

به نام خدا

پروژه پایانی درس پردازش تصویر

تهیه کننده:

رحیم انتظاری

۹۳۷۲۲۰۲۵

زیر نظر:

آقای دکتر فتحی

مقدمه

این پروژه شامل پیدا کردن جهت سر ، در ویدئوی ورودی می شود. پیدا کردن جهت سر کاربرد های فراوان از جمله در پیدا کردن جهت نگاه دارد. خروجی این پروژه می تواند در کنار خروجی جهت چشم، به طور ترکیبی جهت نگاه را مشخص کنند.

مجموعه دادگان در این پروژه، مجموعه دادگان HPEG است. (A Natural Head Pose and Eye Gaze Dataset) این مجموعه دارای دو Session است. در این پروژه از session a استفاده شده است. session b شامل جهت نگاه نیز می شود.

در session a هر ویدئو ۶ ثانیه طول دارد. برای هر ویدئو فایلی به نام data وجود دارد شامل زاویه سر.(yaw).

استخراج ویژگی

اولین گام در این پروژه ،استخراج ویژگی است. در این مرحله از مقاله زیر استفاده شده است:

Head Pose Estimation Using Histogram of SIFT Descriptors

در این مقاله از SIFT به عنوان ویژگی استفاده شده است.

برای استخراج ویژگی ها جهت استفاده در مرحله آموزش(مدل گرافیکی) از مجموعه دادگان، به ترتیب مراحل زیر انجام شده است:

۱. فریم کردن هر ویدئو
۲. تغییر نام ویدئو ها
۳. استخراج صورت از هر ویدئو
۴. استخراج ویژگی ها : اعمال SIFT و BOW

در این بخش گام های ذکر شده به ترتیب توضیح داده شده است:

۱. فریم کردن هر ویدئو

برای این منظور در سیستم عامل لینوکس اوبونتو از دستور زیر استفاده می کنیم:

```
ffmpeg -i 1_a.avi -r 30 -ss 1 -f image2 %5d.jpg
```

که در آن ۳۰ نشان دهنده تعداد فریم در ثانیه است و خروجی به صورت عدد ۵ رقمی با پسوند jpg است.

برای نصب ffmpeg در لینوکس اوبونتو نیز از این لینک استفاده شده است:

<http://trac.ffmpeg.org/wiki/CompilationGuide/Ubuntu#ffmpeg>

۲. تغییر نام ویدئو ها

به منظور استفاده از کد BOW در مراحل بعدی، کپی فریم های بدست آمده را یک بار تغییر نام داده و در یک پوشه همه آن ها را قرار

می دهیم.

برای تغییر نام از کد پوسته زیر استفاده کرده ایم:

```
#bin/bash
```

```
i=1
```

```
find -type f -iname "*.jpg" | while read x; do
```

```
mv "${x}" "${printf "%05d" ${i}}.jpg"
```

```
i=$((i+1))
```

```
Done
```

این کد تمامی فایل های موجود در یک پوشه را که پسوند jpg دارند یافته و سپس آن ها را از شماره i به بعد شماره گذاری می کند. این کار را برای همه پوشه های مجموعه دادگان انجام می دهیم. حاصل پوشه ای است شامل همه ی فریم های همه ویدئو ها در یک پوشه که نام آن ها اعداد ۵ رقمی از ۱ تا ۱۸۳۵ است.

۳. استخراج صورت از هر ویدئو

در این گام برای استخراج ویژگی ها ابتدا باید صورت ها را از فریم ها بدست آوریم. کد C++ به نام facedet2.cpp این کار را انجام می دهد

۴. استخراج ویژگی ها : اعمال SIFT و BOW

در این مرحله باید برای تصاویر بدست آمده ابتدا ویژگی SIFT را استخراج کنیم. بدین منظور از کد BOW1.cpp استفاده کرده ایم. در این کد مانند مقاله از پنجره لغزانی استفاده شده است که ۸*۸ بوده و به اندازه ۴ پیکسل همپوشانی دارد. تقریباً ۳۷۰۰۰ تصویر با این کار ایجاد می شود. سپس از هر کدام از این تصاویر SIFT (بردار ۱۲۸ تایی) استخراج می کنیم.

خروجی این مرحله اولاً تصاویری است که نقاط SIFT در آن ها نمایش داده شده است.

ثانیاً فایلی است به نام dictionary.yml که در مرحله بعد از آن استفاده می شود. این فایل شامل مراکز کلاستر است. در اینجا تعداد کلاستر ها ۹ تا انتخاب شده است.

در گام دوم این مرحله، فریم های هر ویدئو به طور جداگانه به کد داده می شود. کد، ابتدا فایل dictionary.yml را خوانده، برای هر تصویر بررسی می کند که به کدام مرکز کلاستر نزدیکتر است. با استفاده از الگوریتم BOW، خروجی بدست آمده از این مرحله، برای هر ویدئو فایلی است به نام descriptor.yml و video.csv که برای هر فریم آن ویژگی های ۹ تایی (عضویت در کلاستر) بدست آمده است.

مدل گرافیکی

در این مرحله ویژگی های بدست آمده از مرحله قبل (یک فایل برای هر ویدئو شامل بردار های ۹ تایی برای هر فریم) به عنوان ورودی مدل گرافیکی به مدل CRF داده می شود. این مدل نمونه ای از مدل های بدون جهت محسوب می شود. در یک ویدئو مانند مثال های مجموعه دادگان استفاده شده، که از تعدادی فریم تشکیل شده است به هر کدام از فریم ها برچسبی نسبت داده می شود که مشخص کننده کنش انجام شده در آن فریم است.

همچنین برای یادگیری پارامترها (وزن توابع ویژگی) از روش L-BFGS استفاده شده است.

پیاده سازی با زبان C++ در سیستم عامل لینوکس و با استفاده از کتابخانه HCRF2.0b می باشد.

بعد از کامپایل برنامه در نهایت یک فایل اجرایی تولید می شود که با اجرای آن و انتخاب گزینه مورد نظر می توان آموزش و تست را انجام داد.

داده های آموزشی، داده های تست و برچسب های آن ها به ترتیب در فایل های dataTrain.csv ، dataTest.csv ، labelsTest.csv، labelsTrain.csv قرار دارند. (برای باز کردن این فایل ها از notepad از استفاده می کنیم، excel تغییراتی در محتوای فایل ایجاد می کند.)

برای انجام آموزش و تست، از دستور زیر استفاده می کنیم:

```
TestHCRF.exe [-t] [-T] [-d filename] [-l filename] [-D filename] [-L filename] [-m filename] [-f filename]
[-r filename] [-o cg|bfgs|asa] [-a crf|ldcrf|hcrf|ghcrf]
```

که به صورت زیر استفاده می شود:

```
testHCRF.exe -t -T -d dataTrain.csv -l labelsTrain.csv -D dataTest.csv -L labelsTest.csv -m model.txt -f
fetures.txt -r results.txt -c stats.txt -o lbfgs -a crf
```

آزمایشات و نتایج

برای آموزش هر ویدئو همراه با فایل برچسب ها به HCRF داده می شود.

در این پروژه مقادیر درجه جهت سر در مجموعه دادگان هر ۱۵ درجه، چندی^۱ شده است.

سپس از ۶ برچسب برای جهت سر استفاده شده است. نکته قابل توجه این است که برچسب ها باید اعداد مثبت باشند. برای این منظور زوایای ۴۵- تا ۴۵+، با فاصله ۱۵ درجه به ترتیب ۷ تا ۱۱ برچسب دارند.

خروجی مرحله قبل شامل فایلی است به نام stats.txt که نتیجه آزمون در آن ذخیره می شود.

در مرحله آزمایش از 10 cross fold استفاده شده است. یعنی از ۱۰ ویدئوی داده شده، ۹ تا برای آموزش و یکی برای تست استفاده شده است. این کار ۱۰ مرتبه با ویدئوهای متفاوت انجام شده است. نتیجه میانگین ۱۰ خروجی بدست آمده را نشان می دهد که دقت^۲ ۶۲٪ بدست آمده است.

دقت بدست آمده کافی به نظر نمیرسد. ممکن است ویژگی های بدست آمده مناسب نبوده باشند. هر چند که ویژگی ها بنا به مقاله ذکر شده انتخاب و استخراج شدند. حتی برای بهبود نتایج از پنجره لغزان استفاده شده است اما نتیجه بهتر نشد.

^۱ quantize

^۲ precision