

视频特征提取

- [视频分类](#)
- [视频特征提取](#)
 - [DT](#)
 - [IDT](#)

视频分类

1. 视频分类：旨在让计算机自动识别视频内容的语义类别。
2. 基于手工特征的视频分类：在深度学习兴起之前，传统视频分类方法利用手工特征进行视频表示，然后利用支持向量机（SVM）等传统分类器完成视频分类。
 - 流程：样本 -> 感知 -> 预处理 -> **特征提取** -> 机器学习算法 -> 预测结果。

视频特征提取

1. 视频特征：包括二维平面和一维时序的三维立体特征。
 - 分解为静态信息（视频帧）、动态信息（如**光流**）。
2. 光流：可以确定（也许是所有的）图像点上运动方向和运动速率，**反映了在时间间隔 d_r 内由于运动所造成的图像变化。**
 - 一般而言，光流是由于场景中**前景目标本身的移动、相机的运动或者两者的共同运动**所产生的。
 - 如第 t 帧时 A 点的位置是 (x_1, y_1) ，第 $t + 1$ 帧时 A 点的位置是 (x_2, y_2) ，则 A 点运动了 $(u_x, u_y) = (x_2, y_2) - (x_1, y_1)$ 。
 - 缺点：
 - 传统稠密光流方法计算开支比较大；
 - 理论的基础建立在同一物体亮度恒定以及物体位移较小的假设上；
 - **只能表示短时序的运动特征。**
3. 长时序特征：**轨迹特征**。
 - 常见提取方法：（1）**DT(Dense Trajectory)**；（2）**IDT(Improved Dense Trajectory)**。

DT

4. DT:

- 方法步骤：

1. 密集采样：

- 在每帧图像的多个空间尺度采样特征点（每隔5个像素，图像尺度以 $\frac{1}{\sqrt{2}}$ 增加）；
- 除去无用的特征点：某些背景单一的区域在时序中可能没有变化，这类特征点不需要去跟踪（根据特征值进行过滤）。

2. 特征点轨迹跟踪：

- 根据关键点的光流 w ，估计下一帧的特征点所在位置：

$$P_{t+1} = (x_{t+1}, y_{t+1}) = (x_t, y_t) + (M * w_t)|_{(x_t, y_t)}, \quad M \text{ is median filter.}$$

- 重复上述方法，形成轨迹；
- 每隔 L （可取 $L = 15$ ）帧需要重新进行密集特征点采样，降低漂移现象。

3. 轨迹形状特征描述子：HOG、HOF、MBH

- 对于长度为 L 的轨迹，其形状特征可表示为：

$$T = \frac{(\Delta P_t, \dots, \Delta P_{t+L-1})}{\sum_{j=t}^{t+L-1} \|\Delta P_j\|}, \quad \Delta P_t = (P_{t+1} - P_t) = (x_{t+1} - x_t, y_{t+1} - y_t)$$

- HOG：计算灰度图像的梯度直方图，通过计算和统计视频帧局部区域梯度方向的直方图，描述视频的静态信息；
- HOF：计算光流的直方图，通过计算和统计光流方向的直方图，描述视频的运动信息。
- MBH：可理解为在光流图像上计算的HOG特征。
- 结论：HOG注重静态图像的细节、HOF注重局部运动的信息、MBH可降低相机运动带来的影响。

4. 利用BoF（Bag of Features）方法对特征进行编码；

5. 基于编码结果训练SVM分类器。

IDT

5. 改进方面：

- 在无关运动估计方面：IDT特征通过估计相机运动以消除背景光流。
- 在特征编码方面：IDT特征采用费雪向量（Fisher Vector, FV）模型代替DT特征中的BoF模型。