دانشگاه فردوسی مشهد
دانشکده مهندسی - گروه کامپیوتر
آزمایشگاه بینایی ماشین

عنوان پروژه: MoSIFT Improvements

نام دانشجو: سیدحسام الدین حسینی

ID:4001366350

موعد تحویل: 1400/11/15

تاریخ تحویل:1400/11/15

خلاصه:

مقاله MoSIFT Recognizing Human Actions in Surveillance Videos مدلی را معرفی میکرد به عنوان MoSIFTکه با روشی الهام گرفته از SIFTمی کوشید فعالیت در حال انجام در ویدیو ها را شناسایی کند. روش شناسایی ویژگی ها و توصیف آنها برگرفته از SIFTبود، اما تغییراتی روی آن اعمال شده بود تا با در نظر گرفتن تغییرات در طول زمان ماهیت روش برای ویدیو مناسب تر شود .در این پروژه هدف بالا بردن دقت روش MoSIFTاست. یکی از این ایده ها اینست که از مرحله پیش پردازش اضافی داده های ورودی ارائه شود. در این راستا، معرفی یک مرحله اضافی از بهبود تصویر در الگوریتم MoSIFTبرای تشخیص اقدامات در ویدیو ضروری است. دومین ایده مورد نظر، MoSIFT را با PDI[1] ترکیب کنیم تا به نمایش کامل تری از اعمال(اکشن) انسان دست یابیم.

[1] positional distribution information (PDI)

شرح تکنیکال

در سال های اخیر، تشخیص اکشن(اعمال) انسان مبتنی بر بصری به تدریج به یک موضوع تحقیقاتی بسیار فعال تبدیل شده است. تجزیه و تحلیل اعمال انسان در ویدیوها به دلیل کاربردهایی مانند تعامل انسان و رایانه، بازیابی ویدیو مبتنی بر محتوا، نظارت بصری، تجزیه و تحلیل رویدادهای ورزشی، یک مشکل بسیار مهم در بینایی رایانه در نظر گرفته می شود. به دلیل پیچیدگی عمل، مانند پوشیدن و عادات مختلف بدن که منجر به مشاهدات متفاوت از یک عمل می شود، حرکت دوربین در محیط خارجی، تغییر نور، سایه ها، دیدگاه، این تأثیرات عوامل باعث تشخیص عمل می شود. هنوز یک پروژه چالش برانگیز است.

نمایش حرکت انسان در سکانس های ویدیویی برای تشخیص عمل بسیار مهم است. به غیر از وجود تمایز کافی بین دسته‌های مختلف، ویژگی‌های حرکت قابل اعتماد نیز برای مقابله با چرخش، تبدیل مقیاس، حرکت دوربین، پس‌زمینه پیچیده، سایه و غیره مورد نیاز است. در حال حاضر، متداول‌ترین ویژگی‌های مورد استفاده در تشخیص اعمال بر اساس حرکت، مانند جریان نوری ، مسیر حرکت ، یا بر اساس شکل ظاهری، مانند کانتور. ویژگی های قبلی به شدت تحت تأثیر روشنایی و سایه قرار می گیرند. ویژگی‌های اخیر به مکان‌یابی دقیق، تفریق پس‌زمینه یا ردیابی متکی هستند و نسبت به نویز، انسداد جزئی و تغییرات در دیدگاه حساس‌تر هستند.

SIFT روشی است که ویژگی های محلی تصویر را شناسایی و توصیف می کند. توصیفگرهای به‌دست‌آمده بر اساس این روش عبارتند از: تغییر ناپذیر مقیاس و تغییرناپذیر چرخش، تبدیل‌های وابسته به هم و نویز.

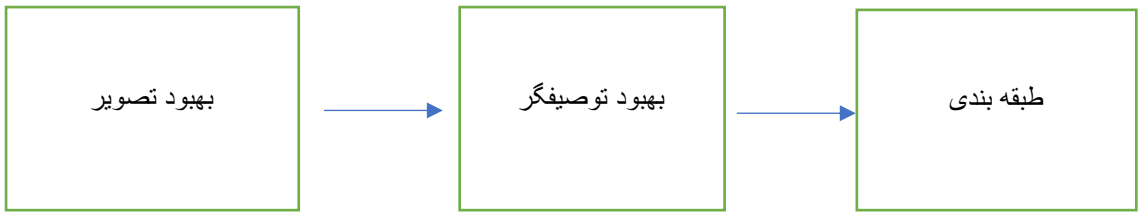
الگوریتم ها قادر به تشخیص و توصیف موقعیت مکانی-زمانی نقاط ویژگی هستند. آنها بر اساس سه مرحله اصلی هستند: تشخیص نقاط مورد علاقه، ساختن یک توصیفگر برای این نقاط و ایجاد یک طبقه بندی.

شرح نتایج

در این پروژه، رویکردی برای شناسایی اعمال انسان با استفاده از مرحله پیش پردازش اضافی داده های ورودی ارائه می شود. حجم زیادی اطلاعات ویدیویی همیشه اجازه نمی دهد تا کیفیت داده ها را در سطح بالایی پشتیبانی کند. این می تواند باعث ایجاد محدودیت در پردازش بیشتر داده های دیجیتال شود.

ایده اول)

- در این راستا، معرفی یک مرحله اضافی از بهبود تصویر در الگوریتم برای تشخیص اقدامات در ویدیو ضروری است. روش پیشنهادی شامل سه مرحله اصلی است: بهبود تصویر[2]، بهبود توصیفگر MoSIFT[3] و طبقه بندی[4]. مرحله بهبود تصویر ارائه شده بر اساس پردازش تصویر محلی و جهانی ترکیبی در حوزه فرکانس است. ایده اصلی در استفاده از روش ریشه‌یابی آلفای محلی[5]، اعمال آن بر روی بلوک‌های مختلف با اندازه‌های مختلف است. برای بهبود توصیفگر MoSIFT، از یک الگوریتم سه بعدی تفاوت چگال میکروبلاک (D DMD3)[6] استفاده می‌شود که با گرفتن دقیق میکروبلاک‌ها در هر ناحیه در جهت‌گیری‌ها و مقیاس‌های مختلف، نمایشی بسیار جهت‌دار از مناطق تصویر ارائه می‌دهد. D DMD3 دارای چندین مزیت نسبت به روش های دیگر است: راندمان بالاتر در مقایسه با روش های موجود. حداقل هزینه های محاسباتی هنگام استفاده از یک تصویر یکپارچه؛ بعد کم؛ سهولت اجرا؛ نیازی به تنظیمات ندارد. این بهبود ارائه شده اجازه می دهد تا بهره وری را 2-4٪ افزایش دهد.



[2] Image Enhancement
[3] Descriptor
[4] Classification
[5] local alfa-rooting method
[6] a three-dimensional microblock dense difference (3D DMD) algorithm

ایده دوم)

- بسیاری از روش‌های تشخیص اکشن موجود عمدتاً از توصیف‌گرهای مکانی-زمانی نقطه‌ی تکی استفاده می‌کنند، در حالی که اطلاعات یکپارچه بالقوه‌شان، مانند اطلاعات توزیع مکانی[7] را نادیده می‌گیرند.

با ترکیب ویژگی مکانی-زمانی محلی و اطلاعات توزیع موقعیت جهانی (PDI)[8] نقاط مورد علاقه، یک توصیفگر بهبود یافته در این پروژه پیشنهاد شده است.

روش پیشنهادی نقاط علاقه را با استفاده از روش تشخیص نقطه بهبودیافته شناسایی می کند. سپس، توصیفگر MoSIFT برای هر نقطه علاقه استخراج می‌شوند.

به منظور به دست آوردن یک توصیف فشرده و محاسبات کارآمد، روش تجزیه و تحلیل مؤلفه اصلی (PCA)[9] دو بار در توصیفگرهای SIFT سه بعدی فریم تک فریم[10] و چند فریم[11] استفاده می شود.

به طور همزمان، PDI نقاط مورد علاقه محاسبه شده و با ویژگی های فوق ترکیب می شود. ویژگی‌های ترکیبی کمی‌سازی و انتخاب می‌شوند و در نهایت با استفاده از الگوریتم تشخیص ماشین بردار پشتیبان[12] (SVM) روی مجموعه داده عمومی KTH آزمایش می‌شوند.

نتایج آزمایش نشان داده است که میزان تشخیص به طور قابل توجهی بهبود یافته است و ویژگی های پیشنهادی می توانند حرکت انسان را با سازگاری بالا با سناریوها با دقت بیشتری توصیف کنند.

[7] spatial distribution information.
[8] global positional distribution information (PDI)
[9] principal component analysis (PCA)
[10] single frame
[11] multiple frames
[12] support vector machine (SVM)

نقاط مورد علاقه:

در بینایی کامپیوتری، نقاط علاقه نشان دهنده مکانی است که دارای تغییرات شدید در ابعاد مکان و زمان است و برای عمل ثبت شده در یک ویدیو برجسته یا توصیف کننده در نظر گرفته می شود. در میان روش‌های مختلف تشخیص نقطه مورد علاقه، بیشترین مورد استفاده برای تشخیص اکشن(اعمال)، روشی است که مقادیر پاسخ تابع را بر اساس ترکیب فیلتر گابور و فیلتر گاوسی محاسبه می‌کند و مقادیر شدید پاسخ محلی را می‌توان به عنوان نقاط علاقه مکانی-زمانی در ویدیو در نظر گرفت.

به دلیل سایه و نویز ویدیو مستعد تشخیص نادرست است و نقاط جالب کاذب به راحتی در پس‌زمینه رخ می‌دهند. به ویژه در حرکت دوربین یا زوم دوربین بی تاثیر است. برخی از اشکالات در مثال ها به عنوان برش های مربع قرمز نشان داده شده در شکل ۲ مشخص شده است.

این اشکالات ناشی از کاستی‌های آشکارساز آن است، به‌ویژه فیلتر گابور که قابلیت استخراج را فقط در محور زمان دارد و در عین حال حرکت پویا در چشم‌انداز را نادیده می‌گیرد.

آشکارساز ما تشخیص قابل توجهی را تسهیل می کند و شامل سه مرحله زیر می شود:

مرحله 1. تشخیص قاب (فریم) برای شناسایی مناطق مورد علاقه در شکل ۳ به عنوان مثال نشان داده شده است.

مرحله 2. استفاده از فیلتر گبور۲بعدی برای تولید پنج جهت متفاوت (0 درجه، 22 درجه، 45 درجه، 67 درجه، 90 درجه) در شکل ۴ نشان داده شده است.

مرحله 3. فیلتر کردن در مناطق شناسایی شده مورد علاقه در مرحله 2.

شکل ۲ نمونه هایی از نتایج ما را نشان می دهد.

بازنمایی اعمال(رفتار):

نمونه را از مجموعه داده KTH انتخاب کرده ایم، سپس ناحیه توزیع نقاط مورد علاقه و ناحیه مکان بدنه در هر فریم شناسایی شده اند درشکل ۵ نشان داده ایم.

اگر تعداد نقاط علاقه در هر فریم N باشد، بعد ویژگی ها $256 \times N$ است تا اطلاعات مکانی-زمانی در این قاب را نشان دهد. بعد ویژگی آنقدر زیاد است که نمی تواند اطلاعات توزیع PDI به دست آمده را بطور منطقی ترکیب کند. بنابراین، کاهش ابعاد بر روی اطلاعات این قطعه انجام می شود. علاوه بر این، به منظور حذف داده‌های اضافی و مختصرتر کردن ویژگی‌ها، ویژگی‌های ترکیبی با استفاده از کوانتیزاسیون و انتخاب پردازش می‌شوند. پنج مرحله زیر ذکر شده است:

مرحله اول) کاهش ابعاد تک فریم

مرحله 2) کاهش ابعاد چند فریم

مرحله 3) ترکیب ویژگی ها

مرحله۴)کوانتیزاسیون ویژگی

مرحله۵)انتخاب ویژگی

تجزیه و تحلیل نتایج

آزمایش‌هایی روی مجموعه داده KTH با ویژگی مکانی-زمانی بهبود یافته انجام می‌شود. با مقایسه با مقاله قبلی MoSIFT مرتبط با ویژگی ها و مجموعه داده های مرتبط، عملکرد برجسته الگوریتم پیشنهادی در این بخش نشان داده شده است.

SVM به عنوان طبقه‌بندی داده‌های روش یادگیری آماری، دارای تفسیر هندسی بصری و قابلیت تعمیم خوبی است، بنابراین در تشخیص الگوی بصری محبوبیت پیدا کرده است.

بر اساس این تئوری، SVM از تئوری به حداقل رساندن ریسک ساختاری، که به صورت نشان داده شده است، توسعه یافته است.

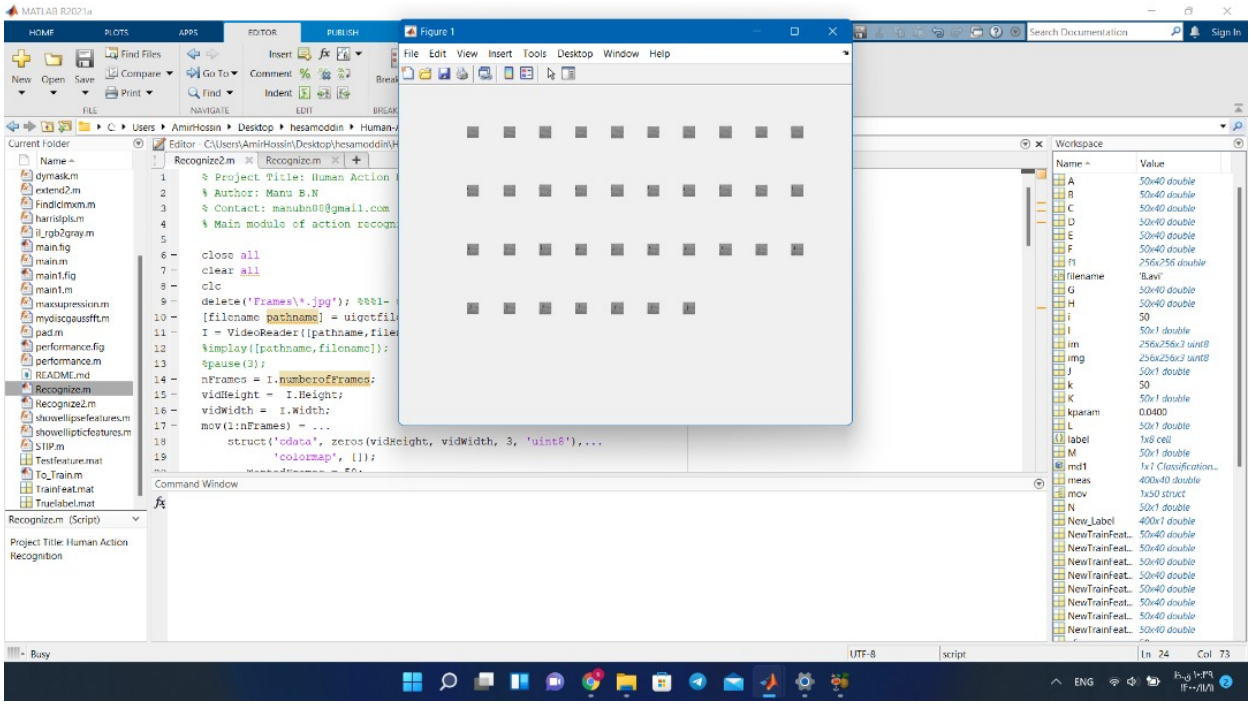
Scenario	MoSIFT	PDI	MoSIFT+PDI
outdoors (SC1)	0.9600	0.8000	0.9600
outdoors with scale variations (SC2)	0.8867	0.8268	0.9200
outdoors with different clothes (SC3),	0.8542	0.7569	0.9167
Indoors with lighting variations (SC4)	0.9600	0.9000	0.9600

نتایج نشان می‌دهد که ویژگی MoSIFT توانایی تشخیص بهتری نسبت به ویژگی PDI دارد. SC1 و SC4 پایدارتر از دو سناریو دیگر هستند. با استفاده از MoSIFT و ویژگی های ترکیبی (D SIFT+PDI3) به میزان تشخیص (96٪) رسیده ایم. همچنین نشان می دهد که MoSIFT سازگاری و استحکام خوبی نسبت به جهت حرکت، موقعیت، سرعت و غیره دارد.

نرخ تشخیص با استفاده از ویژگی های ترکیبی (MoSIFT SIFT+PDI) در مقایسه با MoSIFT بسیار افزایش می یابد.

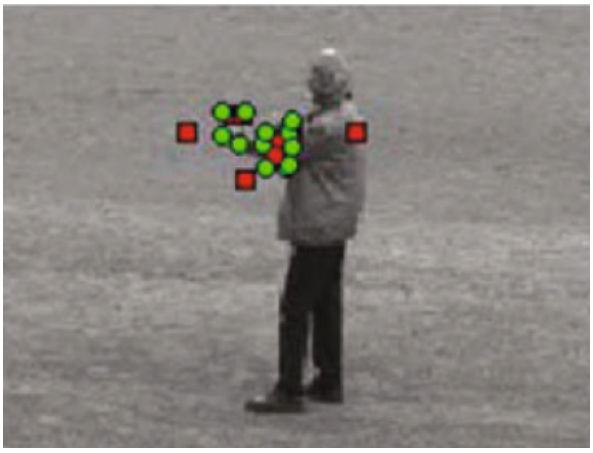
نتایج

شکل۱- دنباله ویدیو ها را نشان می دهد.



شکل۱- دنباله ای از ویدیو

شکل۲- نتایج تشخیص (نقاط سبز)، برخی از اشکالات در مثال ها به عنوان برش های مربع قرمز نشان داده شده است.



شکل۲- نتایج تشخیص

شکل۳- تشخیص قاب (فریم) برای شناسایی مناطق مورد علاقه

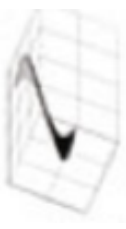
شکل۳- تشخیص قاب (فریم)

شکل۴- استفاده از فیلتر گبور۲بعدی برای تولید پنج جهت متفاوت (0 درجه، 22 درجه، 45 درجه، 67 درجه، 90 درجه)

صفر درجه

۲۲ درجه

۴۵ درجه



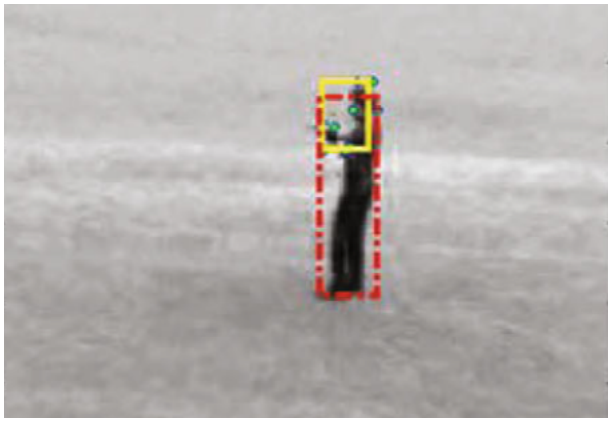
۶۷ درجه

۹۰ درجه

شکل۴- پنج جهت مختلف با استفاده از گبور فیلتر دو بعدی

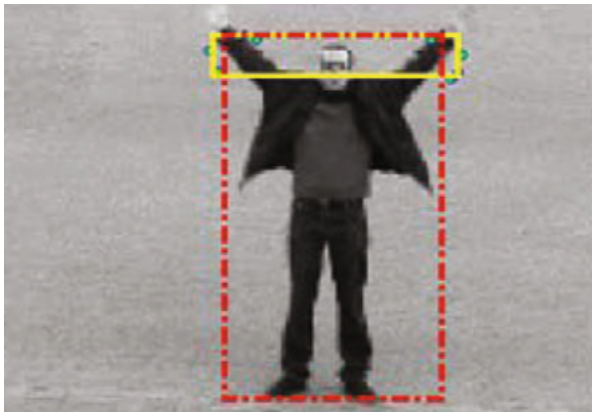
شکل۵- شناسایی ناحیه در هرفریم

Boxing

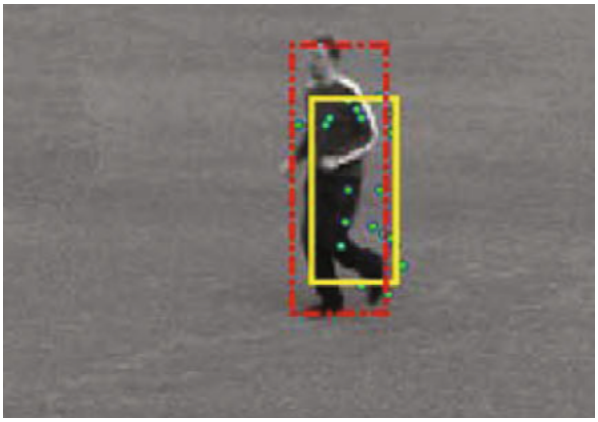


Handclapping

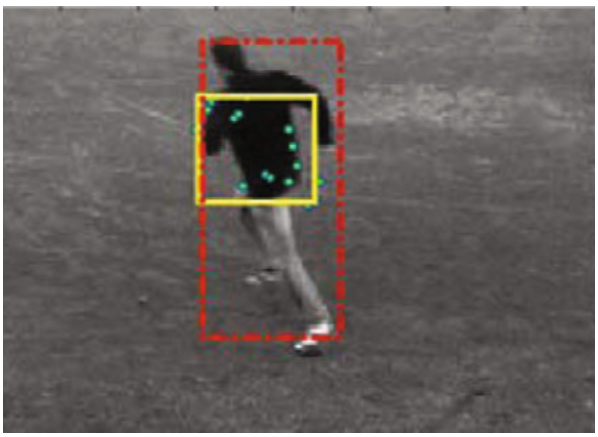
Handwaving



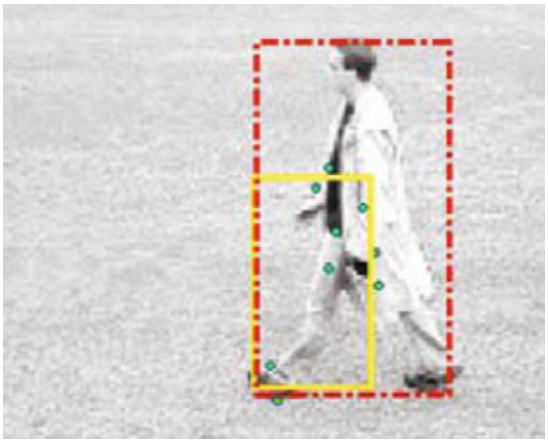
Jogging



Running



Walking



شکل5- شناسایی ناحیه در هر فریم

```
% Project Title: MoSIFT Improvements
% Author: Hesamoddin Hosseini
% Contact: hesamoddin.hosseini@mail.um.ac.ir
% Websit:http://hesamoddin.hosseini.student.um.ac.ir

close all
clear all
clc
delete('Frames\*.jpg');
[filename pathname] = uigetfile({'*.avi'},'Select A Video File');
I = VideoReader([pathname,filename]);

nFrames = I.numberofFrames;
vidHeight =  I.Height;
vidWidth =  I.Width;
mov(1:nFrames) = ...
    struct('cdata', zeros(vidHeight, vidWidth, 3, 'uint8'),...
           'colormap', []);
       WantedFrames = 50;
for k = 1:WantedFrames
   mov(k).cdata = read( I, k);
   mov(k).cdata = imresize(mov(k).cdata,[256,256]);
   imwrite(mov(k).cdata,['Frames\',num2str(k),'.jpg',num2str(k),'.jpg']);%%%ADDRESS
EDITED
end
for I = 1:WantedFrames
   im=imread(['Frames\',num2str(I),'.jpg',num2str(I),'.jpg']); %%%ADDRESS EDITED
    figure(1),subplot(5,10,I),imshow(im);
end
clc
for i=1:WantedFrames
    disp(['Processing frame no.',num2str(i)]);
  img=imread(['Frames\',num2str(i),'.jpg']);%%%ADDRESS EDITED
  f1=il_rgb2gray(double(img));
  [ysize,xsize]=size(f1);
  nptsmax=40;
  kparam=0.04;
  pointtype=1;
  sxl2=4;
  sxi2=2*sxl2;
  % detect points
  [posinit,valinit]=STIP(f1,kparam,sxl2,sxi2,pointtype,nptsmax);
  Test_Feat(i,1:40)=valinit;

end


load('TrainFeat.mat')
X = meas;
Y = New_Label;
Z = Test_Feat;

md1 = ClassificationKNN.fit(X,Y);
Type = predict(md1,Z);
Type = mode(Type); %%% NEW
if (Type == 1)
    disp('Boxing');
    helpdlg(' Boxing ');
elseif (Type == 2)
    disp('Hand Clapping');
```

```
    helpdlg('Hand Clapping');
elseif (Type == 3)
    disp('Hand Waving');
    helpdlg('Hand Waving');
elseif (Type == 4)
    disp('Jogging');
    helpdlg('Jogging');
elseif (Type == 5)
    disp('Running');
    helpdlg('Running');
elseif (Type == 6)
    disp('Walking');
    helpdlg('Walking');
elseif (Type == 7)
    disp('Cycling');
    helpdlg('Cycling');
elseif (Type == 8)
    disp('Surfing');
    helpdlg('Surfing');
end
```