计算机视觉系统期中汇报

第12组

E202008000106邹雪丹

201708010207 杨佳政

201708010214 唐协成

201508010628 李鑫

目录

CONTENTS

01. 问题定义

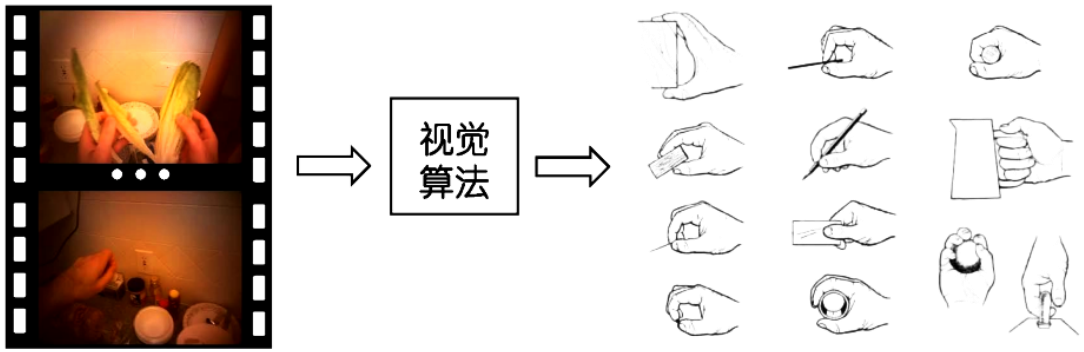
02. 拟采用的方法

03. 拟采用的工具

04. 时间规划

1.问题定义

编程任务-需求



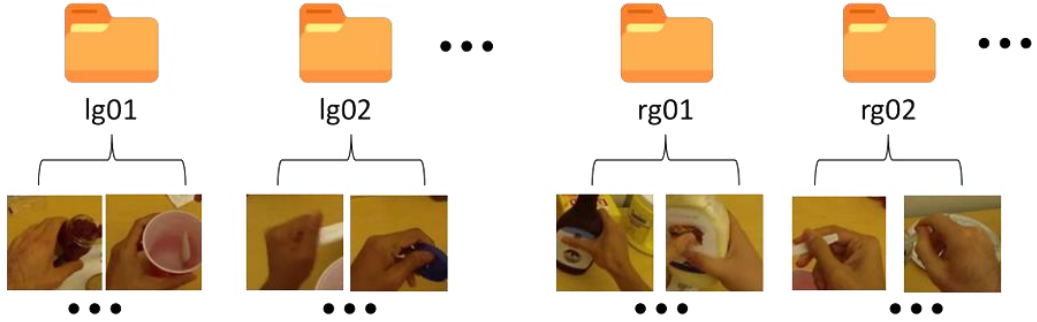
手部动作图像集　　抓取手势类型集

需求描述：
从大量的视频数据中自动发掘常见的抓取手势类型集，在机器人学和医疗康复等领域有重要的应用价值。此次编程任务希望通过设计计算机视觉算法，对包括手部动作的图像集合进行处理和分析，从中自动发现不同的抓取手势类型。

编程任务-具体要求

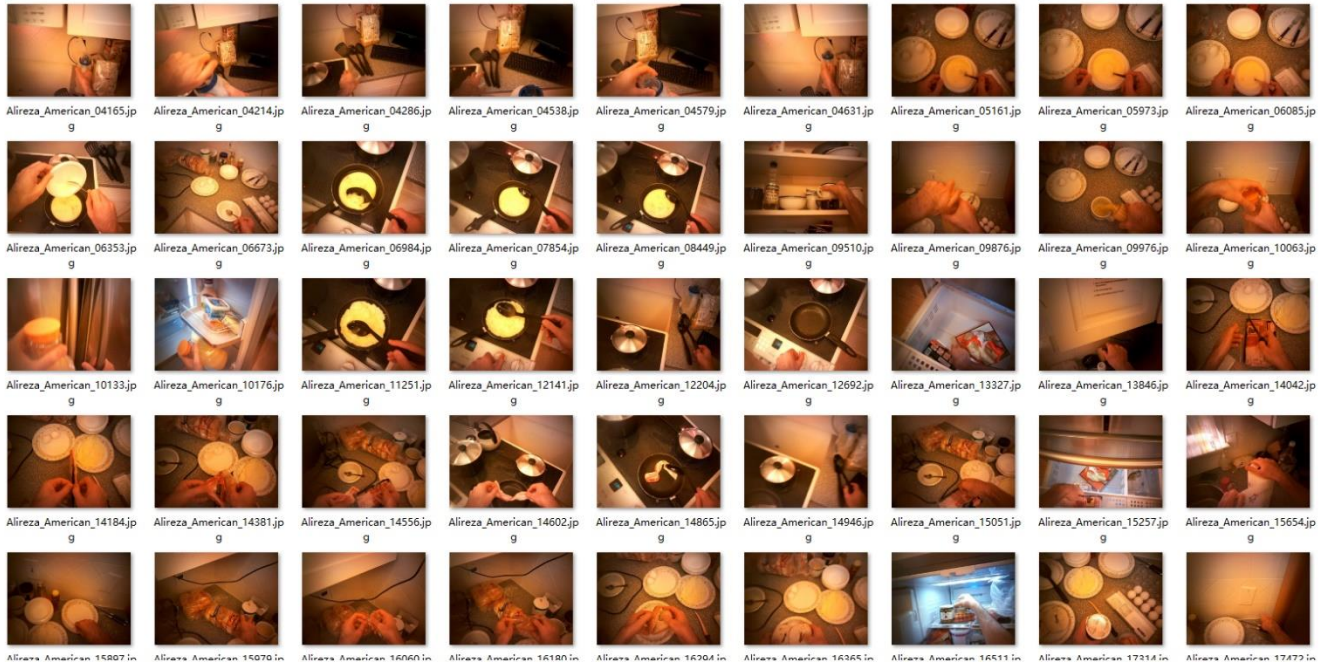
输入：包含手部动作的图像集

输出：对应不同抓取手势的手部图像集合(建议10-15个类型)



注意：输出图像保留原始图像的文件名

1.问题定义



Download The Dataset

The table below contains the following data for each *activity* performed by each *subject*: Egocentric Video (V), Audio (A), Gaze (G), Action Annotations (N)

Subject/Activity	American Breakfast	Pizza (Special)	Afternoon Snack	Greek Salad	Pasta Salad	Turkey Sandwich	Cheese Burger	Download all subject data
Yin	V, A, G, N	V, A, G, N	V, A, G, N	V, A, G, N		V, A, G, N	V, A, G, N	
Alireza	V, A, G, N	V, A, G, N	V, A, G, N	V, A, G, N	V, A, G, N	V, A, G, N	V, A, G, N	
Carlos	V, A, G, N	V, A, G, N	V, A, G, N	V, A, G, N	V, A, G, N	V, A, G, N	V, A, G, N	
Rahul	V, A, G, N	V, A, G, N	V, A, G, N	V, A, G, N	V, A, G, N	V, A, G, N	V, A, G, N	
Shaghayegh	V, A, G, N	V, A, G, N	V, A, G, N					

观察给出的图片数据集，我们发现所有的图片均是第一人称视角拍摄的（FPV）。
经调查发现，这其实是EGTEA Gaze+数据集，原数据集其实是一个第一人称视角视频录像，这里进行了截图。

http://ai.stanford.edu/~alireza/GTEA_Gaze_Website/GTEA_Gaze+.html

http://cbs.ic.gatech.edu/fpv/#egtea_gaze_plus

1.问题定义




手只出现在了边缘角落，较差的光照条件导致和背景色差异不明显

动态模糊

两只手

清晰的图片

1.问题定义

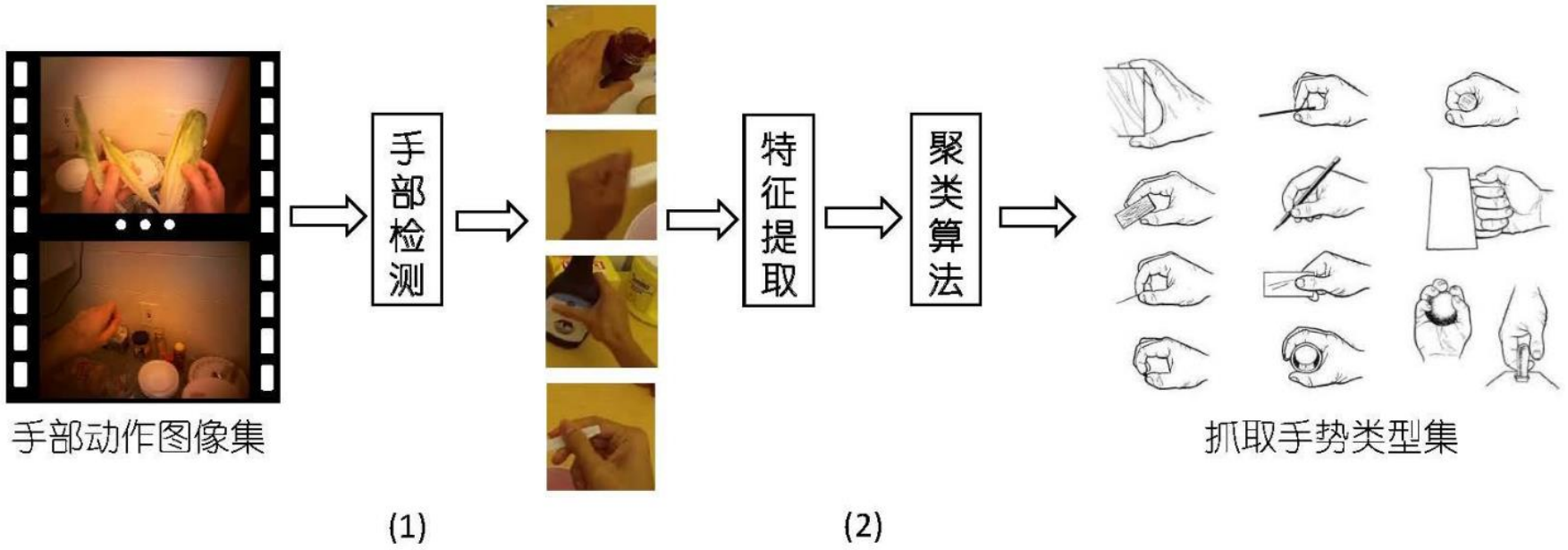
我们随机挑选数据图片用百度AI平台的手部测试和关键点标记算法做测试，事实证明即使是想拿自家算法卖钱的百度AI都没有能力去识别出部分数据图片中的手部。

我们认为这个问题是一个**研究级**难度的问题。

我们要识别**第一人称摄像头**拍摄的各种**手部抓取姿势**并将他们分类。

https://ai.baidu.com/tech/body/hand

2.拟采用的方法



手部识别

↓

图像分割

↓

特征提取

↓

图像聚类

我们认为这个问题可以分解为识别手部并分割对应图片，对分割到的手部进行特征提取，最终聚类手部这几个子问题。对于每一个子问题的解决应该是相对独立的，我们可以单独地提升其中的每个步骤的效果。

2.拟采用的方法

A. 手部检测和分割

1、Grabcut

2、基于肤色的检测方法

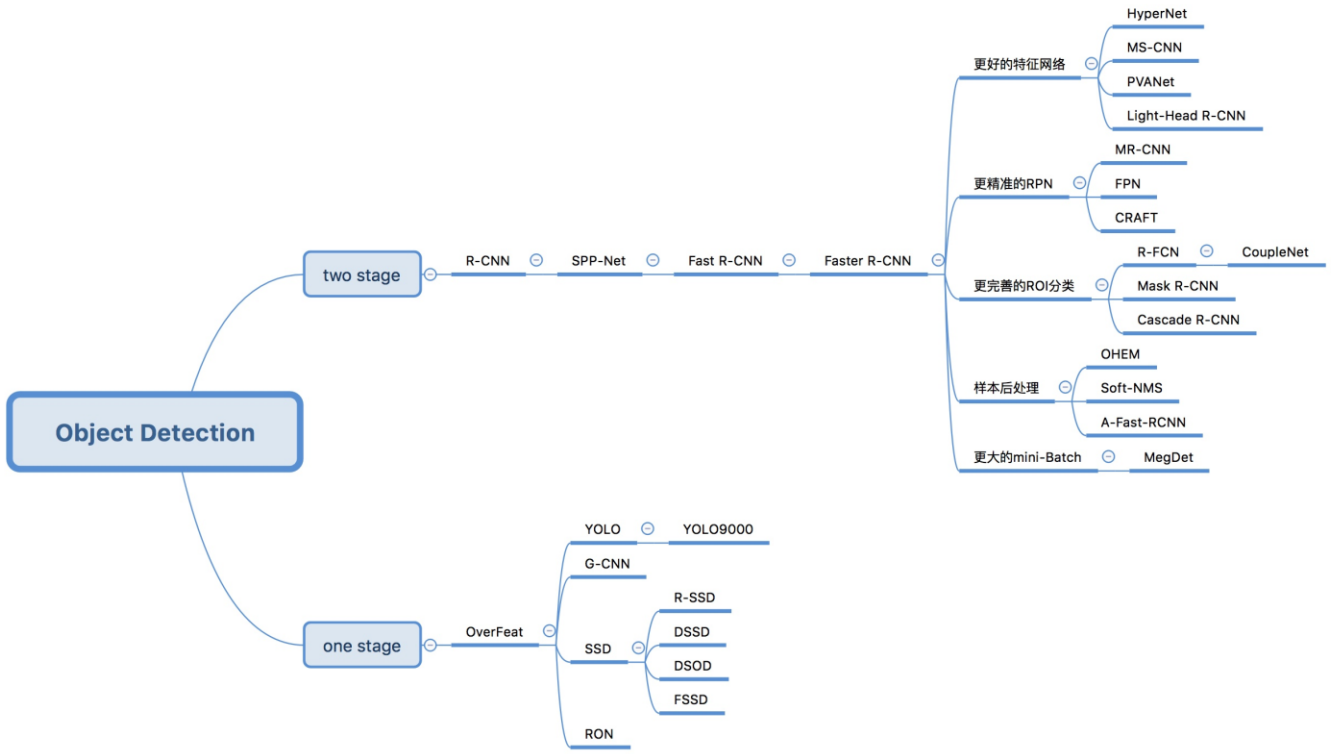
3、Yolo（one-stage）

4、R/Fast/Faster-CNN (two-stage)

5、深度图方法

6、手动设计特征法（Hand-crafted Feature）

Assist: 分水岭，腐蚀膨胀.....



深度学习中常见的目标检测方法

2.拟采用的方法

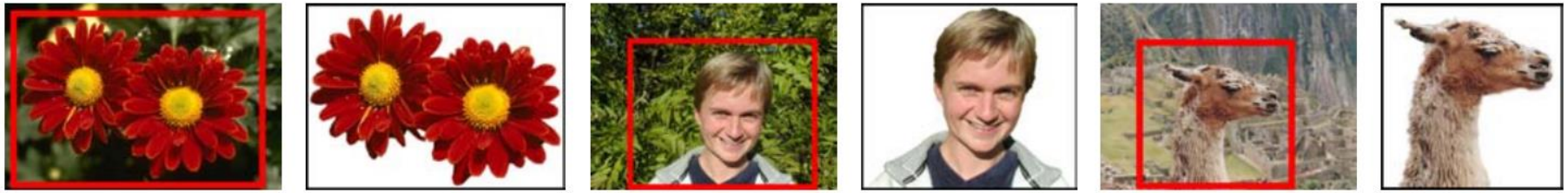
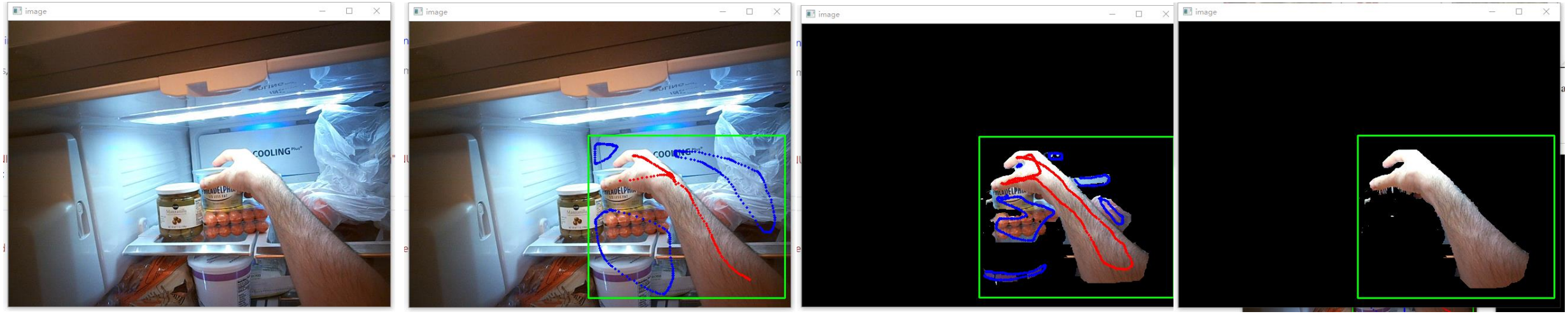


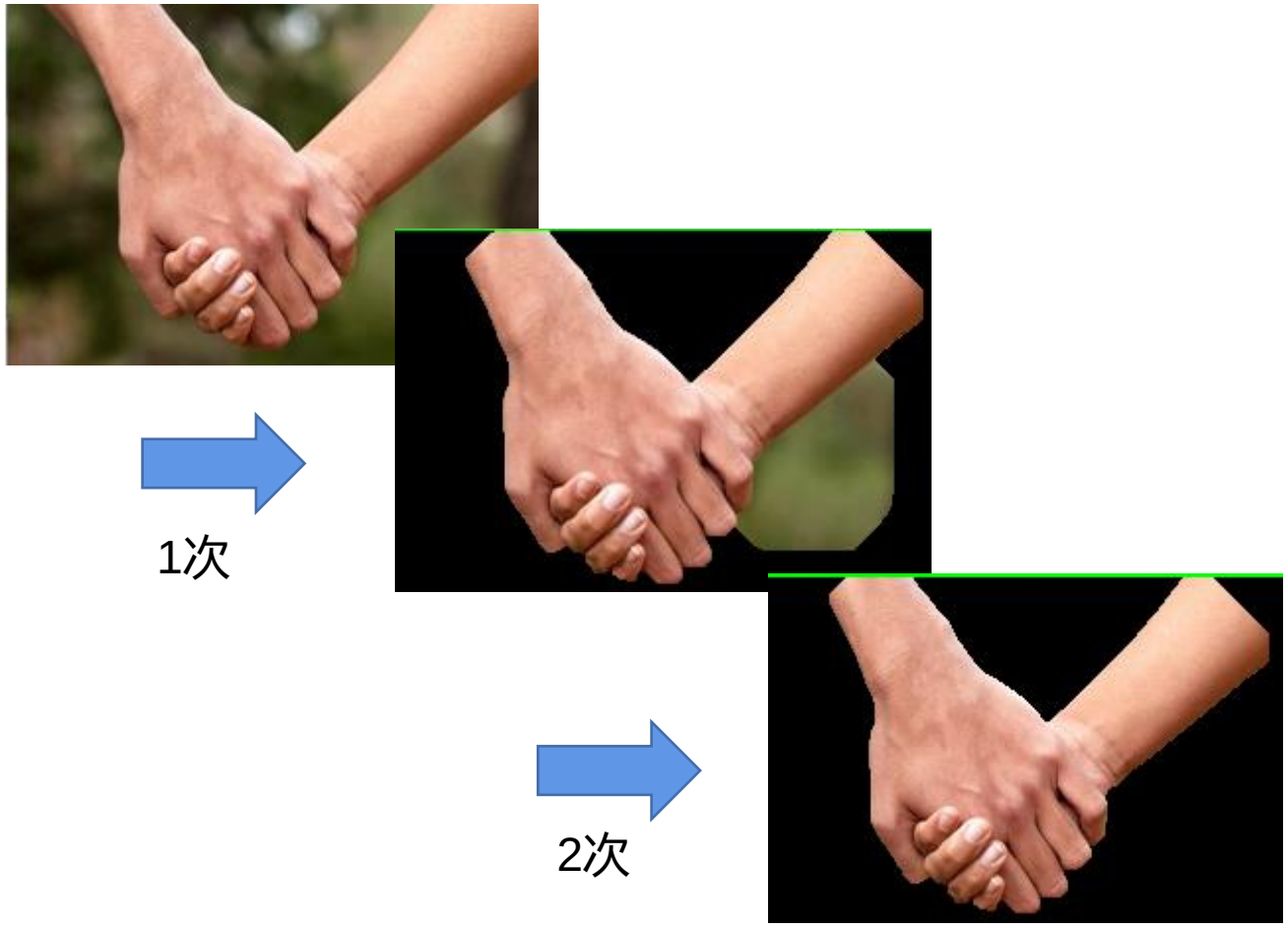
Figure 1: **Three examples of GrabCut.** The user drags a rectangle loosely around an object. The object is then extracted automatically.



一般意义的图像分割算法，比如经典的**GrabCut**方法，要求用户交互式地预先给出前景区和背景区，我们随意挑选了一张数据集中的图片来测试，如果选取太大，因为背景杂乱，迭代两次交互后成功分割出来了手部图像。

Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. 2004. "GrabCut": interactive foreground extraction using iterated graph cuts. In ACM SIGGRAPH 2004 Papers (SIGGRAPH '04). Association for Computing Machinery, New York, NY, USA, 309–314. DOI:https://doi.org/10.1145/1186562.1015720

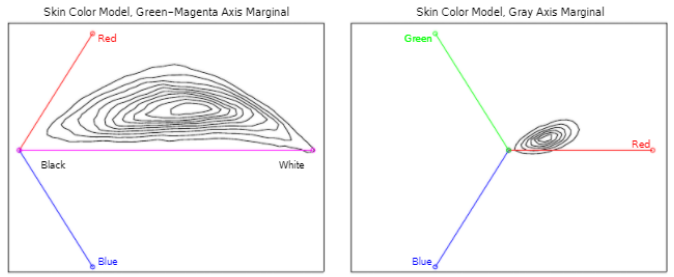
2.拟采用的方法



对于背景简单的图片来说，Grabcut算法迭代足够多次还是可以自动地分割出主体物（前景）的。
不过我们的数据图片大部分手压根不是主体，而且背景杂乱，也不可能全部都人为交互处理。

2.拟采用的方法

一种特定地分割手部的方法是最早由Jones等人提出肤色检测方法。一种常见的操作是:
1）把原图转到YUV色彩空间
2）对其中的U分量进行高斯滤波
3）做自二值化阈值分割OSTU法（大津算法）



(a) Contour plot for skin model, marginalized along the green-magenta axis.

(b) Contour plot for skin model, marginalized along the gray axis.

(c) Contour plot for non-skin model, marginalized along the green-magenta axis.

(d) Contour plot for non-skin model, marginalized along the gray axis.

Figure 2: Contour plots for marginalizations of the skin and non-skin color models. The top row shows the skin model, the bottom row shows the non-skin model. The left column uses the viewing direction from Figure 1 (c) while the right column uses the view from Figure 1 (d).

Jones, Michael J., and James M. Rehg. "Statistical color models with application to skin detection." International Journal of Computer Vision 46.1 (2002): 81-96.

2.拟采用的方法

之后我们可以使用opencv自带的基于Satoshi等人提出的方法实现的findContours()函数来找到所有的轮廓，并选取最大轮廓保留为我们的手部。遗憾的是数据图片中大部分时候分割出来的手往往并不是最大轮廓或者手部无法被有效轮廓分割出来。

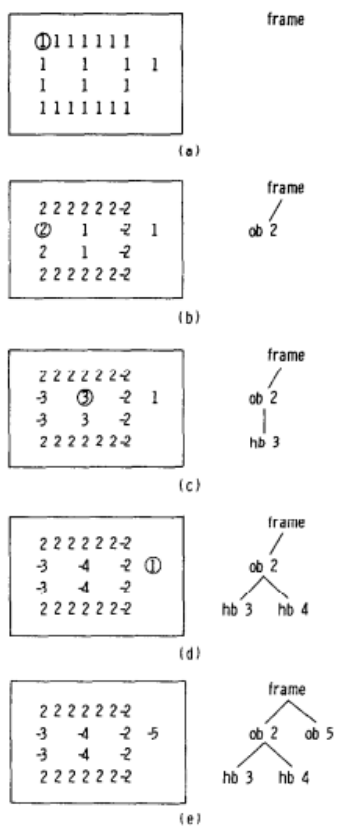
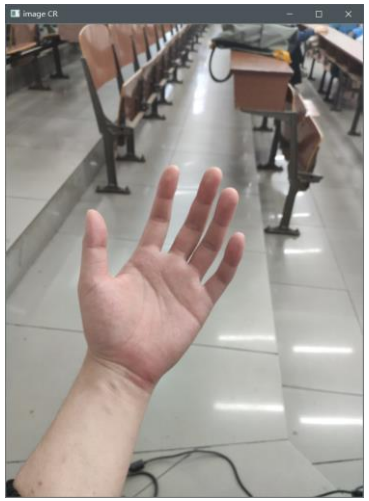
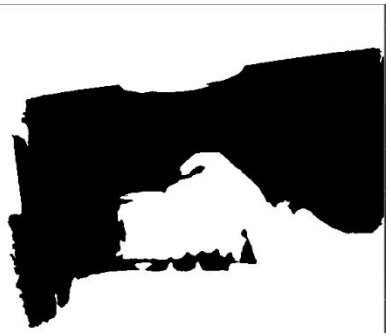


FIG. 3. An illustration of the process of Algorithm 1. The left-hand figures show the pixel values and the right-hand the extracted structures among borders (ob: outer border, hb: hole border). The circled pixels are the starting points of border following.





Satoshi Suzuki and others. Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics, and Image Processing, 30(1):32–46, 1985.

2.拟采用的方法

(a) (b) (c) (d) (e)

Figure 1: **Overview of the method.** (a) Original image. (b) Some of the hypotheses proposed by hand and context detector. Bounding boxes in red are proposed by the hand detector and in green by the context detector. (c) Skin detection and hypotheses generation. (d) Super-pixel segmentation of the image with combined hypothesised bounding boxes from the three proposal schemes. Using super-pixel based non-maximum suppression (NMS), overlapping bounding boxes are suppressed. (e) Final detection after post-processing.

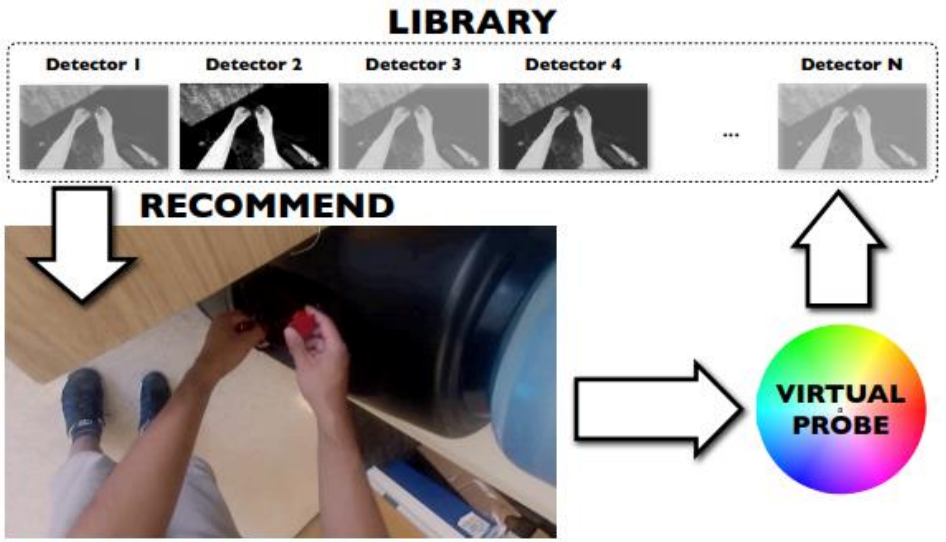


Figure 1. Ego-centric hand detection as a model recommendation task. Virtual probe features are extracted at test time to recommend the best detector performance.

其它论文文献有将这个分割任务做成三个分类器（手形，语意和皮肤）然后评估整体得分，或者当做带推荐系统的分类器问题，还有的文献利用先验知识加上光流处理(DTR)，或者构建决策树森林,LSVM等等。

Mittal, Arpit , A. Zisserman , and P. Torr . "Hand detection using multiple proposals." British Machine Vision Conference 2011.
C. Li and K. Kitani, “Model Recommendation with Virtual Probes for Egocentric Hand Detection,” in 2013 IEEE International Conference on Computer Vision, Sydney, 2013, pp. 2624–2631, IEEE Computer Society.

2.拟采用的方法

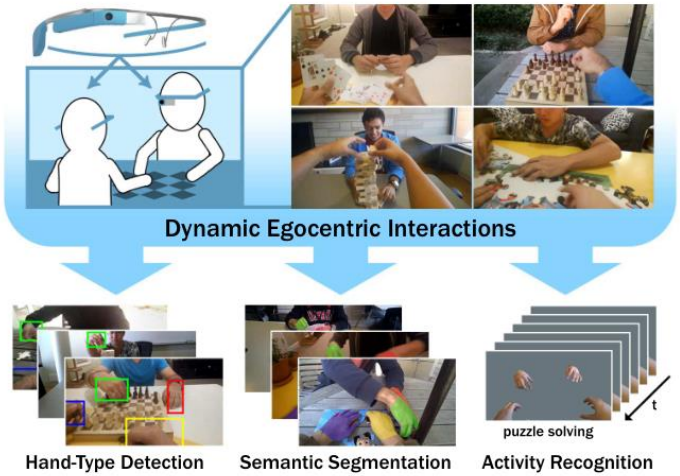


Figure 1: We present a CNN-based technique for detecting, identifying, and segmenting hands in egocentric videos of multiple people interacting with each other. To illustrate one specific application, we show that hand segments alone can be used for accurate activity recognition.

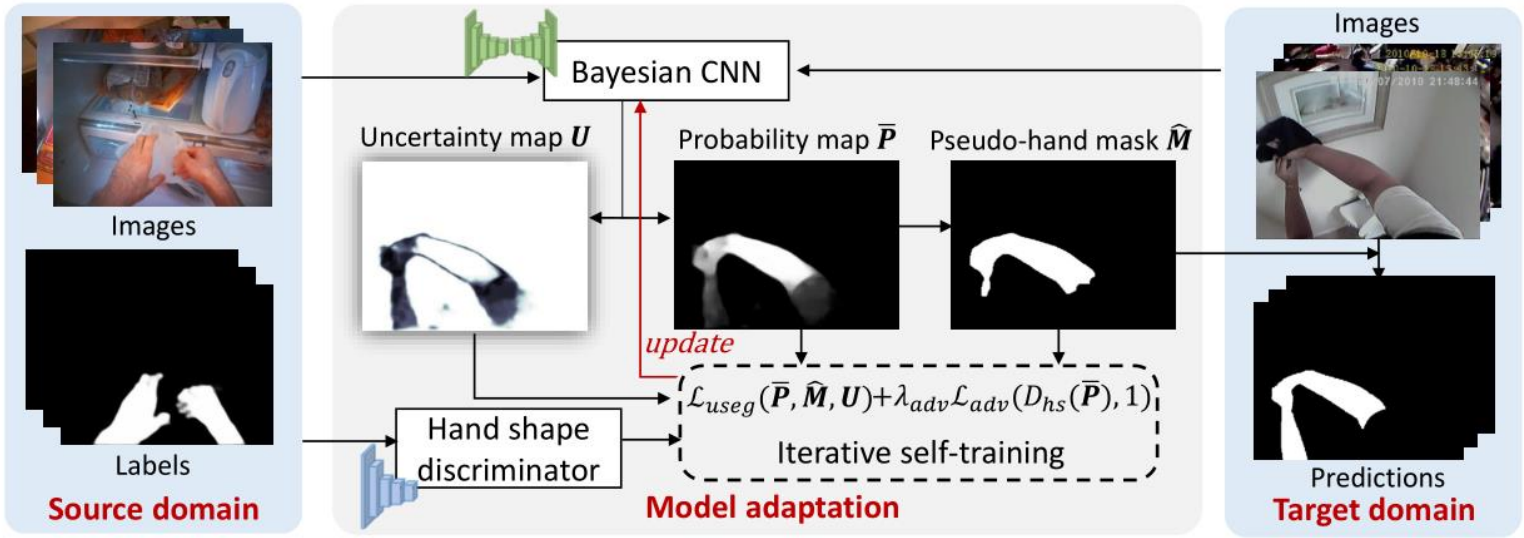


Figure 3. Overview of the proposed uncertainty-guided model adaptation.

最近基于深度学习中卷积神经网络CNN提出的一些网络模型被运用在处理FPV的手部识别问题上，一些最新的论文表明他们训练得到的网络能够很好地识别第一人称拍摄的视频图像中的手部位置,并且不再是传统CNN方法的two-stage，并且检测的结果通常自带左右手判别。

Bambach, Sven, et al. "Lending a hand: Detecting hands and recognizing activities in complex egocentric interactions." ***Proceedings of the IEEE International Conference on Computer Vision*****. 2015.**

Cai, Minjie, Feng Lu, and Yoichi Sato. "Generalizing Hand Segmentation in Egocentric Videos With Uncertainty-Guided Model Adaptation." ***Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*****. 2020.**

2.拟采用的方法

B. 特征分析和手势聚类

1、手动设计特征法（Hand-crafted Feature）
eg: Gabor, HOG,SIFT, DTR, 模板匹配，统计分析…

2、深度学习法（各种CNN）

常用的手势特征:轮廓、边缘、图像矩、图像特征向量、区域直方图特征…

如何判定手势（抓取姿态）类型?
手形状，物体属性，手和物体的交互…

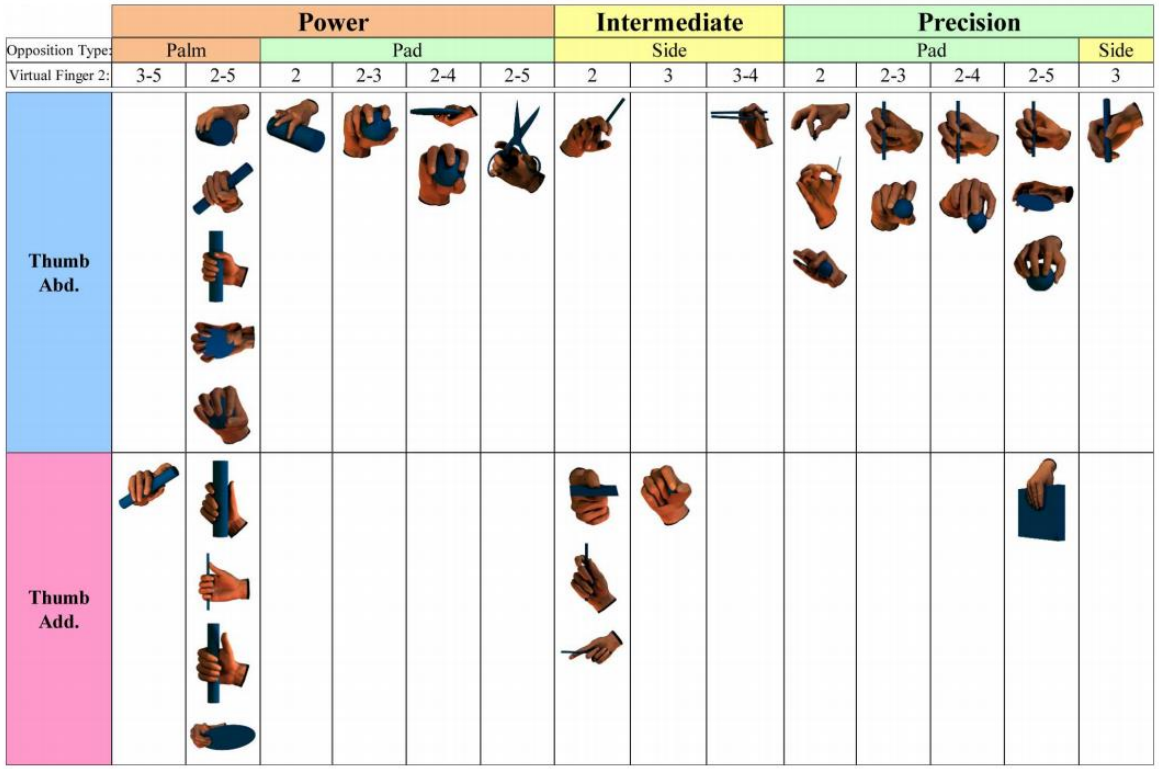


Fig. 2. Comprehensive Grasp Taxonomy which includes 33 grasp types.

抓取姿势分类例

T. Feix, R. Pawlik, H.-B. Schmiedmayer, J. Romero, and D. Kragic, “A comprehensive grasp taxonomy,” in Robotics, Science and Systems: Workshop on Understanding the Human Hand for Advancing Robotic Manipulation, 2009, pp. 2–3.

2.拟采用的方法

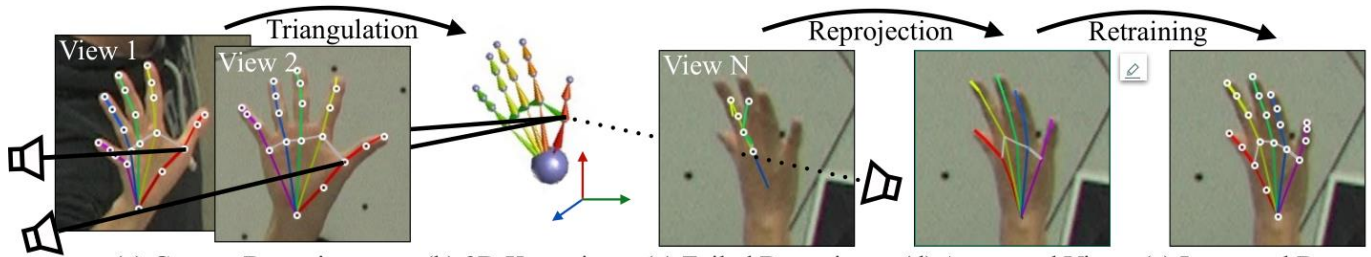
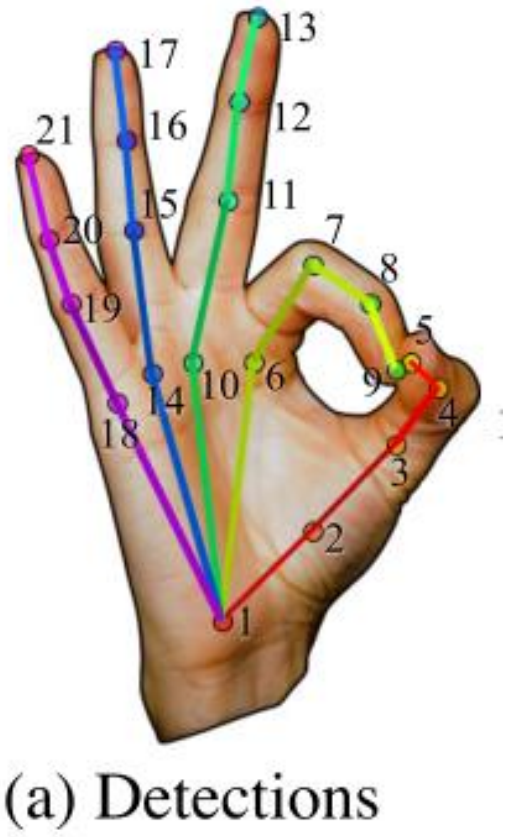


Figure 3: Multiview Bootstrapping. (a) A multiview system provides views of the hand where keypoint detection is easy, which are used to triangulate (b) the 3D position of the keypoints. Difficult views with (c) failed detections can be (d) annotated using the reprojected 3D keypoints, and used to retrain (e) an improved detector that now works on difficult views.



近年有论文通过标记少量多视角下的手部图像为数据集，训练CPM网络，使得可以对单张二维手部图像标记出对应的关键点。

T. Simon, H. Joo, I. Matthews and Y. Sheikh, "Hand Keypoint Detection in Single Images Using Multiview Bootstrapping," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),

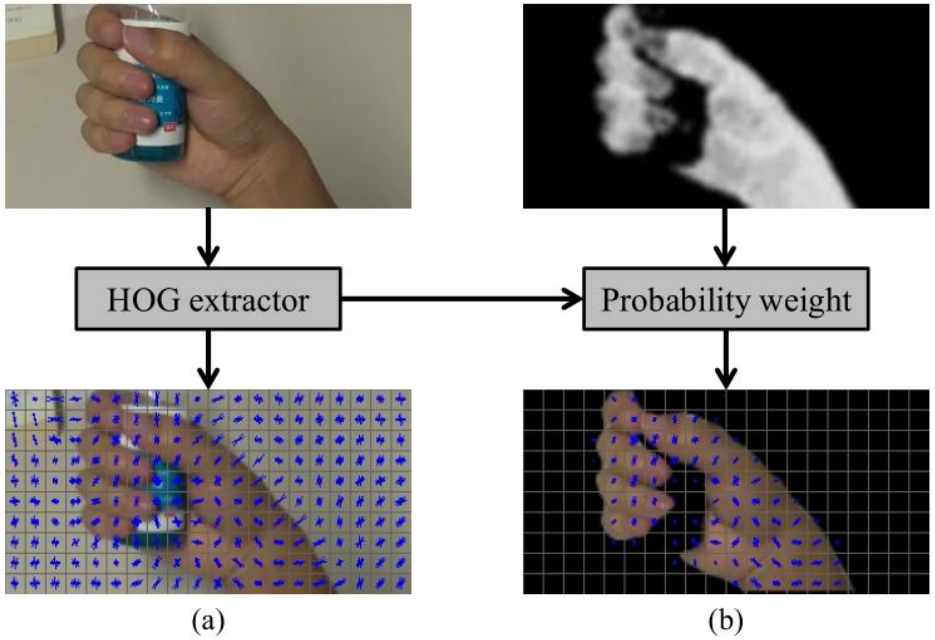
2.拟采用的方法

常见的聚类算法

1. 基于层次的方法 (Hierarchical Methods)
2. 基于划分的方法 (Partition-based Methods)
   eg: K-means
3. 基于密度的方法 (Density-based Methods)
   eg: DBSCAN
4. 基于网络的方法 (Grid-based Methods)
5. 基于模型的方法 (Model-based Methods)
   eg: SOM
6. 基于模糊的方法 (FCM)

.....

2.拟采用的方法



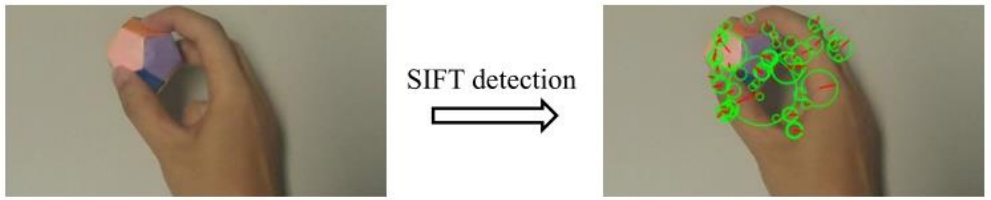
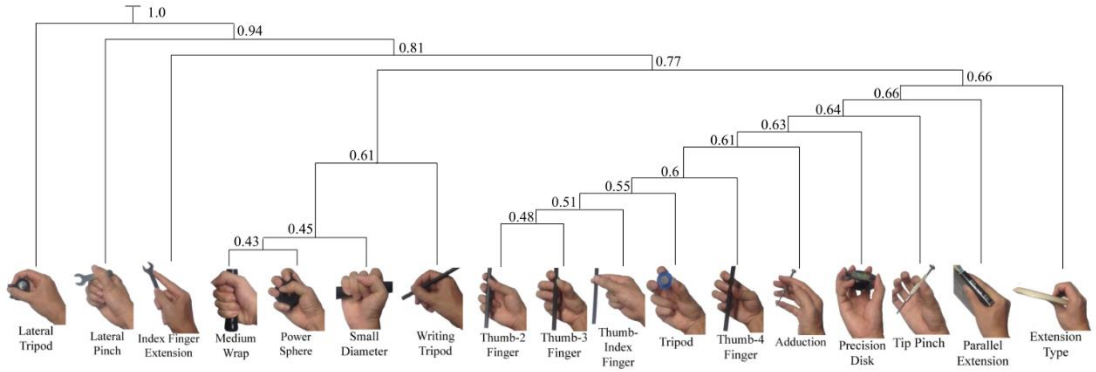


Fig. 4. Visualization of SIFT keypoints.



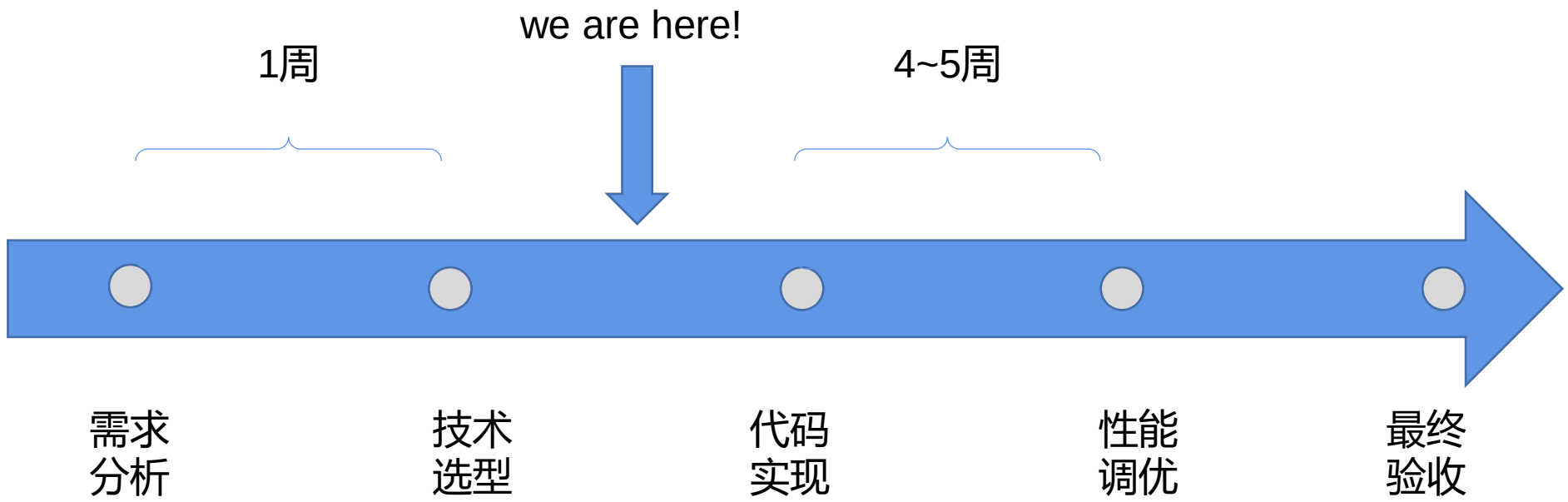
近年有研究者通过计算手部区域的HOG，手部交互物体的SIFT特征，然后训练了一个一对多多类别分类器来分析抓取手势的类别。

Cai, Minjie , K. M. Kitani , and Y. Sato . "A scalable approach for understanding the visual structures of hand grasps." ***Proceedings IEEE International Conference on Robotics & Automation*** **2015(2015):1360-1366.**

3.拟采用的工具

- 语言：Python3.7 (C++, Matlab)
- 软件库：OpenCV3，numpy ，（Pillow,tqdm,...）
- （可能用到各种深度学习相关模块：Tensonflow，PyTorch, CUDA... ）
- 系统: Ubuntu 20.x (Win10)
- GPU: GTX1660Ti 6G
- 其它平台：Google Colab, 极链AI，矩阵云..... 学校实验室电脑（？）

4.时间规划



做中学，搭积木， 多尝试，取最优

Question?

第12组

E202008000106邹雪丹
201708010207 杨佳政
201708010214 唐协成
201508010628 李鑫