Person re-identification by Local Maximal Occurrence representation and metric learning

Liao Shengcai, Hu Yang, Zhu Xiangyu, Li Stan Z.

Experiment Presenter: Zhenpei Yang
Person Re-identification: Given an image of a person from one camera, identifying the person from images taken from different cameras

(a) Pedestrian Detection

(b) Person Re-identification

Slides credit: liangzheng
Person re-identification is a challenging problem because:

- Big Intra-class variance due to pose, viewpoint, illumination change.
- Need a proper metric to compute cross-class distance.

Contribution

- Extract good features \textit{Local Occurrence Maximum (LOMO)}
- Use good distance metric \textit{Cross-view Quadratic Discriminant Analysis (XQDA)}
About distance metric

Examine Images

Which image is more likely correspond to image Q? A or B?

Model the distribution for *intra-class distance and extra-class distance*!
Discriminative model

Intuition: model the covariance for Intra-class distance and extra-class distance respectively using gaussian

\[ f_k(x) = \frac{1}{(2\pi)^{p/2} |\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)\right) \]

- Linear Discriminant Analysis (LDA)
- Quadratic Discriminant Analysis (QDA)
- Cross-view Quadratic Discriminant Analysis (XQDA)

\[ \sum_k = \sum_k, \forall k \]

\[ d = \log \frac{\pi_k}{\pi_l} - \frac{1}{2}(\mu_k + \mu_l)^T \Sigma^{-1} (\mu_k - \mu_l) + x^T \Sigma^{-1} x (\mu_k - \mu_l), x = \Delta \]

\[ d = -\frac{1}{2} x^T \Sigma^{-1} x - \frac{1}{2} \log \left| \frac{\Sigma_k}{\Sigma_l} \right| + \log \frac{\pi_k}{\pi_l} \]
Intuition: Original feature space is too high dimension. Maybe it’s helpful to consider the problem in subspace.

Cross-view Quadratic Discriminant Analysis (XQDA)

\[ P(\Delta|\Omega_I) = \frac{1}{(2\pi)^{d/2} |\Sigma_I|^{1/2}} e^{-\frac{1}{2} \Delta^T \Sigma_I^{-1} \Delta}, \]

\[ P(\Delta|\Omega_E) = \frac{1}{(2\pi)^{d/2} |\Sigma_E|^{1/2}} e^{-\frac{1}{2} \Delta^T \Sigma_E^{-1} \Delta}, \]

Hard to measure precisely in high dimension space. Measure this in subspace!

Which subspace? What about PCA?
Cross-view Quadratic Discriminant Analysis (XQDA)

The two distributions for intra-class and extra-class distance both have zero means.

\[
\max_w \frac{\sigma_E(w)}{\sigma_w(w)}
\]
The QXDA chose subspace that maximize the two classes’ variance ratio.
Viewpoint Invariance Analysis

- Video taken by hand-hold camera
- #Total 23 seconds/705 frames(48*128)
- 0-360 degree view

Slides credit: my roommate
Viewpoint Invariance Analysis

Choose feature

Extract features on each frame

Learn a distance metric $d(\cdot)$ using XQDA

Measure the $d(f_t, f_1)$

Distance?
Investigated Features

- Local Maximum Occurrence (LOMO)
- LOMO without Maximum Operator
- Convolutional Neural Network Feature (CNN)
Distance Metric

- Quadratic Discriminant Analysis (XQDA)
- Cosine Similarity
- The max operation in LOMO makes it more robust to viewpoint change
- XQDA can learn more robust metric against viewpoint variation
Which region contribute mostly?

- Conduct training on four different body parts
- Compute the matching performance using each body parts
The upper body is the most distinguishable part
Sensitivity to Occlusion

Parameter: the size of occlusion area
The performance degrades monotonously as occlusion becomes more severe.
1st rank accuracy degrades monotonously as occlusion become more severe.
Conclusion

- XQDA find the subspace that maximize the covariance odds of intra-class and extra-class distance.
- Doesn’t robust to occlusion.
- LOMO feature has some viewpoint invariance due to the max operation.
- XQDA can learn more robust metric against viewpoint variation.
- Upper body is the most distinct part for person-reidentification.