دانشکده مهندسی

پایان‌نامه کارشناسی ارشد در رشته مهندسی کامپیوتر (هوش مصنوعی)

یادگیری ویژگی های مکانی و زمانی برای تشخیص فعالیت انسان

توسط:

نجمه هادی برحق طلب

استاد راهنما:

دکتر زهره عظیمی فر

شهریور ۱۳۹۲

به نام خداوند دوست

که هر چه دارم از اوست

به نام خدا

اظهارنامه

اینجانب ..... (.....) دانشجوی رشته ی ..... گرایش ..... دانشکده‌ی ..... اظهار می کنم که این پایان‌نامه حاصل پژوهش خودم بوده و در جاهایی که از منابع دیگران استفاده کرده‌ام، نشانی دقیق و مشخصات کامل آن را نوشته‌ام. همچنین اظهار می‌کنم که تحقیق و موضوع پایان نامه‌ام تکراری نیست و تعهد می نمایم که بدون مجوز دانشگاه دستاوردهای آن را منتشر ننموده و یا در اختیار غیر قرار ندهم. کلیه حقوق این اثر مطابق با آیین نامه مالکیت فکری و معنوی متعلق به دانشگاه شیراز است.

نام و نام خانوادگی:

تاریخ و امضا:

به نام خدا

یادگیری ویژگی های مکانی و زمانی برای تشخیص فعالیت انسان

به کوشش

نجمه هادی برحق طلب

پایان‌نامه

ارائه شده به تحصیلات تکمیلی دانشگاه شیراز به عنوان بخشی

از فعالیت‌های تحصیلی لازم برای اخذ درجه کارشناسی ارشد

در رشته‌ی:

مهندسی کامپیوتر – هوش مصنوعی

از دانشگاه شیراز

شیراز

جمهوری اسلامی ایران

ارزیابی کمیته پایان‌نامه، با درجه‌ی: عالی

دکتر زهره عظیمی فر، استادیار بخش مهندسی و علوم کامپیوتر (استاد راهنما) .....

دکتر رضا بوستانی، استادیار بخش مهندسی و علوم کامپیوتر (استاد مشاور) .....

دکتر ستار هاشمی، استادیار بخش مهندسی و علوم کامپیوتر (داور متخصص داخلی) .....

شهریور ماه ۱۳۹۲

تقدیم به

آنان که دوستشان دارم.

سپاسگزاری

سپاس خدای را که سخنوران، در ستودن او بمانند و شمارندگان، شمردن نعمت های او ندانند و کوشندگان، حق او را گزاردن نتوانند. و سلام و دورد بر محمّد و خاندان پاک او، طاهران معصوم، هم آنان که وجودمان وامدار وجودشان است؛ و نفرین پیوسته بر دشمنان ایشان تا روز رستاخیز...

بدون شک جایگاه و منزلت معلم، اجّل از آن است که در مقام قدردانی از زحمات بی شائبه-ی او، با زبان قاصر و دست ناتوان، چیزی بنگاریم. اما از آنجایی که تجلیل از معلم، سپاس از انسانی است که هدف و غایت آفرینش را تامین می کند و سلامت امانت هایی را که به دستش سپرده اند، تضمین؛ بر حسب وظیفه و از باب "من لم یشکر المنعم من المخلوقین لم یشکر اللّه عزّ و جلّ": ازپدر و مادر عزیزم...این دو معلم بزرگوار که همواره بر کوتاهی و درشتی من، قلم عفو کشیده و کریمانه از کنار غفلت هایم گذشته اند و در تمام عرصه های زندگی یار و یاوری بی چشم داشت برای من بوده اند؛ از استاد با کمالات و شایسته؛ سرکار خانم دکتر زهره عظیمی فر که در کمال سعه صدر، با حسن خلق و فروتنی، از هیچ کمکی در این عرصه بر من دریغ ننمودند و زحمت راهنمایی این رساله را بر عهده گرفتند؛ و از استاد فرزانه و دلسوز؛ جناب آقای دکتر بوستانی که زحمت مشاوره این رساله را متقبل شدند؛ از استاد صبور و با تقوا، جناب آقای دکتر هاشمی، مدیریت محترم کرسی گروه، که زحمت داوری این رساله را متقبل شدند؛ کمال تشکر و قدردانی را دارم. باشد که این خردترین، بخشی از زحمات آنان را سپاس گوید.

در پایان به پاس قدر دانی از قلبی آکنده از عشق و معرفت که محیطی سرشار از سلامت و امنیت و آرامش و آسایش برای من فراهم آورده است؛ همدلی که با واژه‌ی نجیب و مغرور تلاش آشنایی دارد و تلاش راستین را می شناسد و عطر رؤیایی آن را استشمام می کند و مرا

در راه رسیدن به اهداف عالی یاری می رساند؛ همو که حس تعهد و مسئولیت را در زندگی مان تلالویی خدایی داده است؛ این پایان نامه تقدیم همسر مهربانم می گردد.

چکیده

یادگیری ویژگی های مکانی و زمانی برای تشخیص فعالیت انسان

به کوشش

نجمه هادی برحق طلب

امروزه با افزایش رو به رشد دوربین های دیجیتال، روزانه حجم عظیمی از داده های ویدئویی تولید می شوند که بر روی وب و پایگاه های داده ی بزرگ قابل دسترس هستند. دسته بندی این داده ها بر اساس محتوایشان به صورت اتوماتیک و با دقت بالا هدف غایی بسیاری از کاربردها در این زمینه است. در این پایان نامه ما سیستمی طراحی کرده ایم که بتواند داده های ویدئویی حاوی عمل انسان را با دقت خوب و در زمان نسبتا کوتاهی برچسب گذاری کرده و دسته بندی کند. در این پایان نامه از روش های دیداری بهره گرفته شده است. روش های دیداری برای تشخیص فعالیت انسان در واقع فرایند تشخیص انسان در ویدئو و عملی که انجام می دهد با بهره گیری از تکنیک های بینایی ماشین می باشد.

در این راستا سیستم طراحی شده از روش یادگیری بدون نظارت برای استخراج ویژگی و کد کردن آن ها استفاده کرده است. روش های یادگیری ویژگی نه تنها به راحتی قابل تعمیم به حوزه ههای دیگر هستند بلکه کارایی قابل توجهی بر روی داده های ویدئویی جمع آوری شده از دنیای واقعی دارند.

در پایان برای ارزیابی این سیستم به دسته بندی داده های ویدئویی پایگاه های داده‌ی KTH، UCF و YouTube پرداخته ایم و به نتایج قابل توجهی دست پیدا کرده ایم.

کلمات کلیدی: تشخیص عمل، ویژگی محلی، تحلیل زیر فضای مستقل، کدگذاری خطی محلی، ادغام سه بعدی.

فهرست مطالب

عنوان صفحه

فهرست جدول‌ها

فهرست شکل‌ها

فصل اول

مقدمه

۱- مقدمه

در این فصل به شرح کلیاتی درباره ی انگیزه ی انتخاب موضوع، تشخیص عمل انسان و همچنین شرحی بر مسئله و کاربردها و چالش های آن ارائه می شود. در انتهای فصل نیز توضیحاتی درباره ی نوع داده های در دسترس جهت تست سیستم های تشخیص عمل انسان داده می شود.

۱-۱- مقدمه

در طول چندین دهه‌ی گذشته، کامپیوترها و شبکه های جهانی زندگی ما را به شدت تحت تأثیر قرار داده اند. کامپیوترها کارهای تکراری و محاسباتی وسیعی انجام می دهند و امکان ارتباطات را گسترش می دهند. همراه با پیشرفت عمومی تکنولوژی در زمینه‌ی کامپیوتر، داده های ویدئویی بیش از پیش قابل دسترس بوده و نقش مهم و فزاینده ای در زندگی روزمره‌ی ما ایفا می کنند. امروزه حتی سخت افزارهای مصرفی متداول مانند نوت بوک ها، تلفن های همراه و دوربین های دیجتال داده های ویدئویی تولید می کنند. به طور همزمان دسترسی سریع تر به اینترنت و افزایش ظرفیت ذخیره سازی به ما این امکان را می دهد که داده های ویدئویی را منتشر کرده و با دیگران به اشتراک بگذاریم .

به طور مثال ۳۶ میلیون کاربر اینترنت در آلمان (٪۴۴ از جمعیت آلمان) بیش از ۶ بیلیون ویدئو را در آگوست ۲۰۰۹ به صورت آنلاین تماشا کرده‌اند و این آمار در مقایسه با آگوست ۲۰۰۸ رشد ۳۸ درصدی داشته است. میزان ویدئویی که در هر دقیقه روی سایت You Tube آپلود می شود در ماه می ۲۰۰۸ نسبت به اواسط سال ۲۰۰۷ از ۶ ساعت به ۲۰ ساعت افزایش یافته است و این به معنای افزایش ۳۳۰ درصدی در طول ۲ سال می باشد.

۱-۲- انگیزه

به هر حال بر خلاف افزایش اهمیت داده های ویدئویی امکان آنالیز آن ها به طور اتوماتیک بسیار محدود است. سیستم های کامپیوتری دیداری از توانایی های دیداری انسان خیلی عقب تر هستند. برای نمونه جستجوی یک ویدئو در میان حجم عظیم از داده های ویدئویی در حال حاضر فقط با برچسب گذاری پر هزینه‌ی دستی امکان پذیر می باشد. موتورهای جستجوی وب عموما برای بازیابی ویدئوهای مربوطه وابسته به داده های متنی مانند توصیف ها یا برچسب ها هستند .

مثال دیگر کاربرد های نظارتی می باشد. تا به امروز شهر لندن حدود یک میلیون دوربین های مدار بسته CCTV camera (closed circuit television) با هزینه ای بالغ بر ۲۰۰ میلیون پوند نصب کرده است. با این وجود آن ها فقط توانستند به ازای هر ۱۰۰۰ دوربین نظارتی به کشف یک جرم کمک کنند. بر اساس گزارشات داخلی دوربین های نظارتی CCTV منجر به هزینه های گزاف و حداقل بازدهی شده اند .

محققان بخش دولتی home office به این نتیجه رسیدند که CCTV تقریبا نتوانست به کاهش جرم کمکی کند و به نظر می رسد بیشترین تاثیر را در جلوگیری از تخلف وسایل نقلیه هنگام پارک کردن داشته است. در واقع با حجم عظیم داده های ویدئویی، گلوگاه اصلی ضرورت ذخیره و تحلیل دستی این داده ها می باشد .

یک حیطه ی کاربردی مهم که آنالیز ویدئو در آن به عنوان یک رابط پیچیده‌ی انسان و کامپیوتر بسیار مورد توجه می باشد بازی های کامپیوتری هستند. یک پروژه ی در حال انجام پروژه ی Natal شرکت میکروسافت می باشد. چهارچوب این پروژه قادر به گرفتن حرکات سه بعدی کل بدن، تشخیص چهره، تشخیص صدا و مکان یابی منبع تولید صدا می باشد. برای این کار باید اطلاعات بدست آمده از چندین سنسور با هم ترکیب شود. سنسورها شامل دوربین های ویدئویی، سنسور تشخیص عمق بر اساس الگوهای اشعه فروسرخ و یک میکروفون می باشد. بدین ترتیب کاربر می تواند بازی های ویدئویی را بدون هیچ ابزار کنترلی و با تعاملات دنیای دیداری به کمک کل اعضای بدنش به طور طبیعی انجام دهد. ضبط و گرفتن حرکات بازیگر برای شخصیت های کارتونی در فیلم های انیمیشینی (شکل۱-۱) و همچنین جلوه های

شکل ۱-۱ ضبط و گرفتن حرکات برای تولید فیلم انیمیشنی در استدیو

ویژه ی سینمایی به یک استاندارد بالفعل تبدیل شده است. با این حال، تحلیل حرکت انسان هم می تواند نقش مهمی در کاربردهای رسانه ای ایفا کند. همچنین در تحلیل و بهبود حرکات ورزشی ورزشکاران نقش تأثیر گذاری داشته باشد.

این مثال ها نشان می دهند امروزه نیاز گسترده ای به سیستم های کامپیوتری دیداری برای درک و پردازش داده های ویدئویی به طور اتوماتیک وجود دارد. آن ها همچنین نشان می دهند که تکنولوژی بینایی ماشین توانایی بالا و بالقوه ای بر روی آینده‌ی ما دارد. طراحی سیستم های هوشمندی که این پردازش ها را به صورت اتوماتیک انجام دهد کار بس ارزشمندی است که می تواند در کاربرد های صنعتی و تجاری مورد بهره برداری قرار گیرد.

از جمله دیگر کاربردها که نیاز به تحلیل و پردازش داده های ویدیوئی دارند تشخیص عمل[1] می باشد. شکل ۱-۲ نمونه ی تعدادی عمل را که از چندین فیلم استخراج شده را نشان می دهد. تشخیص عمل انسان می تواند در سیستم های نظارتی[2] مثل اتاق های کنترل هوشمند، تجزیه و تحلیل فیلم مسابقات ورزشی و کلیّه‌ی سیستم های هوشمند مبتنی بر تعامل انسان و ابزار های الکترونیکی مورد استفاده قرار گیرد. به طور مثال سیستم های نظارتی در اماکن عمومی مانند فرودگاه ها و ایستگاه های مترو می توانند اعمال غیرعادی و مشکوک از قبیل قرار دادن کیف در سطل زباله توسط یک شخص را به طور اتوماتیک تشخیص داده و به

[1] Action recognition
[2] Surveillance system

شکل ۱-۲ نمونه چند عمل در فیلم ها

شکل ۱-۳ کاربردهای تشخیص عمل انسان: الف) سیستم های نظارتی ب) دستگاه کینکت ج) حاشیه نویسی مسابقات ورزشی

قسمت امنیتی هشدار دهند. برای نمونه، کینکت[3] یک محصول تجاری است که از قدرت شناسایی عمل انسان بهره می برد. یک کنسول کنترلگر بازی از مایکروسافت که معمولا حرکات بدن انسان را دنبال می کند، سیستم حاشیه نویسی ورزشی می تواند تحلیل حرکت پیچیده بازیکن را انجام دهد و اطلاعات استراتژی بازی را از ویدیو پخش زنده بازی ورزشی به صورت بلادرنگ استخراج کند. شکل ۱-۳ برنامه های کاربردی مختلفی را معرفی می کند. شکل ۱-۳ (الف) یک مثال از سیستم نظارت تجاری را نشان می دهد. شکل ۱-۳ (ب) کنسول کنترلگر بازی براساس حرکت انسان را نشان می دهد که کینکت از شرکت مایکروسافت نامیده می شود. شکل ۱-۳ (ج) یک سیستم حاشیه نویسی ورزشی را نشان می دهد که برای هوشمندسازی یادداشت نویسی بازی های ورزشی استفاده می شود.

۱-۳- شرح مسئله

این پایان نامه بر روی مسئله ی تشخیص فعالیت انسان در ویدئو های واقعی مانند ویدئوهای تهیه شده از فیلم ها، اینترنت، دوربین های نظارتی و غیره متمرکز شده است. برای آن که موضوع را دقیق تر بررسی کنیم لازم است تعریف واژه های "عمل پایه"، "عمل" و "فعالیت" به روشنی بیان شود. زبان انسان از جمله ها تشکیل شده است که جمله به نوبه ی خود شامل فاعل، فعل و مفعول می باشد. برای توصیف مفهوم دیداری ویدئو وجود ساختاری شبیه به آنچه برای زبان انسان بیان شد ضروری می باشد. از دیدگاه الگوریتمی می توان تشخیص عمل انسان را چنین بیان کرد: تشخیص ۱) فاعل (کننده ی کار) که معمولا انسان ها هستند؛ ۲) مفعول که می تواند انسان های دیگر باشد یا محیطی که فاعل کار را در آن انجام می دهد؛ ۳) فعل عملی را که فاعل انجام می دهد توصیف می کند، همچنین می تواند توصیف تعامل بین مفعول و فاعل باشد. به این معنا عمل می تواند در بازه ی کوتاه زمانی انجام شود. همچنین می شود به رخدادی که در بازه ی زمانی طولانی تر انجام می شود اطلاق شود.

[3] Kinect