کردن آن می‌باشد. در تئوری، دو سیگنال با واریانس معمولی برابر تعداد بیت‌های مختلفی احتیاج دارند، اگر بی‌نظمی واریانس متفاوتی داشته باشند.

ضعف رابطه (۸-۴) عدم توجه به ویژگی‌های چشم انسان است. در واقع، سیستم بینایی چشم انسان بسیار غیرخطی است و نمی‌توان آن را تنها با میانگین مربعات خطا تخمین زد. به عنوان مثال، تحمل خطا در نواحی مختلف برای چشم انسان یکسان نیست. برای در نظر گرفتن این مسئله، اعوجاج ناحیه مشخصی از ویدئو، t ، در مقالات به شکل دیگری بازنویسی شد [۵۶، ۶۸]:

^۱ Entropy Variance

$$D'_i = w_i \cdot D_i \quad (10-4)$$

که در آن w_i وزن نسبت داده شده به ناحیه i در نقشه توجه و D_i اعوجاج مبتنی بر میانگین مربعات خطا است. در این رساله، قصد داریم این رابطه را گسترش دهیم تا ویژگی دیگری از سیستم بینایی چشم به نام گودی مرکزی شبکه که در فصل پیشینه تحقیق به آن اشاره شد را نیز در نظر بگیریم. البته روشی که از آن استفاده می‌کنیم برای بررسی تاثیر سایر ویژگی‌های چشم نیز می‌تواند مورد استفاده قرار گیرد. در واقع هدف، پیدا کردن تابع گسترش‌یافته $f(\cdot)$ برای وزن‌دهی به اعوجاج مبتنی بر میانگین مربعات خطا می‌باشد:

$$D''_i = f(\cdot) \cdot D_i \quad (11-4)$$

نحوه بدست آوردن تابع $f(\cdot)$ در ادامه این فصل تشریح خواهد شد. اما از آنجا که این تابع از دل مسئله اختصاص بیت بدست می‌آید، ابتدا به نحوه‌ی استفاده از مدل نرخ-اعوجاج رابطه (11-4) در اختصاص بیت به نواحی مختلف ویدئو پرداخته می‌شود.

اختصاص بیت را می‌توان در قالب یک مسئله بهینه‌سازی فرمول‌بندی کرد:

$$\begin{cases} \min D(R_1, R_2, \dots, R_N) \\ s. t. \sum_{i=1}^N \frac{S_i \cdot R_i}{S} \leq R \end{cases} \quad (12-4)$$

که در آن S مساحت کل فریم، S_i مساحت ناحیه i ام، N تعداد نواحی در یک فریم، R_i نرخ بیت (بیت بر پیکسل) استفاده شده برای کدکردن ناحیه i ام و R نرخ بیت هدف کل فریم می‌باشد. اعوجاج فریم نیز این گونه تعریف می‌شود [68]:

$$D(R_1, R_2, \dots, R_N) = \sum_{i=1}^N \frac{S_i}{S} D_i(R_i) \quad (13-4)$$

مسئله قیددار رابطه (12-4) را می‌توان با استفاده از روش بهینه‌سازی لاگرانژ به یک مسئله بی‌قید تبدیل کرد. طبق این روش، ابتدا باید تابع لاگرانژ را تعریف کنیم:

$$g(R_1, R_2, \dots, R_N) = D(R_1, R_2, \dots, R_N) - \lambda \left(R - \frac{S_1 \cdot R_1 + S_2 \cdot R_2 + \dots + S_N \cdot R_N}{S} \right) \quad (14-4)$$

برای کمینه‌کردن $g(\cdot)$ باید داشته باشیم:

$$\frac{dg}{dR_1} = \frac{dg}{dR_2} = \dots = \frac{dg}{dR_N} = \frac{dg}{d\lambda} = 0 \quad (15-4)$$

با حل این رابطه مقدار نرخ بیت هر ناحیه بدست می‌آید:

$$R_i = R + \frac{1}{\alpha N S} \left(\sum_{j \neq i} S_j \cdot \ln \frac{E_i}{E_j} + \sum_{j \neq i} S_j \cdot \ln \frac{f_i(\cdot)}{f_j(\cdot)} \right) \quad (16-4)$$

برای بدست آوردن مدل نرخ-اعوجاج گسترش یافته، یا بطور دقیق تر تابع $f(\cdot)$ ، روال نمایش داده شده در شکل ۲-۴ را طی می کنیم.



شکل ۲-۴ - روال یافتن تابع f

۴-۱-۱- تهیه مجموعه داده

گام نخست این روال، تهیه مجموعه‌ای از ویدئوها می باشد. در اینجا، از ویدئوهای جمع‌آوری شده برای توسعه مدل توجه بازی استفاده کردیم زیرا نقشه‌های توجه این ویدئوها در دسترس بودند.

۴-۱-۲- بدست آوردن نقشه توجه

با توجه به اینکه تمرکز رساله بر روی کاربرد بازی‌های مبتنی بر رایانش ابری است، از مدل توجه مخصوص همین کاربرد باید برای ویدئوهای بازی استفاده شود. بنابراین، در گام دوم مدل توجه بازی را بر روی ویدئوها اعمال می کنیم. نقشه‌های خروجی مدل توجه بازی می توانند شامل هر تعداد سطح از توجه باشند. با این حال، در شبیه‌سازی‌های مربوط به ارزیابی این مدل، سه سطح در نظر گرفته شد. در اینجا نیز از سه سطح استفاده می کنیم. این سه سطح، در واقع، ویدئو را از نظر اهمیت از دیدگاه بازیکن به سه ناحیه تقسیم می کنند.

۴-۱-۳- اختصاص بیت بر اساس توجه

در گام سوم، به هر یک از سه ناحیه‌ی موجود در هر ویدئو طبق مدل نرخ-اعوجاج مبتنی بر توجه پیشنهاد شده در [۵۶]، یعنی بر اساس رابطه (۴-۱۰)، بیت اختصاص داده می شود.

۴-۱-۴- بدست آوردن کیفیت ادراکی بهتر

در گام چهارم، مقدار بیت اختصاص داده شده به هر ناحیه را با تغییر مقدار QP برای آن ناحیه، تغییر می‌دهیم، بطوریکه نرخ بیت کل ویدئو تغییر نیابد. این گام نیازمند تعداد زیادی آزمایش است. زیرا مجموعه QP جدید یافته شده نه تنها باید منجر به نرخ بیت کل برابر گردد، بلکه باید از نظر کیفیت ادراکی نیز بهتر باشد.

جدول ۱-۴- بیت‌های اختصاص داده شده به هر ناحیه بر اساس مدل پیشین

ویدئوی بازی	نرخ بیت		
	ناحیه ۱	ناحیه ۲	ناحیه ۳
A3	917611	606575	279929
A4	202094	275763	458760
A5	375053	583685	1692404
A6	2028578	806352	2372343
A7	114479	183465	1244778
A8	152926	118034	236470
A9	30933	446690	184487
A10	105956	2538502	77937
A11	963003	843180	1157476

جدول ۱-۴ نرخ بیت اختصاص داده شده به نواحی مختلف ویدئوها را نشان می‌دهد. بیت کل هر یک از ویدئوها در گام سه، برابر با بیت کل همان ویدئو در آزمایش‌های ارزیابی کیفی مدل توجه بازی که با مجموعه QP های ۳۰، ۳۲ و ۳۴ به ترتیب برای نواحی پراهمیت، اهمیت متوسط و کم‌اهمیت بدست آمده بود، قرار داده شد.

برای اطمینان از نتایج بدست آمده در گام سه، یک آزمون ارزیابی کیفی بر روی ویدئوهایی که در این گام انتخاب می‌شوند، انجام می‌دهیم. برای ارزیابی کیفی، همانند ارزیابی مدل توجه بازی، از روش Double Stimulus Continuous Quality Scale تشریح شده در توصیه‌نامه ITU-R Recommendation 500 (ITU-R 1974-1997) استفاده گردید. هفده نفر در این ارزیابی شرکت کردند. جدول ۲-۴ اطلاعات دموگرافیک شرکت‌کنندگان را نشان می‌دهد.

هر یک از ویدئوهای موجود در ارزیابی توسط هر یک از شرکت‌کنندگان دیده شد. این روال حدود پانزده دقیقه از وقت آنها را گرفت. ویدئوها با دقت مکانی اصلی خود نشان داده شدند تا هیچ اعوجاجی ناشی از تغییر مقیاس وجود نداشته باشد. فاصله شرکت‌کنندگان تا صفحه نمایشگر طبق توصیه Rec 812 ITU-R، چهار برابر ارتفاع نمایشگر تنظیم شد. ویدئوها با ترتیب تصادفی و سه ثانیه صفحه خاکستری میانشان به شرکت‌کنندگان نشان داده شد و آنها کیفیت این ویدئوها را با مشخص کردن عددی بین ۱ تا ۵ تعیین کردند. قبل از شروع ارزیابی به آنها گفته شد که ارزیابی آنها می‌تواند یک عدد حقیقی در این بازه باشد. در شروع هر جلسه ارزیابی، پنج نمایش خارج از برنامه قرار داده شد. اولین نمایش برای آشنایی شرکت‌کننده با روال ارزیابی و چهار نمایش بعدی برای پایدارسازی نظر

وی. سطوح کیفیت نیز مطابق آزمون ارزیابی کیفی مدل توجه بازی، طبق جدول ۳-۶ در نظر گرفته شد.

جدول ۲-۴- اطلاعات دموگرافیک شرکت کنندگان

مهارت در بازی‌های رایانه‌ای از نظر کاربر				
عالی	خوب	متوسط	ضعیف	خیلی ضعیف
۱۱,۷٪	۱۲٪	۴۱,۱۸٪	۳۵,۳٪	۰٪
میزان ساعت بازی در ماه				
بیش از ۳۰	۲۱-۳۰	۱۱-۲۰	۶-۱۰	کمتر از ۵
۱۸٪	۰٪	۱۷,۶۴٪	۵,۸۸٪	۶۴,۷٪
سکوپی که تاکنون کاربر بر روی آن بازی کرده است				
گوشی همراه	تبلت	کنسول	کامپیوتر	
۵۸,۸۲٪	۱۷,۶۵٪	۵,۸۸٪	۷۶,۴٪	
سبک بازی که کاربر تاکنون بازی کرده است				
ورزشی	ماجراجویی	شلیکگر اول شخص		
۶۴,۷۱٪	۴۷,۰۵٪	۵۸,۸۲٪		
شبیه‌ساز پرواز	جنگی	مسابقه‌ای		
۲۹,۴۱٪	۱۱,۷۶٪	۷۶,۴۷٪		

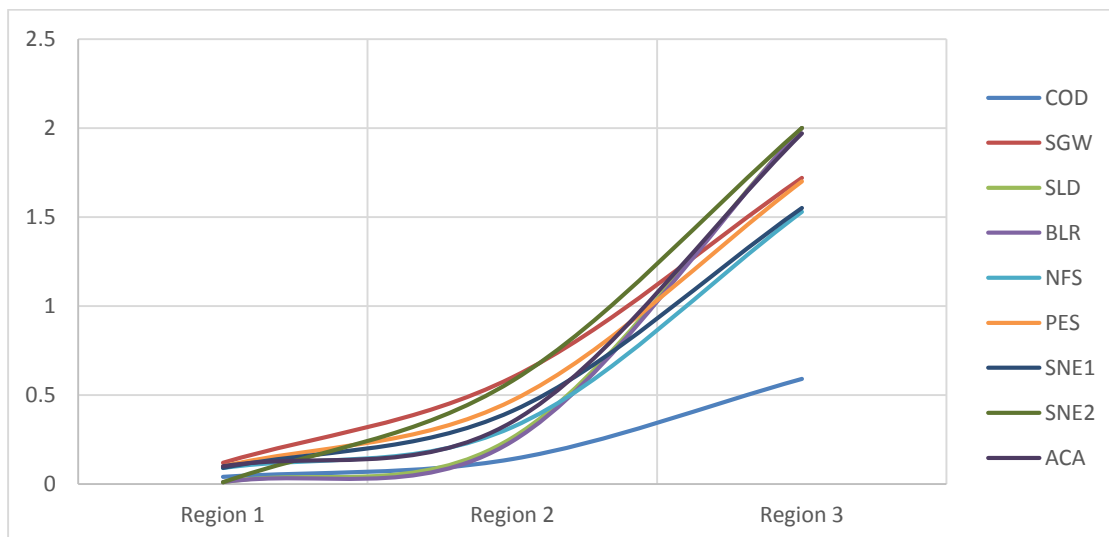
جدول ۳-۴- نتایج آزمایش کیفی برای دو سناریو

	متوسط امتیاز نظر									
	COD	SGW	SLD	BLR	NFS	PES	SNE1	SNE2	ACA	تغییر
مدل [56]	۳,۳۱	۲,۹۵	۳,۳۱	۳,۹۷	۳,۳۱	۲,۸۱	۲,۶۸	۳,۰۶	۴,۳۷	-
ویدئوهای جدید	۳,۶۸	۴,۳۵	۳,۹۶	۳,۹۷	۴,۰۶	۳,۸	۴,۳۳	۴,۲۲	۴,۲۵	۲۵,۹٪

جدول ۳-۴ نتایج آزمایش‌ها برای دو سناریو را نشان می‌دهد. در یک سناریو اختصاص بیت بر اساس مدل نرخ-اعوجاج مبتنی بر توجه پیشین و در سناریوی دیگر اختصاص بیت بصورت دستی ولی با شرایط ذکر شده در بالا انجام گرفت. همانطور که از نتایج بر می‌آید، ویدئوهای جدید در نرخ برابر ۲۵,۹٪ امتیاز بیشتری از شرکت کنندگان دریافت کرده‌اند.

۴-۱-۵- استخراج تابع f

گام پنجم، گام آخر برای بدست آوردن مقدار $f(0)$ می‌باشد. به همین منظور نرخ بیت‌های نواحی مختلف ویدئوهای بدست آمده در گام چهارم را در رابطه (۴-۱۶) قرار می‌دهیم و مقادیر $f(0)$ را به طور معکوس از این رابطه استخراج می‌کنیم. جدول ۴-۴ مقادیر تابع $f(0)$ برای ویدئوهای مختلف را نشان می‌دهد. شکل ۳-۴ همین مقادیر را به شکل نمودار نشان می‌دهد.



شکل ۴-۳ - مقادیر تابع f حاصل از ویدئوهای گام چهارم

جدول ۴-۴ - مقادیر تابع f برای هر ناحیه

	COD	SGW	SLD	BLR	NFS	PES	SNE1	SNE2	ACA
ناحیه ۱	0.04	0.12	0.01	0.01	0.09	0.1	0.09	0.01	0.1
ناحیه ۲	0.14	0.6	0.26	0.24	0.32	0.47	0.41	0.58	0.35
ناحیه ۳	0.59	1.72	2	2	1.53	1.7	1.55	2	1.97

جدول ۴-۵ - بیت‌های اختصاص داده شده به هر ناحیه در گام چهارم

ویدئوی بازی	نرخ بیت		
	ناحیه ۱	ناحیه ۲	ناحیه ۳
A3	649898	704086	377249
A4	511829	225000	328032
A5	856288	481077	1267553
A6	2403402	804958	1872778
A7	289638	299874	931167
A8	276881	68240	135316
A9	93838	448668	90730
A10	590703	1930818	97997
A11	962924	713556	1261260

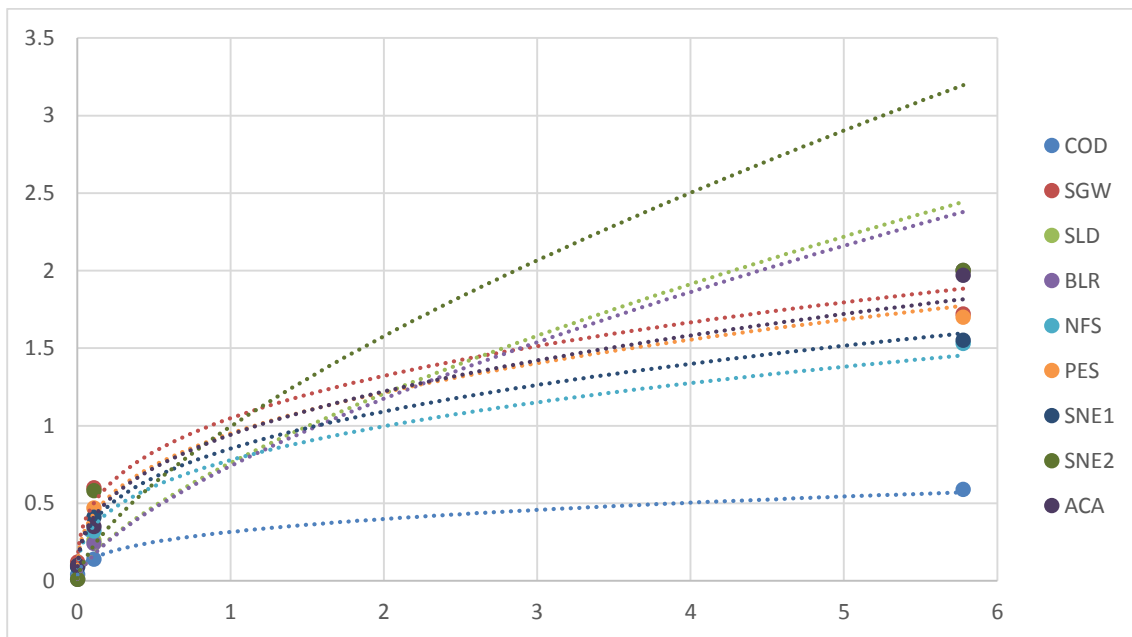
بیت‌های که منجر به رسیدن این مقادیر از تابع f شده‌اند، در نشان داده شده است. از آنجا که تعداد سطوح توجه برای همه ویدئوها یکسان است، وزن‌های توجه نیز یکسان می‌باشند. توجه داشته باشید که از آنجا که وزن توجه به نوعی نماینده تضعیف انرژی سیگنال است، و انرژی سیگنال تابعی لگاریتمی از سطح سیگنال است، در [۶۸] پیشنهاد شده است که سطوح توجه از طریق رابطه (۴-۱۷) به وزن ترجمه شوند:

$$w_i = C \cdot e^{\frac{A_i}{k}} \quad \text{رابطه (۴-۱۷)}$$

که در آن A_i سطح توجه ناحیه i ، w_i وزن توجه ناحیه i و C و k نیز ضریب ثابت هستند. در این رساله ما نیز از همین رابطه استفاده کردیم و به پیروی از پیشنهاددهندگان مقادیر C و k را به ترتیب ۰,۰۰۲ و ۳۲ قرار دادیم. در جدول ۴-۶ وزن هر یک از نواحی مشخص شده است.

جدول ۴-۶- وزن توجه نواحی مختلف

	وزن توجه
ناحیه ۱	0.0020
ناحیه ۲	0.1092
ناحیه ۳	5.7785



شکل ۴-۴- رابطه مقادیر تابع f و وزن‌های توجه در ویدئوها

اکنون به بررسی مقادیر تابع f و رابطه آن‌ها با وزن‌های توجه می‌پردازیم. شکل ۴-۴ منحنی برازش شده بر روی این مقادیر را نشان می‌دهد. انواع منحنی‌ها، از جمله منحنی‌های خطی، لگاریتمی، توانی و نمایی بر روی نمودار هر ویدئو برازش شد. مقدار R^2 منحنی‌های برازش شده در جدول ۴-۷ نمایش داده شده است. همانطور که مشخص است، رابطه مقادیر تابع f و وزن‌های توجه در این دیتاست از نوع توانی است.

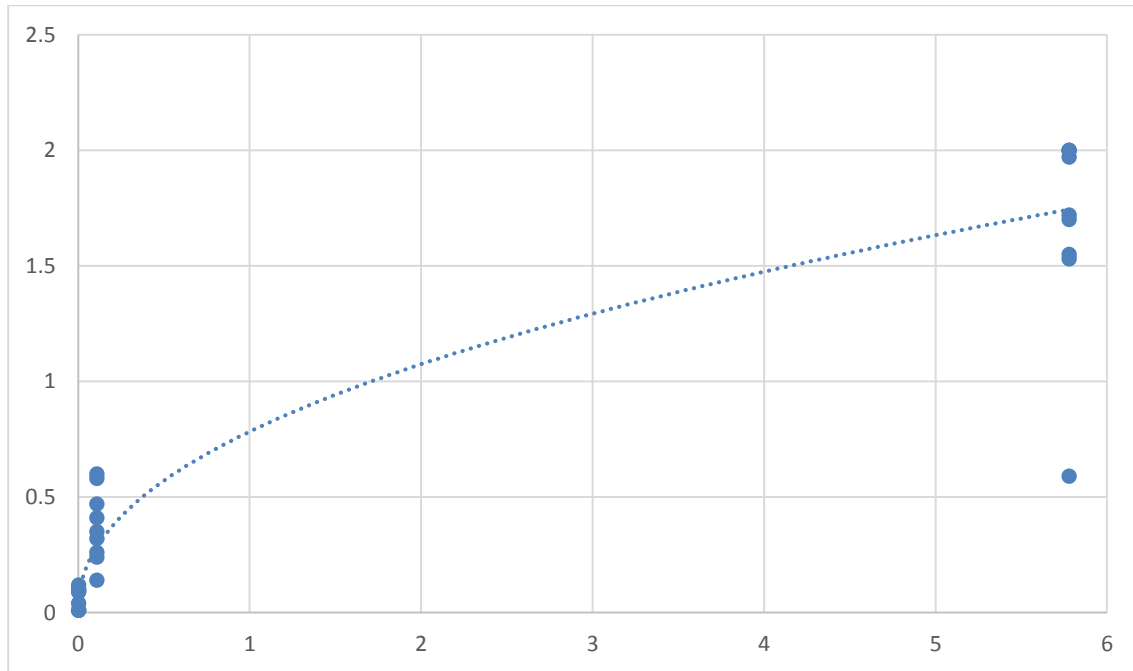
جدول ۷-۴- مقادیر R^2 برای انواع منحنی‌های برازش شده بر تابع f و وزن‌های توجه

	COD	SGW	SLD	BLR	NFS	PES	SNE1	SNE2	ACA	میانگین
خطی	0.9761	0.9234	0.9902	0.922	0.9824	0.958	0.9629	0.9311	0.9885	0.9594
لگاریتمی	0.8796	0.9484	0.8409	0.8337	0.8647	0.9109	0.9035	0.9416	0.8466	0.8855
توانی	0.9982	0.9862	0.9833	0.9874	0.9961	0.9974	0.9987	0.9148	0.9912	0.9837
نمایی	0.7969	0.6554	0.644	0.66	0.8131	0.7172	0.7311	0.4796	0.837	0.7038

برای بدست آوردن یک رابطه مستقل بین تابع f و وزن‌های توجه که مستقل از ویدئوها باشد، یک منحنی توانی بر روی داده‌های کل ویدئوها برازش می‌کنیم:

$$f = 0.7829 w^{0.4566} \quad \text{رابطه (۴-۱۸)}$$

مقدار R^2 برای این منحنی ۰٫۸۱۷ می‌باشد. این منحنی در شکل ۴-۵ نشان داده شده است.

شکل ۴-۵ - منحنی برازش شده بر روی تابع f و وزن‌های توجه

در این مرحله فرضیه بهبود تخمین تابع f با استفاده از ویژگی گودی مرکزی شبکه را بررسی می‌کنیم. به همین منظور، ابتدا تابع φ را معرفی می‌کنیم. این تابع وزن نواحی مختلف ویدئو بر پایه گودی مرکزی شبکه را محاسبه می‌کند. مقدار φ_i ، وزن مبتنی بر گودی چشم در ناحیه i ، بر پایه این یافته روان‌فیزیکی^۱ است که چگالی دریافت‌کنندگان نوری در گودی مرکزی شبکه که حدود ۲ تا

^۱ Psychophysical

۵ درجه حول محور دید را پوشش می‌دهد، در بالاترین حد خود است و با فاصله از محور دید به شدت کاهش می‌یابد [۸۲]:

$$\varphi_i = \max_{j \in Regions} \left(w_i \cdot \frac{1}{\sqrt{2\pi\sigma_x\sigma_y}} e^{-\left(\frac{(x_j-x_i)^2}{2\sigma_x^2} + \frac{(y_j-y_i)^2}{2\sigma_y^2}\right)} \right) \quad \text{رابطه (۴-۱۹)}$$

که در آن (x_i, y_i) مرکز ناحیه i و σ_x و σ_y به ترتیب حوزه‌ی دید^۱ افقی و عمودی هستند. به عنوان مثال، مقدار σ_x برای زاویه دید ۲ درجه در فاصله ۷۰ سانتیمتری از یک مانیتور ۹۶ نقطه بر اینچ^۲ (DPI) برابر ۴۶ پیکسل است. توجه داشته باشید که نقشه توجه محل دقیق نگاه بیننده را مشخص نمی‌کند. در واقع، این نقشه‌ها پایه احتمالاتی دارند. نواحی که بیشترین مقدار در نقشه توجه را دارند، با احتمال بیشتری دریافت‌کننده نگاه بیننده هستند؛ اما نقطه نگاه ممکن است در نواحی دیگر هم قرار گیرد. حتی اگر بیشینه‌های محلی را به عنوان نقاط توجه در نظر بگیریم، تضمینی وجود ندارد که مقادیر موجود در نقشه‌های توجه منطبق بر ویژگی گودی مرکزی شبکه‌ی توزیع شده باشند. بنابراین، ندانستن نقطه دقیق نگاه کار را دشوار می‌سازد. از این رو، در محاسبه φ_i مقدار w_i را به عنوان وزن احتمال در آن ضرب می‌کنیم. جدول ۴-۸ مقادیر محاسبه شده φ برای نواحی مختلف ویدئوها را نشان می‌دهد.

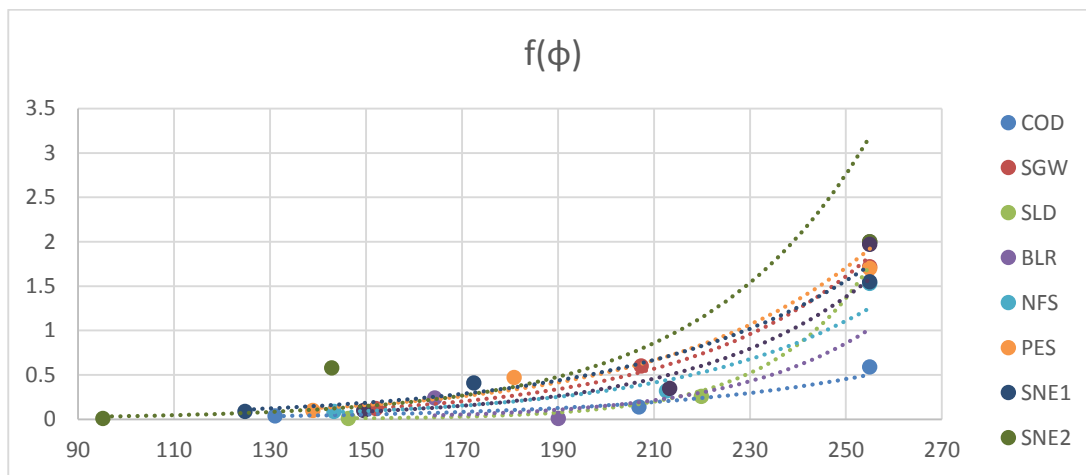
جدول ۴-۸- مقادیر φ برای هر ناحیه

	COD	SGW	SLD	BLR	NFS	PES	SNE1	SNE2	ACA
ناحیه ۱	131.02	152.06	146.31	190.10	143.33	138.93	124.80	95.203	149.50
ناحیه ۲	206.88	207.38	219.91	164.30	212.59	180.86	172.48	142.87	213.29
ناحیه ۳	254.98	254.97	254.99	254.99	254.99	254.97	254.98	254.99	254.99

اکنون به بررسی رابطه تابع f با مقادیر φ می‌پردازیم. شکل ۴-۶ منحنی برازش شده بر روی این مقادیر را نشان می‌دهد. انواع منحنی‌ها، از جمله منحنی‌های خطی، لگاریتمی، توانی و نمایی بر روی نمودار هر ویدئو برازش شد. مقدار R^2 منحنی‌های برازش شده در جدول ۴-۹ نمایش داده شده است. همانطور که مشخص است، رابطه مقادیر تابع f و وزن‌های گودی مرکزی شبکه‌ی در این دیتاست از نوع نمایی است.

^۱ Field of View

^۲ Dot Per Inch (DPI)



شکل ۴-۶ - رابطه مقادیر تابع f و وزن‌های توجه در ویدئوها

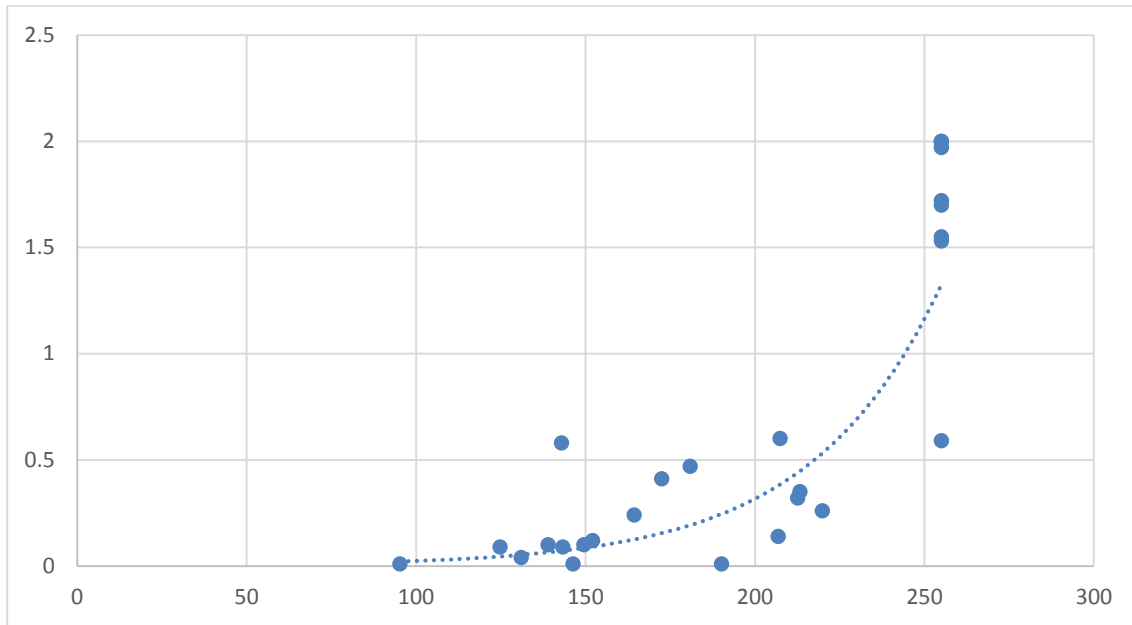
جدول ۴-۹ - مقادیر R^2 برای انواع منحنی‌های برازش شده بر تابع f و وزن‌های گودی مرکزی شبکه

	COD	SGW	SLD	BLR	NFS	PES	SNE1	SNE2	ACA	MEAN
خطی	0.7862	0.9287	0.6719	0.8586	0.7596	0.9799	0.9745	0.9998	0.7527	0.8568
لگاریتمی	0.7145	0.8874	0.613	0.8204	0.6959	0.9501	0.9343	0.9804	0.6913	0.8097
توانی	0.9381	1	0.9828	0.3194	0.9289	0.9835	0.9915	0.8486	0.9233	0.8795
نمایی	0.9719	0.9941	0.9951	0.3689	0.9612	0.9557	0.9639	0.8469	0.9557	0.8903

برای بدست آوردن یک رابطه مستقل بین تابع f و وزن‌های گودی مرکزی شبکه که مستقل از ویدئوها باشد، یک منحنی نمایی بر روی داده‌های کل ویدئوها برازش می‌کنیم:

$$f = 0.0017 e^{0.0261\phi} \quad \text{رابطه (۴-۲۰)}$$

مقدار R^2 برای این منحنی ۰,۶۴۶۲ می‌باشد. این منحنی در شکل ۴-۷ نشان داده شده است.



شکل ۴-۷- رابطه مقادیر تابع f و وزن‌های گودی مرکزی شبکه در ویدئوها

مقدار R^2 منحنی برازش شده بر روی مقادیر تابع f و وزن‌های توجه برابر $0,817$ و مقدار R^2 منحنی برازش شده بر روی مقادیر تابع f و وزن‌های گودی مرکزی شبکه برابر $0,6462$ است. در اینجا سعی می‌کنیم رابطه f با وزن‌های توجه و گودی مرکزی شبکه را بصورت تومان بدست آوریم. به همین منظور سعی می‌کنیم روابط بدست آمده را با یکدیگر ترکیب و ضرایب جدید را بدست آوریم:

$$f = \alpha_1 w^{\alpha_2} + \alpha_3 e^{\alpha_4 \varphi} \quad \text{رابطه (۴-۲۱)}$$

با استفاده از جعبه ابزار برازش منحنی نرم افزار متلب، مقادیر $\alpha_1, \alpha_2, \alpha_3$ و α_4 به ترتیب برابر $0,1634, 1, 0,4353, 6, 36e-6$ و $0,462$ است. مقدار R^2 این منحنی برابر $0,8793$ است که از مقدار R^2 دو منحنی قبلی بیشتر است. جدول ۴-۱۰ مقادیر واقعی f و مقادیر تخمین‌زده شده توسط رابطه (۴-۲۱) را نشان می‌دهد.

جدول ۴-۱۰- مقادیر واقعی و تخمین‌زده شده تابع f

		COD	SGW	SLD	BLR	NFS	PES	SNE1	SNE2	ACA
Region 1	Real	0.0751	0.0707	0.0723	0.0366	0.073	0.0739	0.0757	0.0773	0.0715
	Predicted	0.04	0.12	0.01	0.01	0.09	0.1	0.09	0.01	0.1
Region 2	Real	0.3543	0.3522	0.2806	0.4311	0.3273	0.4168	0.4254	0.439	0.3235
	Predicted	0.14	0.6	0.26	0.24	0.32	0.47	0.41	0.58	0.35
Region 3	Real	1.6737	1.6741	1.6734	1.6733	1.6731	1.6742	1.6736	1.6733	1.6732
	Predicted	0.59	1.72	2	2	1.53	1.7	1.55	2	1.97

برای بررسی بیشتر ویدئوهای گام‌های سوم و چهارم، فریم‌های مشابه از دو ویدئوی PES و SNE2، که اتفاقاً در آزمون کیفی مدل توجه بصری نسبت به سایر ویدئو امتیاز کمتری گرفته بودند، را با یکدیگر

مقایسه می‌کنیم. شکل ۴-۸ و شکل ۴-۹، به ترتیب دو فریم متناظر در ویدئوهای PES و SNE2 را نشان می‌دهند. در مورد فریم‌های بازی PES مشاهده می‌شود که مرز چمن‌های زمین بازی در ویدئوی گام سوم، بیشتر از بین رفته است و مقدار اعوجاج ناشی از کدکردن ویدئو به شکل بلاک به بلاک^۱ در آنها بیشتر است. هر چه این اعوجاج‌ها دورتر از محل توجه بازیکن باشند، کمتر به چشم می‌آیند. اما تحمل خطا در نواحی نزدیک به نقطه نگاه کمتر می‌شود. در مورد فریم گام سوم هم اعوجاج‌های نزدیک به بازیکن حامل توپ باعث افت کیفیت ادراکی می‌شود. بنابراین باید مقدار اعوجاج یک واحد کدشونده را متناسب با فاصله آن از نقطه نگاه در نظر گرفت که تابع φ کمک خوبی به این امر است. در مورد فریم‌های بازی SNE2 نیز همین مشاهده وجود دارد. با اینکه بازیکن به نقاط مهم بازی توجه می‌کند و افت کیفیت در آن نقاط کمتر به چشم او می‌آید، اگر مقدار اعوجاج در نقاط غیر مهم از حدی بیشتر گردد، باعث حواس‌پرتی بازیکن شده و کیفیت ادراکی را کاهش می‌دهد. این موضوع وقتی فاصله اعوجاج از محل توجه کم باشد، شدیدتر می‌شود.

^۱ Blockiness



شکل ۴-۸ - دو فریم متناظر از ویدئو PES - فریم بالا از گام سوم، فریم پایین از گام چهارم



شکل ۹-۴ - دو فریم متناظر از ویدئو SNE2 - فریم بالا از گام سوم، فریم پایین از گام چهارم

در ادامه برای ارزیابی عملکرد تابع f ، با استفاده از این تابع فرایند اختصاص بیت را بر روی تعدادی ویدئو غیر از ویدئوهای آموزش اجرا کرده و کیفیت ادراکی آن‌ها را اندازه می‌گیریم. برای این کار نرخ بیت نواحی مختلف از فریم‌های باقیمانده‌ی ویدئوهای PES، SNE1، SNE2 و ACA را بر اساس تابع f بدست آمده محاسبه کردیم. همچنین، با حفظ نرخ بیت کل، یک بار دیگر اختصاص بیت را بر اساس مدل [56] انجام دادیم و کیفیت ویدئوهای بدست آمده را با استفاده از آزمون ارزیابی کیفی مقایسه کردیم. جدول ۴-۱۱ نرخ بیت‌های اختصاص داده شده به هر یک از نواحی در ویدئوهای مختلف را نشان می‌دهد. از آنجا که نقشه توجه در ابتدای هر GOP بروزرسانی می‌شود، برای هر ویدئو بازی، نرخ بیت‌های اختصاص داده شده به نواحی مختلف در هر GOP گزارش شده است.

جدول ۴-۱۱- نرخ بیت‌های اختصاص داده شده به نواحی مختلف با مدل پیشنهادی و مدل [56]

ویدئوهای PES						
	مدل پیشنهادی			مدل [56]		
	ناحیه ۱	ناحیه ۲	ناحیه ۳	ناحیه ۱	ناحیه ۲	ناحیه ۳
1	257977	84984	110776	145888	88682	221717
2	263786	97545	106430	126306	117464	243479
3	242152	91543	105316	122768	109151	233405
4	255243	55812	106527	124626	58360	254990
5	245126	81006	127653	163820	71739	243747
6	245126	81006	127653	138227	55423	373731
7	351862	78287	113521	113942	101302	332939

ویدئوهای SNE1						
	مدل پیشنهادی			مدل [56]		
	ناحیه ۱	ناحیه ۲	ناحیه ۳	ناحیه ۱	ناحیه ۲	ناحیه ۳
1	102607	355376	195956	65903	424191	180369
2	36599	305003	221786	22097	367953	199688
3	98422	284957	224367	85705	344302	204155
4	96965	329395	207733	88886	397439	183043
5	160637	329874	201093	124216	390548	201051
6	120637	375449	219934	94631	448382	198395
7	73591	378918	222715	37115	458033	200598

ویدئوهای SNE2						
	مدل پیشنهادی			مدل [56]		
	ناحیه ۱	ناحیه ۲	ناحیه ۳	ناحیه ۱	ناحیه ۲	ناحیه ۳
1	703347	1571097	53487	117668	2373275	119461
2	460777	1408865	162499	226641	1683485	145694
3	272322	1342923	59668	98771	1520487	104096
4	643475	1382748	73963	174682	1775068	155640
5	324104	1378113	49895	150500	1545969	63709
6	511174	1305192	124011	303246	1700157	123591
7	373815	1064762	60772	152426	1406093	90576

ویدئوهای ACA						
	مدل پیشنهادی			مدل [56]		
	ناحیه ۱	ناحیه ۲	ناحیه ۳	ناحیه ۱	ناحیه ۲	ناحیه ۳
1	614507	1202852	1189782	547632	1293613	1209589
2	518434	1372518	1174053	518762	1398084	1262490
3	367015	1260294	973779	333170	1262168	973186
4	116937	974568	836069	92993	960120	931391
5	128497	930503	760025	118192	840273	832716
6	556743	1955767	1182647	627085	1959467	1183394
7	392301	1141806	1062617	417861	1137853	1059265

جدول ۴-۱۲- نتایج آزمایش کیفی

	متوسط امتیاز نظر				تغییر
	PES	SNE1	SNE2	ACA	
مدل [56]	۲,۹۳	۳,۰۵	۳,۱۸	۴,۳۷	-
مدل پیشنهادی	۳,۹۸	۳,۶	۴,۲۵	۴,۱۸	٪۲۰,۷

جدول ۴-۱۲ نتایج آزمون کیفی برای ویدئوهای تست را نشان می‌دهد. همانطور که دیده می‌شود، متوسط امتیاز نظر برای ویدئوهایی که اختصاص بیت‌شان بر اساس مدل بدست آمده انجام پذیرفته، حدود ۲۰٪ بیشتر است. البته طبق نتایج، کیفیت تفاوت در ویدئوی ACA برای شرکت‌کنندگان قابل تمیز نبوده است. این مشاهده را می‌توان مربوط به نوع بافت در این ویدئو دانست. در واقع، بافت‌های موجود در این ویدئو بگونه‌ای هستند که تفاوت نرخ بیت اختصاص داده شده از طریق مدل پیشنهادی و مدل [56] منجر به تفاوت قابل تمیز در ویدئو نمی‌شود.

۴-۲- جمع‌بندی

مدل‌های نرخ-اعوجاج مبتنی بر توجه سعی می‌کنند میزان اعوجاج در نواحی مختلف ویدئو را بر اساس میزان توجه بیننده به آنها کنترل نمایند بطوریکه نواحی پراهمیت، اعوجاج کمتر و نواحی کم‌اهمیت، اعوجاج بیشتری داشته باشند. با این وجود، همانطور که نتایج آزمایشات نشان می‌دهد، مقدار اعوجاج در نواحی کم‌اهمیت نمی‌تواند از حدی بیشتر کاهش یابد زیرا خود باعث تاثیر منفی بر روی کیفیت ادراکی بیننده می‌شود. در این فصل، مدل جدیدی ارائه شد که با در نظر گرفتن ویژگی گودی مرکزی شبکه چشم انسان علاوه بر سازوکار توجه میزان اعوجاج یک ناحیه را به نسبت فاصله آن ناحیه از محل‌های توجه محدود کرده و در نتیجه باعث کنترل بهتری بر روی میزان اعوجاج در نواحی مختلف می‌گردد. نتایج ارزیابی کیفی نشان می‌دهد با بکارگیری این مدل در اختصاص بیت، متوسط امتیاز نظر حدود ۲۰٪ افزایش می‌یابد. روش استفاده شده در ساخت مدل، همچنین امکان بررسی تاثیر همزمان سایر ویژگی‌های HVS مانند حساسیت چشم در رابطه نرخ-اعوجاج را فراهم می‌کند و می‌تواند برای انجام کارهای آتی مدنظر قرار گیرد.

فصل پنجم

نتیجه گیری و کارهای آینده

بازی‌های مبتنی بر رایانش ابری یک دامنه جذاب و نوظهور در صنعت سرگرمی محسوب می‌شود. با استفاده از این رویکرد جدید، بازیکن‌ها دیگر نیازی به خرید سخت‌افزارهای قوی و پرهزینه ندارند. آنها همچنین نگران ناسازگاری‌های احتمالی در نصب بازی نخواهند بود. با استفاده از این رویکرد هر فرد می‌تواند در هر زمان و بر روی هر دستگاهی به انجام بازی مورد علاقه خود که در محیط رایانش ابری قرار دارد بپردازد. افزایش دسترسی به اینترنت پرسرعت و پیشرفت تکنولوژی‌های سخت‌افزاری بخصوص در زمینه پردازنده‌های گرافیکی (GPU) و کدکننده‌های نرم‌افزاری یا سخت‌افزاری ویدئو شرایط را برای تحقق این رویکرد فراهم ساخته‌است.

این سرویس اگرچه مزایای زیادی در بردارد، اما چالش‌هایی هم دارد که باید مورد توجه قرار گیرد. یکی از این چالش‌ها، پهنای باند است. در این سرویس به ازای هر کاربر یک ویدئو آن هم با کیفیت HD با نرخ بیتی حدود ۵ مگابیت بر ثانیه بر روی شبکه ارسال می‌شود. بنابراین در صورت عدم استفاده از روش‌های فشرده‌سازی مناسب، شبکه‌ای که این سرویس بر روی آن ارائه می‌شود باید پهنای باند بالایی داشته باشد. بنابراین بررسی و ارائه راه‌حل در این زمینه کمک بزرگی به توسعه کاربرد بازی‌های مبتنی بر رایانش ابری خواهد کرد.

روش‌های فشرده‌سازی سعی می‌کنند با حذف افزونگی‌های آماری و ادراکی، نرخ فشرده‌سازی را افزایش دهند. استانداردهای به‌روز کدکردن ویدئو هر دو افزونگی را در نظر می‌گیرند. به عنوان مثال، تخمین حرکت، پیش‌بینی مد داخل فریمی^۱ و کدکردن بی‌نظمی^۲ از نمونه روش‌های حذف افزونگی‌های آماری هستند. نمونه‌گیری رنگ^۳، فیلتر رفع بلاک^۴، و ماتریس‌های پیمانانه‌سازی تضعیف فرکانس‌های بالا نیز از جمله روش‌های حذف افزونگی‌های ادراکی هستند. با این وجود، سامانه بینایی چشم انسان ویژگی‌های بسیار مهمی دارد که با استفاده از آنها می‌توان نرخ فشرده‌سازی را افزایش داد. بکارگیری این ویژگی‌ها در فرایند کدکردن، حوزه‌های مختلفی در کدکردن ادراکی ویدئو، از جمله کدکردن مبتنی بر توجه، پدید آورده است. هدف از این رساله، ارائه یک روش کدکردن ادراکی ویدئوی موثر برای کاربرد بازی‌های مبتنی بر رایانش ابری می‌باشد.

از آنجا که مدل‌های ادراکی فعلی در کاربرد بازی عملکرد مناسبی ندارند، در این رساله دو مدل ادراکی مخصوص بازی ارائه گردید. مدل اول، مدل توجه بازی نام دارد. این مدل هر دو سازوکار توجه پایین-به-بالا و بالا-به-پایین را در نظر گرفته و با ترکیب آنها محل توجه بازیکن را پیش‌بینی می‌کند. پیمانانه توجه بالا-به-پایین مدل پیشنهادی، به اشیا بازی بر اساس فعالیت فعلی بازیکن از جنبه‌های مختلف امتیاز داده و آنها را اولویت‌دهی می‌کند. نتایج آزمایش‌ها نشان می‌دهد، تاثیر پیمانانه توجه بالا-به-پایین برای کاربرد بازی بیشتر از پایین-به-بالا می‌باشد. این نتایج همچنین نشان می‌دهند با

^۱ Intra-frame

^۲ Entropy Coding

^۳ Chroma subsampling

^۴ Deblocking Filter

بکارگیری این مدل می‌توان با حفظ کیفیت ادراکی کاربر، نرخ بیت مودنیاز برای ویدئوی بازی را بطور متوسط ۲۵٪ کاهش داد. مدل دوم، به منظور نشان دادن تاثیر مهارت در تغییر الگوی توجه بازیکن‌ها ارائه گردید. این مدل، با خوشه‌بندی داده‌های جمع‌آوری شده از دستگاه ردیاب چشم سعی می‌کند، الگوهای متمایز توجه میان بازیکن‌ها با مهارت‌های مختلف را شناسایی کرده و برای هر سطح از مهارت یکنقشه توجه جداگانه تهیه نماید. برای تشخیص مهارت بازیکن‌ها در عمل نیز از امتیاز آنها استفاده می‌کند. نتایج آزمایش‌ها نشان می‌دهد، با در نظر گرفتن مهارت در مدل توجه بصری می‌توان دقت آن را تا حدود ۱۵٪ افزایش داد.

در نهایت یک مدل نرخ-عوجاج برای اعمال مدل‌های ادراکی در مسئله اختصاص بیت به نواحی مختلف ویدئو ارائه گردید. این مدل علاوه بر سازوکار توجه، سازوکار گودی مرکزی شبکه را نیز در نظر می‌گیرد. نتایج آزمایش‌ها نشان می‌دهد کیفیت ادراکی ویدئو با استفاده از این مدل در مقایسه با مدلی که فقط سازوکار توجه را در نظر می‌گیرد، از نظر متوسط امتیاز نظر ۲۰٪ بیشتر است. برای سرعت بخشیدن به آزمایش‌های آتی در زمینه فشرده‌سازی ادراکی ویدئو برای کاربرد بازی‌های مبتنی بر رایانش ابری دیتاست‌های تهیه شده در این رساله برای عموم منتشر می‌شوند. همچنین با توجه به پیمانه‌بندی ضعیف سامانه آزمون GA که حل مشکلاتی از قبیل قلاب انداختن به بازی برای گرفتن فریم‌های آن را سخت می‌کند، یک سامانه آزمون متن باز مبتنی بر گراف فیلتر ارائه گردید که از پیمانه‌بندی بالاتری برخوردار است.

۵-۱- کارهای آینده

کارهای آتی این رساله به دو دسته تقسیم می‌شوند. دسته اول مربوط به تکمیل و بهبود مدل‌های ارائه شده می‌باشد. در حالی که دسته دوم مربوط به وارد کردن جنبه‌های کاملاً جدید به تحقیقات انجام شده است.

بهبود مدل‌ها و سامانه آزمون CG در دسته اول قرار می‌گیرند. در مورد مدل توجه بازی، انتخاب بهینه روش‌های پایین-به-بالا و بالا-به-پایین از نظر پیچیدگی پردازش، دقت پیش‌بینی محل توجه بازیکن و کیفیت ادراکی ویدئو بازی در برنامه کارهای آتی قرار دارند. با توجه به اینکه، این مدل اولین تلاش برای در نظر گرفتن توجه بالا-به-پایین در کاربرد بازی است، در این رساله، طراحی مدل، انتخاب پیمانه‌های موجود در آن و ارتباط آن‌ها در مقایسه با انتخاب بهینه روش‌های بکار رفته برای پیمانه‌های آن در اولویت قرار داشت. در مورد مدل توجه مبتنی بر مهارت، در گام خوشه‌بندی از معیار میانگین و انحراف معیار همبستگی برای انتخاب خوشه‌ها استفاده کردیم. به عنوان کار آتی، قصد داریم معیار NSS را در خود گام‌های مدل بکار گیریم. به عنوان مثال، می‌توانیم در انتخاب خوشه‌ها علاوه بر میانگین و انحراف معیار همبستگی، تاثیر نهایی آنها در دقت نقشه‌ها را با معیار NSS بسنجیم. همچنین می‌توانیم با استفاده از آن موقعیت‌های بازی متمایز را شناسایی کرده و برای آنها

نقشه جداگانه طراحی کنیم. البته برای انجام این کار نیاز است تا دیتاست فعلی را توسعه داده و چنین تغییر موقعیت‌هایی را در آن بگنجانیم. یکی دیگر از کارهای آتی مربوط به مدل توجه مبتنی بر مهارت، اندازه‌گیری تاثیر ورودی‌های کاربر در پیش‌بینی الگوی توجه یا به طور دقیق‌تر خوشه توجه وی می‌باشد. در مورد سامانه آزمون CG نیز، در آینده قصد داریم با پیاده‌سازی فیلترهای جدید و بهبود فیلترهای فعلی سامانه را توسعه دهیم.

دسته دوم کارهای آتی، مربوط به وارد کردن جنبه‌های جدید به تحقیقات این رساله است. یکی از مهمترین این جنبه‌ها در نظر گرفتن پیچیدگی کدکردن در گام وارد کردن مدل‌های ادراکی به کدکننده می‌باشد. به طور دقیق‌تر، با توسعه مدل‌های پیچیدگی-اعوجاج می‌توان بعد از آنکه مقدار اعوجاج قابل تحمل در نواحی مختلف ویدئو بدست آمد، پارامترهای کدکننده را طوری تنظیم کرد که پیچیدگی کدکردن نواحی نیز متناسب با اهمیت آنها باشد. به عنوان مثال، می‌توان با محدود کردن مقدار پارامترهایی مانند بازه جستجو در تخمین حرکت در نواحی کم اهمیت، پیچیدگی کدکردن در این نواحی را کاهش داد. اگرچه این کار کیفیت و نرخ بیت در آن نواحی را تغییر می‌دهد، اما اگر تغییرات منطبق با مدل ادراکی باشد، می‌تواند در مجموع باعث کاهش بار پردازشی شود. در [۸۳]، نتایج اولیه آزمایش‌ها در این زمینه ارائه شده است.

ضمیمه الف

سامانه آزمون بازی های مبتنی بر رایانش ابری

کاربرد CG همراه با مزایای فوق‌العاده‌اش برای بازیکن‌ها، طراحان بازی و ارائه‌دهندگان خدمت، چالش‌هایی نیز دارد. پهنای باند زیاد، تاخیر کم، مقیاس‌پذیر بودن، مقاومت در برابر خطا، توزیع منابع و غیره از نکاتی هستند که در طراحی سامانه‌های CG باید مدنظر قرار گرفته شوند. البته این چالش‌ها نه تنها مانع از رشد این کاربرد نشده‌اند، بلکه رقابت بزرگی میان ارائه‌دهندگان سرویس بوجود آورده‌اند تا با فائق آمدن بر آنها و کسب جایگاه نخست در بازار به سودهای هنگفت دست یابند. به همین دلیل، تحقیقات زیادی بر روی جنبه‌های نرم افزاری و سخت افزاری CG انجام می‌شود.

یکی از نیازهای اساسی این تحقیقات، وجود یک سامانه آزمون CG است تا دانشمندان بتوانند آزمایش‌های خود را بر روی آن پیاده‌سازی کرده و عملکرد ایده‌هایشان را در عمل بسنجند. سامانه آزمون باید به اندازه کافی به واقعیت نزدیک باشد بطوریکه بتوان نتیجه آزمایش‌ها را با انجام کمی تنظیمات به شکل واقعی اجرا کرد. متن باز بودن، گسترش‌پذیر بودن و قابل تنظیم بودن از دیگر ویژگی‌های موردنیاز چنین سامانه آزمون است.

برای انجام یک ارزیابی ایده‌آل باید یک سامانه آزمون CG راه‌اندازی کرد و از افراد خواست تا در آن بستر یکبار بدون اعمال تغییرات و یکبار با اعمال تغییرات به انجام بازی بپردازند و تجربه خود را گزارش کنند تا عملکرد مدل‌ها بطور دقیق‌تری ارزیابی شود. تنها سامانه آزمون متن باز، که در اواسط اجرای این رساله منتشر گردید، سامانه (GA) GamingAnywhere [۲] می‌باشد که جزئیات آن در همین بخش مورد بررسی قرار می‌گیرد. در کنار تمام ویژگی‌های خوبش، یکی از نقاط ضعف این سامانه استفاده از فن قلاب^۱ برای گرفتن فریم‌های بازی می‌باشد. سامانه GA بطور مستقیم قلاب را وارد فایل‌های اجرایی بازی می‌کند. اگر قلاب با موفقیت متصل شود، از سایر روش‌ها مانند خواندن بافر گرافیک به صورت دوره‌ای، عملکرد بهتری خواهد داشت. با این وجود، این عملیات در اکثر مواقع درست عمل نمی‌کند. به عنوان مثال، GA برای بازی‌هایی که بیش از یکبار در طول بازی دستگاه Direct3D را اخذ می‌کنند، به درستی عمل نمی‌کند. این در حالی است که اکثر بازی‌ها ابتدا یک دستگاه اخذ می‌کنند و فایل‌های موردنیاز بازی، مانند فایل‌های بافت و مش، را در حافظه بارگذاری می‌کنند، سپس برای اجرای خود بازی دستگاه دیگری با ویژگی‌های مناسب‌تر اخذ می‌کنند. در واقع فن قلاب برای هر بازی چالش‌های خاص خود را دارد. بنابراین بهتر است بخش گرفتن بازی به‌عنوان یک پیمان‌نامه مستقل در سامانه طراحی شود، تا بتوان این مشکل را مدیریت کرد.

در این رساله، ما یک سامانه آزمون جدید طراحی و پیاده‌سازی کردیم که بر پایه گراف فیلتر DirectShow عمل می‌کند. این گراف‌ها اجازه می‌دهند هر تعداد فیلتر با هر ترتیب دلخواهی را به هم متصل کنیم و یک عملیات چندرسانه‌ای انجام دهیم. بنابراین نه تنها در بخش فن قلاب، بلکه در سایر بخش‌ها نیز پیمان‌بندی بهتری از آنها حاصل می‌شود. به علاوه، هر فیلتر می‌تواند قبل و در حین اجرای گراف تنظیم گردد. بنابراین، این سامانه مانند GA گسترش‌پذیر و قابل تنظیم نیز می‌باشد. با این وجود، با توجه به اینکه DirectShow بر روی سیستم‌عامل ویندوز اجرا می‌شود، کارساز این

^۱ Hooking Technique

سامانه قابلیت چند سکویی ندارد. البته توجه داشته باشید که گراف‌های فیلتر چندرسانه با استفاده از GStreamer [۸۴] بر روی سیستم‌عامل‌های لینوکس هم قابل ساخت هستند. مزیت اساسی این سامانه مستقل بودن پیمانه‌های آن از همدیگر است. بطور دقیق‌تر، می‌تواند برای رسیدن به هدف خاص خود پیاده‌سازی یک فیلتر را تغییر دهند بدون آنکه عملکرد سایر فیلترها خدشه‌دار شود. به عنوان مثال، پیمانه فن قلاب می‌تواند برای هر بازی بصورت جدا تنظیم شود. همچنین، دانشمندان می‌توانند با اضافه و کم کردن و تغییر دادن فیلترها عملکرد سامانه را گسترش دهند. به عنوان مثال، با اضافه کردن یک فیلتر اندازه‌گیری تاخیر، می‌توان روش‌های مختلف افزایش سرعت در کد کردن و جویبارسازی ویدئو در سمت کارساز را با یکدیگر مقایسه کرد. به عنوان مثال دیگر، استفاده از یک فیلتر اندازه‌گیری کیفیت می‌توان عملکرد روش‌های کد کردن ادراکی در کاهش پهنای‌بند را اندازه‌گیری کرد.

۱-۱- سامانه GamingAnywhere

اولین سامانه متن باز CG، سامانه GA می‌باشد. علاوه بر باز بودن، این سیستم قابلیت توسعه، انتقال و تنظیم بالایی دارد. در حالیکه کارساز این سیستم بر روی ویندوز، لینوکس و سیستم عامل اکس پیاده سازی شده است، کارخواه آن می‌تواند بر روی سیستم عامل‌های دیگر از جمله iOS و اندروید اجرا شود. تاخیر به ازای هر فریم در این سیستم ۳۴ میلی‌ثانیه می‌باشد که در مقایسه با Onlive و StreamMyGame به ترتیب ۳ و ۱۰ برابر کمتر است.

۱-۱-۱ طراحی

سیستم GA یک بستر باز برای توسعه و مطالعه کاربردهای جویبارسازی چندرسانه‌ای در اختیار محققین قرار داده است. در طراحی این سیستم موارد زیر مدنظر قرار داده شده است:

۱. توسعه‌پذیری^۱: این سیستم بصورت پیمانه‌ای طراحی شده است. هم اجزای وابسته-به-ماشین^۲ مانند گرفتن صدا و تصویر و هم اجزای مستقل-از-ماشین^۳ مانند کدک‌ها و پروتکل‌های شبکه می‌توانند به راحتی اصلاح یا تعویض شوند. توسعه‌دهندگان می‌توانند با دنبال کردن واسط-های برنامه نویسی پیمانه‌های GA قابلیت‌های آن را افزایش دهند. این سیستم محدود به بازی نمی‌باشد و هر کاربرد جویبارسازی چندرسانه‌ای دیگری از جمله پخش زنده^۴ نیز با استفاده از آن قابل انجام است.

^۱ Extensibility

^۲ Platform-dependent

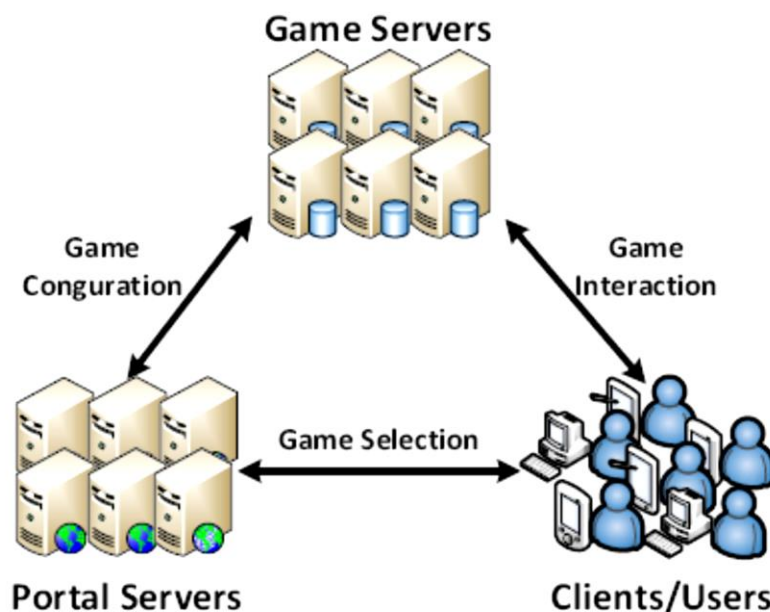
^۳ Platform-independent

^۴ Live casting

۲. انتقال‌پذیری^۱: با گسترش شبکه‌های وایرلس، دستگاه‌های همراه بطور فزاینده مشتری سرویس‌های ابری می‌شوند. به همین دلیل در طراحی GA، انتقال‌پذیری مورد توجه قرار گرفته است. در حال حاضر این سیستم برای ویندوز، لینوکس، OS X و اندروید پیاده‌سازی شده است. کتابخانه‌های بکار رفته در GA نیز بسیار انتقال‌پذیر هستند.
۳. تنظیم‌پذیری^۲: محققین برای مطالعه کاربردهای جویبارسازی چندرسانه‌ای ممکن است بخواهند پارامترهای زیادی را مورد بررسی دهند. به همین دلیل تعدادی زیادی کدک صدا و تصویر در GA قرار داده شده است. امکان تغییر تمامی پارامترهای قابل تنظیم توسط فایل‌های متنی برای کاربر وجود دارد.
۴. باز بودن: فایل‌های اجرایی و متن این سیستم در آدرس <http://gaminganywhere.org/> در اختیار عموم قرار دارد.

۱-۲- معماری

سناریوی استفاده عملی از GA در شکل ۱-ا نشان داده شده است. کاربر ابتدا از طریق درگاه وارد حساب کاربری خود می‌شود. لیست بازی‌ها از طریق درگاه در اختیار کاربر قرار داده شده و کاربر بازی موردنظر خود را انتخاب می‌کند. بعد از مشخص شدن بازی، کارساز درگاه، یک کارساز بازی دردسترس را پیدا کرده و بازی را بر روی آن اجرا کرده و آدرس آن را در اختیار کاربر قرار می‌دهد. بدین ترتیب، کاربر مستقیماً به کارساز بازی وصل شده و شروع به بازی کردن می‌نماید.

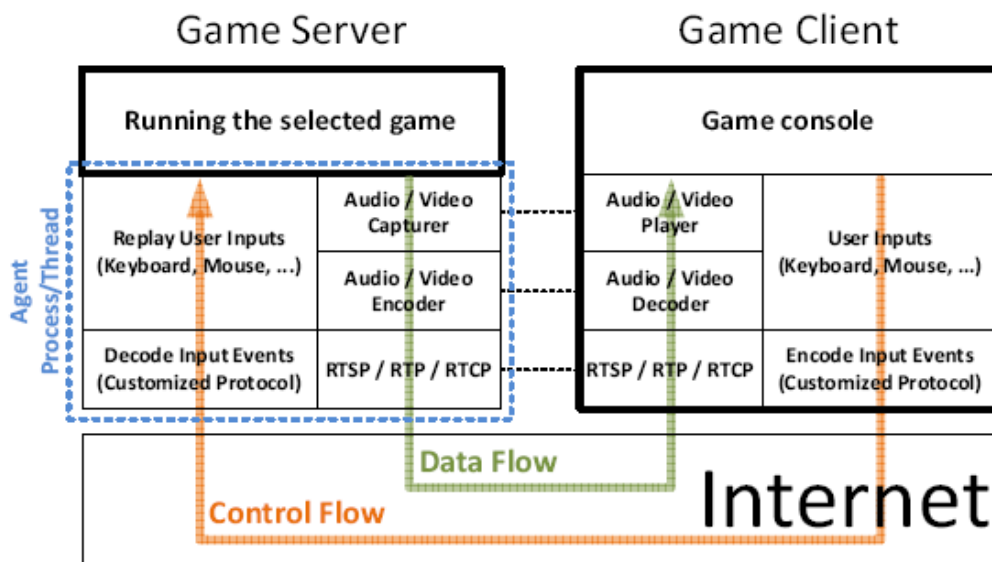


شکل ۱-ا - سناریوی استفاده از سیستم GA [۲]

^۱ Portability

^۲ Configurability

شکل ۲-۱ معماری کارساز بازی GA را نشان می‌دهد. در این معماری دو جریان شبکه‌ای تعریف شده است: جریان داده و جریان کنترل. جریان داده برای جویبارسازی فریم‌های صدا و تصویر از سمت کارساز به کارخواه می‌باشد. به غیر از بازیکن سایر افراد نیز می‌توانند از طریق یک پخش‌کننده مانند VLC (۳۹) به تماشای بازی بپردازند. آنها البته نمی‌توانند در بازی مداخله داشته باشند. جریان کنترل، در جهت عکس، برای ارسال عمل‌های کاربر از سمت کارخواه به کارساز می‌باشد. معماری GA اجازه می‌دهد بازی‌های رایانه شخصی و همچنین بازی‌های مبتنی بر وب بر روی آن اجرا شود.



شکل ۲-۱- نگاه پیمانه‌ای به کارساز و کارگزار GA [۲]

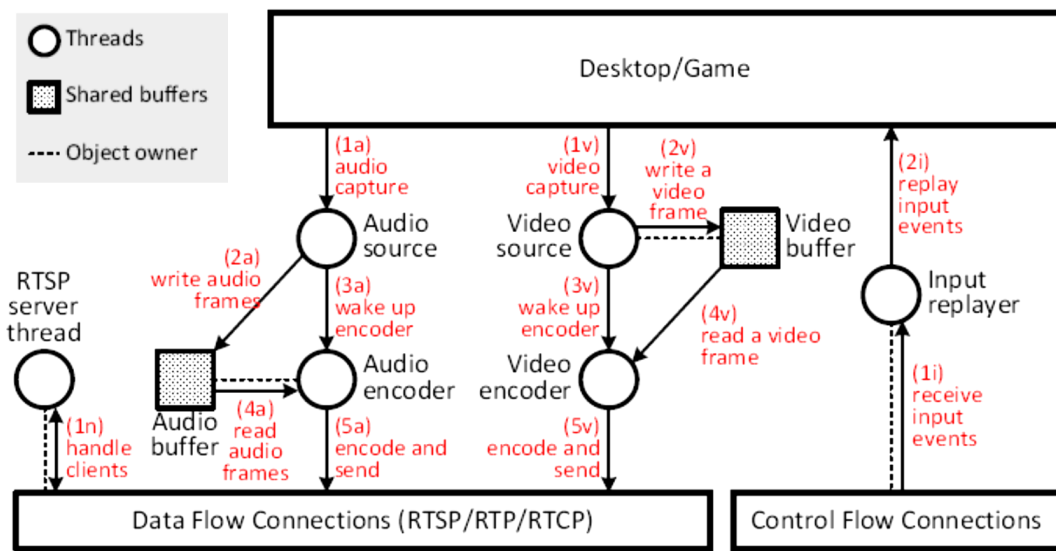
۱-۳- پیاده‌سازی

در پیاده‌سازی این سیستم از چند کتابخانه خارجی از جمله libavcodec/libavformat، live555 و SDL استفاده شده است. کتابخانه‌های libavcodec/libavformat، بخشی از پروژه ffmpeg هستند که برای ضبط، تبدیل و جویبارسازی ویدئو و صدا ساخته شده‌اند. سیستم GA از این کتابخانه برای کد کردن و واگذاری فریم‌های صدا و تصویر استفاده می‌کند. از این کتابخانه همچنین برای مدیریت پروتکل RTP در سمت کارساز استفاده می‌شود. کتابخانه live555 مجموعه‌ای از کتابخانه‌های C++ برای جویبارسازی با پروتکل‌های باز (RTSP، RTP، RTCP و SIP) می‌باشد. سیستم GA از این کتابخانه برای مدیریت پروتکل‌های RTSP/RTP در سمت کارخواه استفاده می‌کند. کتابخانه SDL (Simple DirectMedia Layer) یک کتابخانه فرامشین^۱ است که با استفاده از OpenGL امکان دسترسی مستقیم به سخت افزار صدا، موشواره، صفحه کلید، اهرمک^۲ و گرافیک را فراهم می‌کند. سیستم GA از این کتابخانه برای به تصویر کشیدن صدا و تصویر در کارخواه استفاده می‌کند.

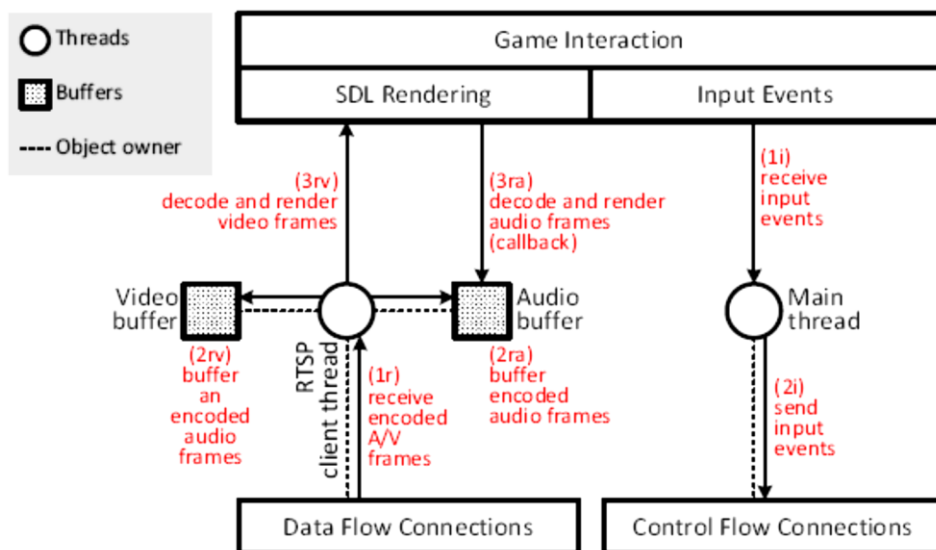
^۱ Cross-platform

^۲ Joystick

سیستم GA با استفاده از پیمانه API intercept توابع گرافیکی را فراخوانی کرده و بافر عقب بازی را مستقیماً و بلافاصله بعد از اتمام فرایند به تصویرکشیدن می‌خواند. این پیمانه بصورت رویدادمحور عمل می‌کند بنابراین نرخ فریم در آن متغیر است. به همین دلیل در آن از یک کنترل‌کننده token bucket استفاده می‌شود. فریم‌های کپچر شده در یک بافر مشترک با کدکننده ذخیره می‌شوند. در حال حاضر از کدکننده x264 در GA استفاده می‌شود. این کدکننده در معماری GA، به ازای هر ببیننده با دقت مکانی ۱۲۸۰ در ۷۲۰ و نرخ فریم ۲۴، ده درصد از یک پردازنده Intel 2.66 GHz i5 را مشغول می‌کند. شکل ۳-ا و شکل ۴-ا ارتباط بین اجزای به ترتیب مرتبط با کارساز و کارخواه را نمایش می‌دهند.



شکل ۳-ا- ارتباط بین پیمانه‌های کارساز، بافرهای مشترک و ارتباطات شبکه در GA [۲]



شکل ۴-ا- ارتباط بین پیمانه‌های کارخواه، بافرهای مشترک و ارتباطات شبکه در GA [۲]

أ-۲- سامانه‌ی ارائه شده

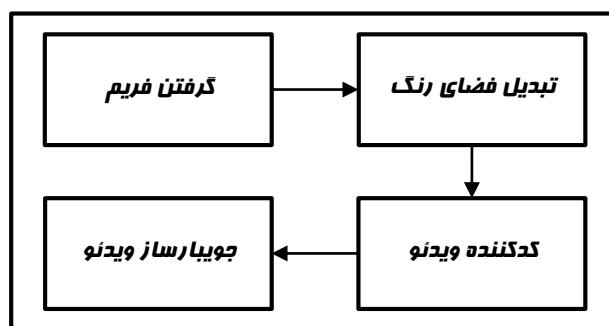
أ-۲-۱- ساختار سامانه

سامانه آزمون CG از دو بخش اساسی تشکیل می‌شود: کارساز و کارخواه. به تصویر کشیدن، کد کردن ویدئو و جویبارسازی ویدئو در سمت کارساز انجام می‌شود. در سمت دیگر، کارخواه جویبار ویدئو را دریافت، واكد و نمایش می‌دهد. همچنین کنترل‌های بازی را گرفته و برای کارساز ارسال می‌کند.

أ-۲-۱-۱- کارساز

کارساز سامانه به زبان C# نوشته شده است و دو ریشه^۱ دارد. یک ریشه مسئول دریافت کنترل‌های بازی و ارسال آنها به موتور بازی است. ریشه دیگر گراف فیلتر DirectShow را اجرا می‌کند. این گراف یک عملیات چندرسانه‌ای پیچیده را به ترتیبی از چند فرایند کوچک‌تر (فیلتر) می‌شکاند. هر فیلتر بخشی از پردازش را انجام می‌دهد. فیلترها پین‌های ورودی و خروجی دارند و از طریق آنها به یکدیگر متصل می‌شوند. بطور دقیق‌تر، برای انتقال داده از یک فیلتر به فیلتر دیگر، پین خروجی فیلتر بالادستی به پین ورودی فیلتر پایین‌دستی متصل می‌شود. هر پین انواعی از داده‌ها را پشتیبانی می‌کند. دو پین وقتی می‌توانند به یکدیگر متصل شوند که از نوع داده مشترکی پشتیبانی کنند. برای ساخت گراف، ابتدا باید نسخه‌ای از هر فیلتر ساخت. سپس پین فیلترهای ساخته شده را به هم متصل کرد.

همانطور که قبلاً هم گفته شد، محدودیتی روی تعداد فیلترها وجود ندارد و کاربر می‌تواند به دلخواه فیلترها را کم یا اضافه نماید تا گراف حاصل منطبق بر هدف آزمایشگاهی وی گردد. با این وجود، پیاده‌سازی ما شامل چهار فیلتر می‌باشد. شکل أ-۵ گراف داخل کارساز سامانه آزمون CG را نشان می‌دهد. مسئولیت هر یک از این فیلترها در ادامه تشریح می‌شود.



شکل أ-۵ - گراف فیلتر در سمت کارساز شامل چهار فیلتر

^۱ Thread

در برخی کاربردها، دانشمندان مایلند برای مطالعه یک رفتار خاص، بازی مشخصی طراحی کنند. اگر بازی خیلی پیچیده نباشد، می‌توان خودش را در قالب یک فیلتر طراحی کرد و در گراف قرار داد. به این صورت سربار گرفتن فریم کمتر می‌شود.

أ-۲-۱-۱-۲- تبدیل فضای رنگ

اکثر کدکننده‌های ویدئو ورودی از نوع YUV را می‌پذیرند. بنابراین بسته به نوع خروجی فیلتر گرفتن فریم، ممکن است به یک فیلتر تبدیل فضای رنگ احتیاج باشد. در پیاده‌سازی ما، از آنجا که خروجی فیلتر گرفتن فریم، RGB است از یک فیلتر تبدیل کننده RGB به YUV استفاده شد.

أ-۲-۱-۱-۳- کدکننده ویدئو

در حال حاضر اکثر شرکت‌های CG از استاندارد H.264/AVC استفاده می‌کنند. به همین دلیل، ما از یک فیلتر کدکننده H.264/AVC مبتنی بر Nvidia CUDA در سامانه استفاده کردیم [۸۶]. با این وجود، امکان استفاده از هر استاندارد دیگری نیز وجود دارد. حتی دانشمندان می‌توانند تغییراتی برای بهبود استانداردهای موجود پیشنهاد داده و آنها را در قالب فیلتر پیاده‌سازی کرده و در گراف از آنها استفاده نمایند.

أ-۲-۱-۱-۴- جویبارسازی ویدئو

بعد از کدشدن ویدئو، باید آن را برای کارساز جویبارساز ارسال کرد. در این بخش از سامانه، از فیلتر کارساز^۱ RTSP نوشته آقای گراینت دیویس [۸۷] که از استانداردهای H.264 و ACC پشتیبانی می‌کند، استفاده می‌کنیم. این فیلتر با جداسازی واحدهای لایه دسترسی شبکه^۲ (NAL) PPS^۳ و SPS^۴ اطلاعاتی مانند نوع و فرمت و غیره را بدست می‌آورد و پروفایل جلسه را می‌سازد. در حین جویبارسازی، بسته‌های رسانه‌ای را درون بسته‌های شبکه‌ای پروتکل RTP^۵ قرار می‌دهد و برای کارخواه ارسال می‌کند.

^۱ Real-Time Streaming Protocol

^۲ Network Access Layer

^۳ Picture Parameter Set

^۴ Sequence Parameter Set

^۵ Real-Time Protocol

أ-۲-۱-۲- کارخواه

هدف نهایی CG آن است که بازیکن‌ها بتوانند بازی مورد علاقه خود را در هر جایی که هستند، در هر زمانی که می‌خواهند و بر روی هر دستگاهی که دارند، اجرا کنند. بدون شک، اگر کارخواه CG قابلیت اجرا شدن بر روی مرورگر را دارا باشد، CG به سرعت به آنچه برای آن طراحی شده است، نزدیک خواهد شد. در واقع، اکنون که دستگاه‌های همراه فراگیر شده‌اند و مرورگرهای وب قدرتمند بر روی آنها نصب شده است. کاربران می‌توانند با استفاده از تگ ویدیوی HTML5 بدون نیاز به نصب وصله^۱ یا افزونه^۲، ویدئو تماشا کنند.

علاوه بر نمایش ویدئو، کارخواه باید توانایی دریافت و ارسال کنترل‌های کاربر به کارساز را نیز داشته باشد. در این خصوص، از چهارچوب متن باز جاوااسکریپت AngularJS [۸۸] برای دریافت موشواره و صفحه کلید استفاده کردیم. اطلاعات لمسی نیز از طریق پیمانه ngTouch این چهارچوب قابل دریافت است [۸۹]. سپس برای انتقال اطلاعات به کارساز، از فناوری Web Sockets استفاده کردیم. این فناوری اجازه می‌دهد در مرورگرهای منطبق با HTML5 از طریق واسط جاوااسکریپت Socket باز کنیم. برنامه‌هایی که بر مبنای HTML5، CSS^۳ و جاوااسکریپت ساخته می‌شوند را می‌توان بهر راحتی به برنامه‌های تلفن همراه تبدیل کرد [۹۰]. این برنامه‌ها در واقع شامل یک مرورگر پنهان در داخل خود می‌باشند.

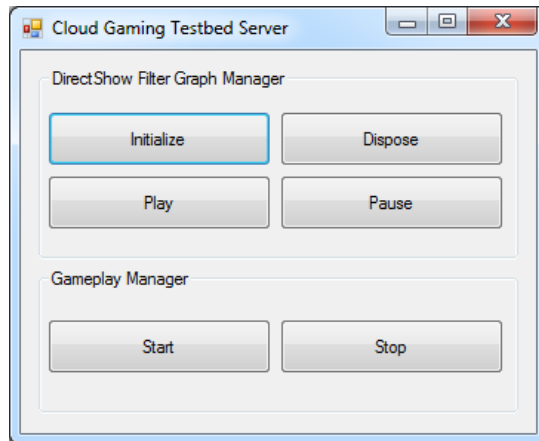
أ-۲-۲- راه‌اندازی

برای اجرای کارساز بر روی یک ماشین باید دو نکته را در نظر گرفت. اول، باید چهارچوب Net. نسخه ۳,۵ یا بالاتر بر روی آن ماشین نصب باشد. دوم، اگر از فیلترهای آماده استفاده می‌شود، باید قبل از اجرای کارساز آنها را بر روی ماشین ثبت کرد. شکل أ-۷ نمای اصلی برنامه کارساز را نشان می‌دهد. همانطور که در این شکل مشخص است، کارساز از دو بخش «مدیریت گراف» و «مدیریت کنترل‌های ورودی» تشکیل شده است. بخش اول، واسطی برای آماده‌سازی، اجرا، توقف و آزادسازی گراف است. بخش دوم، مسئول دریافت کنترل‌های ورودی از سمت کارخواه و شبیه‌سازی آنها بر روی ماشین کارساز است. این کار با استفاده از فراخوانی تابع SendInput در کتابخانه user32.dll انجام می‌شود. به این ترتیب موتور بازی، بصورت شبیه‌سازی شده موشواره و صفحه کلید کارخواه را خواهد داشت.

^۱ Plug-in

^۲ Add-on

^۳ Cascade Style Sheet



شکل ۷-۱ - نمای اصلی برنامه کارساز

وقتی گراف فیلترها ساخته شد، کاربر می‌تواند آن را اجرا یا متوقف کند. اگر خود بازی به شکل یک فیلتر پیاده‌سازی نشده باشد، باید قبل از ساخت گراف فایل اجرایی بازی را اجرا کرد. بعد از اتمام آزمایش‌ها کاربر باید گراف را از بین ببرد تا منابع آزاد گردد.



شکل ۸-۱ - سامانه آزمون در حال اجرا

شکل ۸-۱ سامانه آزمون در حال اجرا را نشان می‌دهد. در این شکل، بر روی ماشین سمت راست بازی «عروسک قشنگ من، سُمی» و کارساز اجرا شده است. ماشین سمت چپ، جوینار ویدئو را دریافت و آن را نمایش می‌دهد.

۳-۱- جمع‌بندی

در کاربرد CG علاوه بر منطق بازی، به تصویر کشیدن ویدئو نیز در سمت کارساز انجام می‌شود. ولی برای عملی شدن این کاربرد امیدبخش باید بر چالش‌های آن غلبه کرد. راه‌حل‌های بالقوه زیادی برای این چالش‌ها وجود دارد اما هیچ راه‌حلی در عمل استفاده نخواهد شد مگر آنکه با آزمایش‌های دقیق و مطمئن برتری آن نسبت به راه‌حل‌های موجود اثبات شود. بنابراین دانشمندان به ابزارهای آزمونی احتیاج دارند تا عملکرد ایده‌های خود را بسنجند و راه‌حل‌های بهینه را بیابند. در حال حاضر، GA تنها سامانه آزمون CG است اما همانطور که در این فصل اشاره شد، محدودیت‌هایی دارد. برای

برطرف کردن بخشی از این محدودیت‌ها، سامانه آزمون جدیدی بر مبنای گراف فیلترهای چندرسانه‌ای ارائه شد. این سامانه با ارائه پیمانه‌بندی ریزتر، رفع اشکال توسعه سامانه را آسان‌تر می‌کند. البته، این سامانه در اوایل مسیر رشد خود قرار دارد و باید بطور مداوم آن را توسعه داد تا جایگاه خود را میان دانشمندان پیدا کند. در آینده قصد داریم با پیاده‌سازی فیلترهای جدید و بهبود فیلترهای فعلی سامانه را توسعه دهیم.

مراجع

- [1] (2015/12/19). *Video Game Industry* [Online]. Available: http://vgsales.wikia.com/wiki/Video_game_industry
- [2] C.-Y. Huang, C.-H. Hsu, Y.-C. Chang, and K.-T. Chen, "GamingAnywhere: an open cloud gaming system," in *Proceedings of the 4th ACM Multimedia Systems Conference*, 2013, pp. 36-47.
- [3] P. E. Ross, "Cloud Computing's Killer App: Gaming," *Spectrum, IEEE*, vol. 46, pp. 14-14, 2009.
- [4] A. Ojala and P. Tyrvaenen, "Developing cloud business models: A case study on cloud gaming," *Software, IEEE*, vol. 28, pp. 42-47, 2011.
- [5] U. Lampe, M. Siebenhaar, D. Schuller, and R. Steinmetz, "A cloud-oriented broker for cost-minimal software service distribution," in *Proceedings of the 2nd ServiceWave Workshop on Optimizing Cloud Services*, 2011.
- [6] S. Wang and S. Dey, "Modeling and characterizing user experience in a cloud server based mobile gaming approach," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, 2009, pp. 1-7.
- [7] M. Jarschel, D. Schlosser, S. Scheuring, and T. Hoßfeld, "An Evaluation of QoE in Cloud Gaming Based on Subjective Tests," in *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2011 Fifth International Conference on*, 2011, pp. 330-335.
- [8] M. Jarschel, D. Schlosser, S. Scheuring, and T. Hoßfeld, "Gaming in the clouds: QoE and the users' perspective," *Mathematical and Computer Modelling*, 2011.
- [9] K. T. Chen, Y. C. Chang, P. H. Tseng, C. Y. Huang, and C. L. Lei, "Measuring the latency of cloud gaming systems," in *Proceedings of the 19th ACM international conference on Multimedia*, 2011, pp. 1269-1272.
- [10] S. Choy, B. Wong, G. Simon, and C. Rosenberg, "The brewing storm in cloud gaming: A measurement study on cloud to end-user latency," in *Network and Systems Support for Games (NetGames), 2012 11th Annual Workshop on*, 2012, pp. 1-6.
- [11] Y.-C. Chang, P.-H. Tseng, K.-T. Chen, and C.-L. Lei, "Understanding the performance of thin-client gaming," in *Communications Quality and Reliability (CQR), 2011 IEEE International Workshop Technical Committee on*, 2011, pp. 1-6.
- [12] A. L. Yarbus, *Eye movements during perception of complex objects*: Springer, 1967.
- [13] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, pp. 185-207, 2013.
- [14] L. E. Nacke, A. Drachen, K. Kuikkaniemi, J. Niesenhaus, H. J. Korhonen, v. d. W. Hoogen, *et al.*, "Playability and player experience research," in *Proceedings of DiGRA*, 2009.
- [15] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," 2009.
- [16] J.-S. Lee and T. Ebrahimi, "Perceptual video compression: A survey," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 6, pp. 684-697, 2012.
- [17] C.-C. Ho, J.-L. Wu, and W.-H. Cheng, "A practical foveation-based rate-shaping mechanism for MPEG videos," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, pp. 1365-1372, 2005.

- [18] W. S. Geisler and J. S. Perry, "Real-time foveated multiresolution system for low-bandwidth video communication," in *Photonics West'98 Electronic Imaging*, 1998, pp. 294-305.
- [19] O. V. Komogortsev and J. I. Khan, "Predictive real-time perceptual compression based on eye-gaze-position analysis," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 4, p. 23, 2008.
- [20] J. Balle, A. Stojanovic, and J.-R. Ohm, "Models for static and dynamic texture synthesis in image and video compression," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, pp. 1353-1365, 2011.
- [21] M. Bosch, F. Zhu, and E. J. Delp, "Segmentation-based video compression using texture and motion models," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, pp. 1366-1377, 2011.
- [22] P. Ndjiki-Nya, D. Bull, and T. Wiegand, "Perception-oriented video coding based on texture analysis and synthesis," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, 2009, pp. 2273-2276.
- [23] F. Zhang and D. R. Bull, "A parametric framework for video compression using region-based texture models," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, pp. 1378-1392, 2011.
- [24] X. Yang, W. Lin, Z. Lu, X. Lin, S. Rahardja, E. Ong, *et al.*, "Rate control for videophone using local perceptual cues," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, pp. 496-507, 2005.
- [25] Z. Wang and A. C. Bovik, "Embedded foveation image coding," *Image Processing, IEEE Transactions on*, vol. 10, pp. 1397-1410, 2001.
- [26] M. E. Raichle, "The brain's dark energy," *Scientific American*, vol. 302, pp. 44-49, 2010.
- [27] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, pp. 1254-1259, 1998.
- [28] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural networks*, vol. 19, pp. 1395-1407, 2006.
- [29] N. Riche, M. Mancas, B. Gosselin, and T. Dutoit, "Rare: A new bottom-up saliency model," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, 2012, pp. 641-644.
- [30] D. Gao, V. Mahadevan, and N. Vasconcelos, "The discriminant center-surround hypothesis for bottom-up saliency," in *Advances in neural information processing systems*, 2008, pp. 497-504.
- [31] N. Bruce and J. Tsotsos, "Attention based on information maximization," *Journal of Vision*, vol. 7, pp. 950-950, 2007.
- [32] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 409-416.
- [33] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, 2006, pp. 545-552.
- [34] W. Wang, Y. Wang, Q. Huang, and W. Gao, "Measuring visual saliency by site entropy rate," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 2368-2375.

- [35] M. Cerf, J. Harel, W. Einhäuser, and C. Koch, "Predicting human gaze using low-level saliency combined with face detection," in *Advances in neural information processing systems*, 2008, pp. 241-248.
- [36] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs to model saliency in images," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 1698-1705.
- [37] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, pp. 1915-1926, 2012.
- [38] A. Borji and L. Itti, "Exploiting local and global patch rarities for saliency detection," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 478-485.
- [39] T. N. Vikram, M. Tscherepanow, and B. Wrede, "A saliency map based on sampling an image into random rectangular regions of interest," *Pattern Recognition*, vol. 45, pp. 3114-3124, 2012.
- [40] T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," 2012.
- [41] J. Zhang and S. Sclaroff, "Saliency detection: A boolean map approach," in *Computer Vision (ICCV), 2013 IEEE International Conference on*, 2013, pp. 153-160.
- [42] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Computer Vision, 2009 IEEE 12th international conference on*, 2009, pp. 2106-2113.
- [43] E. Vig, M. Dorr, and D. Cox, "Large-scale optimization of hierarchical features for saliency prediction in natural images," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, 2014, pp. 2798-2805.
- [44] H. R. Tavakoli, E. Rahtu, and J. Heikkilä, "Fast and efficient saliency detection using sparse sampling and kernel density estimation," in *Image Analysis*, ed: Springer, 2011, pp. 666-675.
- [45] E. Erdem and A. Erdem, "Visual saliency estimation by nonlinearly integrating features using region covariances," *Journal of vision*, vol. 13, p. 11, 2013.
- [46] R. J. Peters and L. Itti, "Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, 2007, pp. 1-8.
- [47] J. Wu, C. Yuen, N.-M. Cheung, J. Chen, and C. W. Chen, "Enabling adaptive high-frame-rate video streaming in mobile cloud gaming applications," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 25, pp. 1988-2001, 2015.
- [48] H.-J. Hong, C.-F. Hsu, T.-H. Tsai, C.-Y. Huang, K.-T. Chen, and C.-H. Hsu, "Enabling adaptive cloud gaming in an open-source cloud gaming platform," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 25, pp. 2078-2091, 2015.
- [49] M. Semsarzadeh, A. Yassine, and S. Shirmohammadi, "Video Encoding Acceleration in Cloud Gaming," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 25, pp. 1975-1987, 2015.
- [50] Y. Liu, S. Dey, and Y. Lu, "Enhancing Video Encoding for Cloud Gaming Using Rendering Information," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 25, pp. 1960-1974, 2015.

- [51] W. Cai, Z. Hong, X. Wang, H. C. Chan, and V. Leung, "Quality-of-Experience Optimization for a Cloud Gaming System With Ad Hoc Cloudlet Assistance," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 25, pp. 2092-2104, 2015.
- [52] H. Tian, D. Wu, J. He, Y. Xu, and M. Chen, "On achieving cost-effective adaptive cloud gaming in geo-distributed data centers," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 25, pp. 2064-2077, 2015.
- [53] S.-P. Chuah and N.-M. Cheung, "Layered coding for mobile cloud gaming," in *Proceedings of International Workshop on Massively Multiuser Virtual Environments*, 2014, pp. 1-6.
- [54] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *Image Processing, IEEE Transactions on*, vol. 13, pp. 1304-1318, 2004.
- [55] M. Nyström and K. Holmqvist, "Effect of compressed offline foveated video on viewing behavior and subjective quality," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 6, p. 4, 2010.
- [56] Z. Li, S. Qin, and L. Itti, "Visual attention guided bit allocation in video compression," *Image and Vision Computing*, vol. 29, pp. 1-14, 2011.
- [57] M. Hemmati, A. Javadtalab, A. A. N. Shirehjini, S. Shirmohammadi, and T. Arici, "Game as Video: Bit Rate Reduction through Adaptive Object Encoding," presented at the NOSSDAV'13, Oslo, Norway, 2013.
- [58] S. Lee and A. C. Bovik, "Fast algorithms for foveated video processing," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, pp. 149-162, 2003.
- [59] J.-S. Lee, F. De Simone, and T. Ebrahimi, "Subjective quality evaluation of foveated video coding using audio-visual focus of attention," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, pp. 1322-1331, 2011.
- [60] Y. Liu, Z. G. Li, and Y. C. Soh, "Region-of-interest based resource allocation for conversational video communication of H. 264/AVC," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, pp. 134-139, 2008.
- [61] G.-L. Wu, Y.-J. Fu, S.-C. Huang, and S.-Y. Chien, "Perceptual quality-regulable video coding system with region-based rate control scheme," *Image Processing, IEEE Transactions on*, vol. 22, pp. 2247-2258, 2013.
- [62] D.-K. Kwon, M.-Y. Shen, and C. J. Kuo, "Rate control for H. 264 video with enhanced rate and distortion models," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, pp. 517-529, 2007.
- [63] C. An and T. Q. Nguyen, "Iterative rate-distortion optimization of H. 264 with constant bit rate constraint," *Image Processing, IEEE Transactions on*, vol. 17, pp. 1605-1615, 2008.
- [64] Z. Chen and C. Guillemot, "Perceptually-friendly H. 264/AVC video coding based on foveated just-noticeable-distortion model," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 20, pp. 806-819, 2010.
- [65] Z.-G. Li, F. Pan, K. P. Lim, G. Feng, X. Lin, and S. Rahardja, "Adaptive basic unit layer rate control for JVT," in *JVT-G012-r1, 7th Meeting, Pattaya II, Thailand*, 2003.
- [66] F. Li and N. Li, "Region-of-interest based rate control algorithm for H. 264/AVC video coding," *Multimedia Tools and Applications*, pp. 1-24, 2015.

- [67] M. Hrarti, H. Saadane, M.-C. Larabi, A. Tamtaoui, and D. Aboutajdine, "A macroblock-based perceptually adaptive bit allocation for H264 rate control," in *I/V Communications and Mobile Network (ISVC), 2010 5th International Symposium on*, 2010, pp. 1-4.
- [68] W. Lai, X.-D. Gu, R.-H. Wang, W.-Y. Ma, and H.-J. Zhang, "A content-based bit allocation model for video streaming," in *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on*, 2004, pp. 1315-1318.
- [69] H. Rahimi, A. Nazari Shirehjini, and S. Shirmohammadi, "Activity-centric streaming of virtual environments and games to mobile devices," in *Haptic Audio Visual Environments and Games (HAVE), 2011 IEEE International Workshop on*, 2011, pp. 45-50.
- [70] P. K. Mital, T. J. Smith, R. L. Hill, and J. M. Henderson, "Clustering of gaze during dynamic scene viewing is predicted by motion," *Cognitive Computation*, vol. 3, pp. 5-24, 2011.
- [71] J. L. Teeters, K. D. Harris, K. J. Millman, B. A. Olshausen, and F. T. Sommer, "Data sharing for computational neuroscience," *Neuroinformatics*, vol. 6, pp. 47-55, 2008.
- [72] (2015/08/08). *Tobii X2 Eye Tracker Portable Lab* [Online]. Available: <http://www.tobii.com/en/eye-tracking-research/global/products/hardware/tobii-x2-30-eye-tracker/>
- [73] (2015/08/08). *Tobii X2-30 Eye Tracker Accuracy and Precision Test Report* [Online]. Available: <http://www.tobii.com/Global/Analysis/Training/Metrics/Tobii%20X2-30%20Eye%20Tracker%20Accuracy%20and%20Precision%20Test%20Report.pdf>
- [74] (2015/12/19). *H.264/AVC Reference Software* [Online]. Available: <http://iphome.hhi.de/suehring/tml>
- [75] W. GAO and F. PAN, "Adaptive Rate Control with HRD Consideration, Document: JVT-H014," in *JVT 8th Meeting*, 2003.
- [76] R. J. Peters, A. Iyer, C. Koch, and L. Itti, "Components of bottom-up gaze allocation in natural scenes," *Journal of Vision*, vol. 5, pp. 692-692, 2005.
- [77] C. Reed. (2015/12/19). *The 5 Most Popular First-Person Shooter Video Games of All Time* [Online]. Available: <http://www.cheatsheet.com/technology/the-5-most-popular-first-person-shooters-of-all-time.html>
- [78] (2015/12/19). *How long is Call of Duty: Black Ops II? - HLTB* [Online]. Available: <http://howlongtobeat.com/game.php?id=1472>
- [79] M. Dick, O. Wellnitz, and L. Wolf, "Analysis of factors affecting players' performance and perception in multiplayer games," in *Proceedings of 4th ACM SIGCOMM workshop on Network and system support for games*, 2005, pp. 1-7.
- [80] T. Berger, "Rate-Distortion Theory," *Encyclopedia of Telecommunications*, 1971.
- [81] A. J. Viterbi and J. K. Omura, *Principles of digital communication and coding*: Courier Corporation, 2013.
- [82] B. A. Wandell, *Foundations of vision*: Sinauer Associates, 1995.
- [83] M. R. Hosseinzadeh Taher, H. Ahmadi, and M. R. Hashemi, "Power-aware analysis of H. 264/AVC encoding parameters for cloud gaming," in *Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on*, 2014, pp. 1-6.

-
- [84] (2015/08/08). *GStreamer: open source multimedia framework* [Online]. Available: <http://gstreamer.freedesktop.org/>
- [85] M. Kartavenkov. (2015/08/08). *DirectShow Virtual Video Capture Source Filter in C#* [Online]. Available: <http://www.codeproject.com/Articles/437617/DirectShow-Virtual-Video-Capture-Source-Filter-in>
- [86] M. Kartavenkov. (2015/08/08). *H.264 CUDA Encoder DirectShow Filter in C#* [Online]. Available: <http://www.codeproject.com/Articles/421869/H-CUDA-Encoder-DirectShow-Filter-in-Csharp>
- [87] G. Davies. (2015/08/08). *DirectShow RTSP Server filter and RTSP Jukebox* [Online]. Available: <http://www.gdcl.co.uk/2013/05/16/RTSP-Jukebox.html>
- [88] (2015/08/08). *AngularJS — Superheroic JavaScript MVW Framework* [Online]. Available: <https://angularjs.org/>
- [89] (2015/08/08). *ngTouch* [Online]. Available: <https://docs.angularjs.org/api/ngTouch>
- [90] J. MORONY. (2015/08/08). *The Step-by-Step Guide to Publishing a HTML5 Mobile Application on App Stores* [Online]. Available: <http://www.joshmorony.com/the-step-by-step-guide-to-publishing-a-html5-mobile-application-on-app-stores/>

واژه‌نامه انگلیسی به فارسی

واژه‌هایی که در کنار آن‌ها علامت • قرار داده شده است، معادل فارسی آنها از پایگاه اطلاع‌رسانی فرهنگستان زبان و ادب فارسی برگرفته شده است.

Add-on	افزونه
Application Programming Interface (API)	واسط برنامه‌نویسی کاربرد
Attention Mechanism	سازوکار توجه
Basic Unit	واحد پایه
Benchmark	محک
Blur	مات
Boolean	دودویی
Bottom-up attention	توجه پایین-به-بالا
Broadcast	پخش جمعی
• Calibration	واسنجی
Cerebral cortex	قشر مغز
Chin Rest Eye Tracker	دستگاه ردیاب چشم چانه-ثابت
Chroma subsampling	نمونه‌گیری رنگ
• Client	کارخواه
• Client-Server	کارخواه-کارساز
Cloud Gaming (CG)	بازی های مبتنی بر رایانش ابری
Coding Units	واحدهای کدشونده
Collectable Items	موارد جمع‌شدنی
Contrast	تضاد
Curve	منحنی
Curve fitting	برازش منحنی
Cut-off	برش
Deblocking Filter	فیلتر رفع بلاک
• Decode	کدگشایی کردن
Dot Per Inch (DPI)	نقطه بر اینچ
Eccentricity	زاویه از محور دید
• Encode	کدبندی کردن
Entropy Coding	کدکردن بی‌نظمی
Entropy Variance	بی‌نظمی وردایی
Exponential	نمایی
Eye Tracker Device	دستگاه ردیاب چشم
False positive	مثبت-غلط
Feedback	بازخورد
Field of View	حوزه‌ی دید

Finger touch	لمس انگشت
Fixation Point	نقطه تمرکز
Fovea	گودی مرکزی شبکیه
Free Viewing	تماشای آزادانه
Full-reference	منبع-دسترس پذیر
Game engine	موتور بازی
Gaze Point	نقطه خیرگی
Grayscale map	نگاشت سطوح خاکستری
Hardware Abstraction Layer (HAL)	لایه انتزاعی سخت افزار
Head Mounted Eye Tracker	دستگاه ردیاب چشم سر-بند
• Histogram	بافت نگاشت
Hooking Technique	فن قلاب
Human Visual Sensitivity	حساسیت بصری انسان
Human Visual System (HVS)	سیستم بینایی انسان
Human-Computer Interaction (HCI)	ارتباط انسان و رایانه
• Jitter	لغزش
• Joystick	اهرمک
Linear Least-Squares	کمینه مربعات خطی
Lossless coding	کد کردن بدون اتلاف
Mean Opinion Score (MOS)	متوسط امتیاز نظر
Mean Squared Error	میانگین مربعات خطا
Module	پیمانه
Multi objective optimization	بهینه سازی چند هدفه
Network Access Layer	لایه دسترسی شبکه
No-reference	منبع-بدون دسترس
Objective	کمی
Open source	متن باز
Orientation	جهت
Perceived Quality	کیفیت ادراکی
• Perception	ادراک
Perceptual Redundancy	افزونگی ادراکی
Perceptual Transcode	باز کد کردن ادراکی
Perceptual Video Coding	کد کردن ادراکی ویدئو
Peripheral Vision	دید جانبی
Plug-in	وصله
Preattentive	پیش توجه
Principle Component Analysis (PCA)	تحلیل مولفه های اصلی
Psychophysical	روان فیزیکی
Quantization	پیمانه بندی کردن

Raster Order Scan	ترتیب پیمایش سطری
Rate-Distortion Optimization (RDO)	بهینه‌سازی نرخ-اعوجاج
Reduced-reference	منبع - نیمه‌دسترس پذیر
• Redundancy	افزونگی
Region Of Interest (ROI)	ناحیه مورد علاقه
Remote Eye Tracker	دستگاه ردیاب چشم راه‌دور
Render	به تصویر کشیدن
Resolution	دقت مکانی
Retina	شبکیه
Saliency	برجستگی
Salient Area Percentage (SAP)	درصد مساحت پراهمیت
• Scrolling	نَورَدش
• Server	کارساز
Side Scrolling Shooter	شلیکگر نَورَدش افقی
Spatial Frequency	فرکانس مکانی
Statistical Redundancy	افزونگی آماری
Stimulus-driven attention	توجه تحریک محور
Subjective	کیفی
Subjective test	ارزیابی کیفی
Surveillance Video	ویدئوی نظارتی
Task-driven attention	توجه وظیفه محور
Telemedicine	پزشکی از راه دور
• Texture	بافت
• Thread	ریسه
Thresholding	حد آستانه گرفتن
Top-down attention	توجه بالا-به-پایین
• Variance	وردایی
Video Streaming	جویبارسازی ویدئو
Visual Attention Model	مدل توجه بصری
Visual Search	جستجوی بصری

Abstract

Flourishing of cloud gaming (CG) and fulfilling its ultimate goal, playing video games wherever, whenever and on every devices, would hardly be possible until its required high bandwidth is reduced. Although conventional video encoders try to remove both statistical and perceptual redundancies, due to high complexity of human visual system (HVS) and content diversity in different applications, identification and deletion of perceptual redundancies still need more research. Developing perceptual models, especially visual attention models, for cloud gaming is very challenging because of unique content of video games, such as their fantasy characters, object placements and complex game logic and design for the sake of immersion and satisfaction. Difference among players' skill levels, playing habits and strategies is another challenge of developing such models. Therefore, the first step of employing perceptual coding in bandwidth reduction of cloud gaming is to develop models which can overcome the aforementioned challenges. It also needs a dataset including recorded players' gaze locations and other data during their gameplay. Since there is not an appropriate game-specific dataset or model yet, in this thesis we work on them. Doing so, two datasets and two perceptual models are proposed. The first dataset includes a variety of video games and their objects. The model built on this dataset is grounded on visual attention mechanism and predicts the player's gaze location based on a combination of low level signal properties and game object prioritization. Experimental results show that this model decreases the required bit rate by nearly 25% on average, while maintaining a relatively high user quality of experience. The second perceptual model, addresses the difference among attention patterns of the players. To develop this model, the recorded eye-tracking data is first clustered. Then, the correlation of clusters and skill levels are shown via statistical and experimental methods. Our analyses show that this model decreases the bandwidth by up to 15% based on the player's skill. The second step is to incorporate the perceptual models into the video encoder by means of perceptual rate-distortion models to assign bits to each region of the video according to its importance to HVS. Since current attention-based bit allocation algorithms do not take other HVS properties into account, in some cases the amount of distortion in less important areas distracts the players and consequently lowers the user perceived quality. Therefore, a new model is proposed which controls the amount of attention in each region based on its distance to important areas by considering both attention and fovea mechanisms. This model results in better user perceived quality (20% increase in mean opinion score). Another contribution of this thesis is to develop a cloud gaming testbed to boost further researches pertaining to cloud gaming.

Keywords: Cloud Gaming, Video Encoding, Perceptual Video Encoding, Visual Attention Model



University of Tehran

College of Engineering

School of Electrical and Computer Engineering

A Content-aware Bitrate Controller for Cloud Gaming

By

Hamed Ahmadi

Supervisor

Dr. Mahmoud Reza Hashemi

Advisor

Dr. Shervin Shirmohammadi

A thesis submitted to the graduate studies office
in partial fulfillment of the requirements for
the degree of Ph.D. in
Information Technology

March 2016