# 浏览式阅读

## 1 自己的总结、评价以及应用

Our model involves a generative module that separately encodes each person into an appearance code and a structure code, and a discriminative module that shares the appearance encoder with the generative module.

## 2 文章的主要问题（abstract、疑问句中）

存在的问题：

第一层： still person images（静态照片）across disjoint camera views

简而言之：Information of a still image is sometimes not enough for recognizing a person

第二层：当采用viedo-based的方法时，又出现了新的问题：some inter-class difference

can be much more obscure than the one when using stillimage-based representation

提出的解决方案：

we propose a top-push distance learning model

## 3 结论（abstract以及conclusion中）

## 4 思路脉络（小标题中的关键句）

1. **Introduction**

Video与still image的不同之处：

Firstly, video is an image sequence containing space-time information（时空信息）,in which motion information is available.

Secondly, appearance cues are more abundant in a sequence than in a still image, which can facilitate extracting more robust appearance features.

Thirdly, occlusion and background influence can be eliminated to some extent.In a sequence, background variation and occlusion can be regarded as removable noises, while in still images they are troubling interferences.

Video-based re-id存在的challenges：

Firstly,video-based person representations are also similar because of similar appearance.

Seondly,it is unfortunate that the walking actions or other motions of different persons may be similar as well. Which means the inter-class variation may be smaller.

the ambiguity of videos

To mine these minor differences in the data, more stringent constraint should be exploited to look for a latent space to maximize the inter-class margin between different persons.

最后：To address the above problem in video-based person reid, we propose a top-push distance learning model (TDL) in this work.

1. Related Work

以前的工作：

most of existing works can be categorized into extracting discriminant/relible features

[6, 32, 7, 4, 25, 14, 36, 35] or learning robust metrics or subspaces for matching

最近的一些工作：

①Dynamic Time Warping (DTW)：, which is a popular sequence matching method widely used for action recognition

②Wang et al：introduced a pictorial video segmentation approach and deployed a fragment selecting and ranking model for person matching.

③Srikrishna et al ... ...

以上各种方法存在的一个不足：

All these works use either multiple images or a selected fragment of a sequence to extract feature, and thus they ignore the integrity and the richness （完整性和丰富性）of video features.

下面提出我们的方法：

我们的不同之处：look for a latent subspace，so that more robust latent features can be exploited.

inter-class variation is much smaller on video level than that on still image level

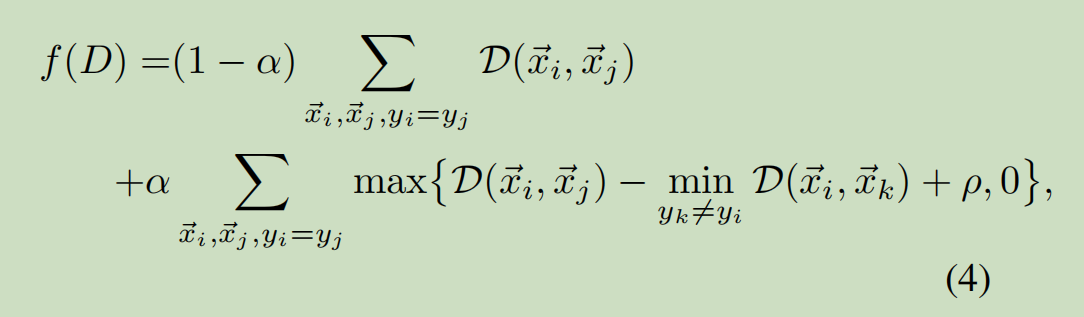
1. Approach

The feature representation of a person video in our model has two main components: space-time features and appearance features.

extracting the space-time features：HOG3D descriptor

extracting the appearance features：color histograms and LBP features

目标函数：



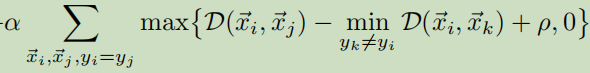
We call the second term the top-push constraint.

**4. Experiment**

1. **Conclusion**

there are more ambiguities in video-based features than still-image-based features.so we introduce a top-push constraint to quantify ambiguous video representation.

下图就是op-push constraint



## 5 难理解点

专业术语：

different illumination：照明

Occlusion：遮挡

context information：上下文信息

probe-gallery pair：探测库对

contextual instance expansion module：上下文实例拓展模块

multi-camera surveillance systems：多摄像机监控系统 surveillance：监视

intra-class variations：组内变异 background clutter：背景混乱

manually cropped image snapshots or video clips：手动裁剪的图像快照或视频剪辑

appearance cue：外观提示 manual annotations：手动注释 semantic group：语义群

spatial and temporal cue：空间和时间线索 texture：纹理

latent subspace：潜在子空间