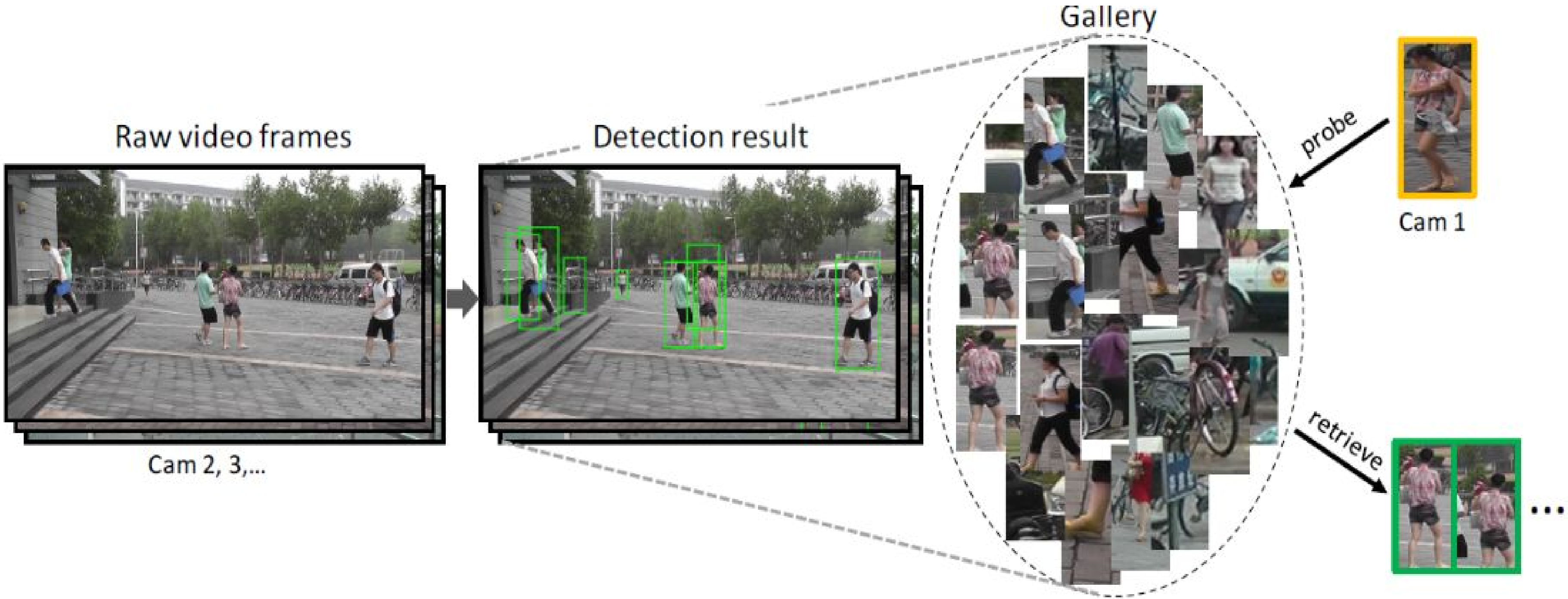
Person re-identification by Local Maximal Occurrence representation and metric learning

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

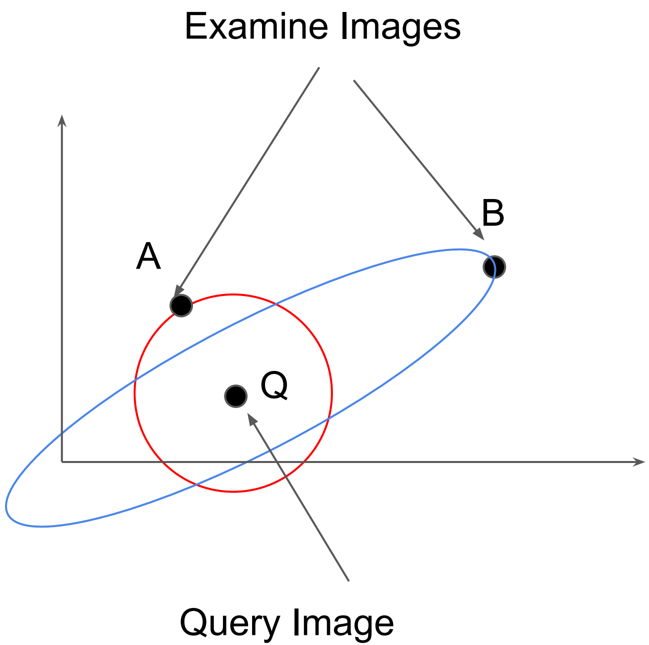
Person re-identification is a challenging problem because:

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

Contribution

- Extract good features ***Local Occurrence Maximum (LOMO)***
- Use good distance metric ***Cross-view Quadratic Discriminant Analysis (XQDA)***

About distance metric



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

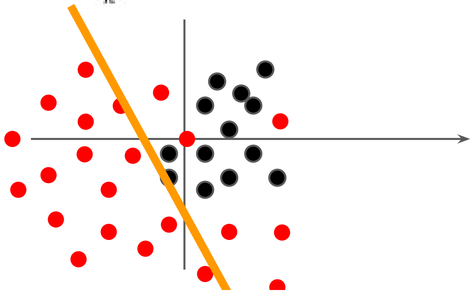
$$f_k(x) = \frac{1}{(2\pi)^{p/2}|\sum_k|^{1/2}} \exp^{-\frac{1}{2}(x-\mu_k)^T \sum_k^{-1}(x-\mu_k)}$$

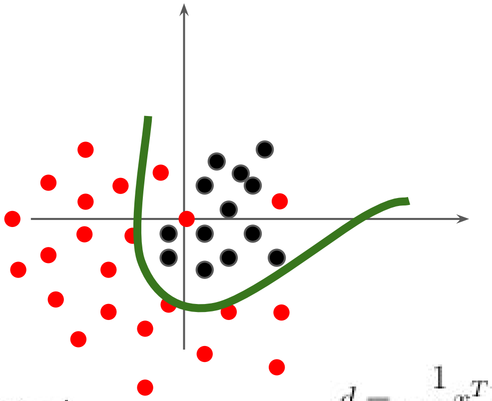
Linear Discriminant Analysis (LDA)

Quadratic Discriminant Analysis (QDA)

Cross-view Quadratic Discriminant Analysis (XQDA)

$$\sum_k = \sum, \forall k$$





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

$$d = -\frac{1}{2}x^T\Sigma^{-1}x - \frac{1}{2}log\frac{|\Sigma_k|}{|\Sigma_l|} + log\frac{\pi_k}{\pi_l}$$

Cross-view Quadratic Discriminant Analysis (XQDA)

Intuition: Original feature space is too high dimension.
Maybe it's helpful to consider the problem in subspace

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

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

Hard to measure precisely in high dimension space

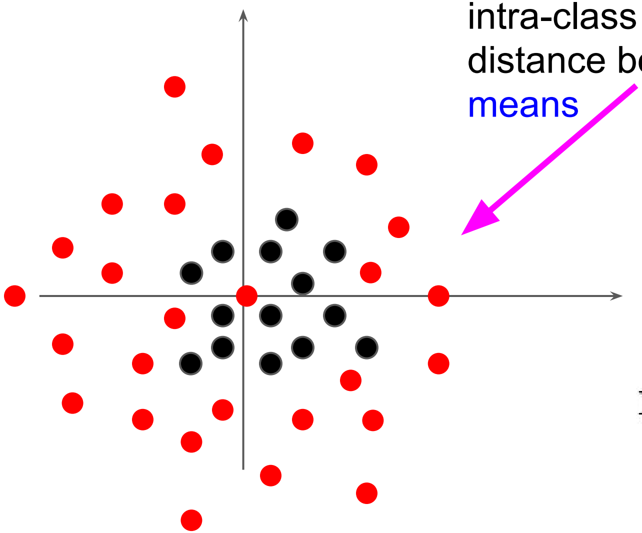
Measure this in subspace!

Which subspace ?

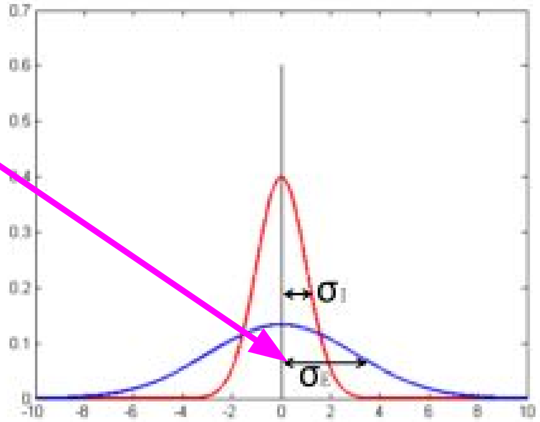
What about PCA ?

Cross-view Quadratic Discriminant Analysis (XQDA)

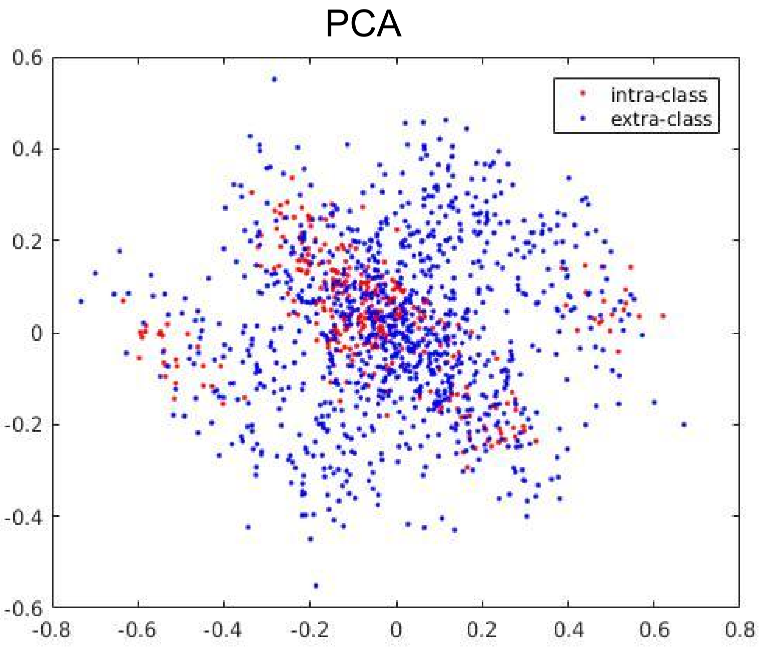
The two distribution for intra-class and extra-class distance both have zero means

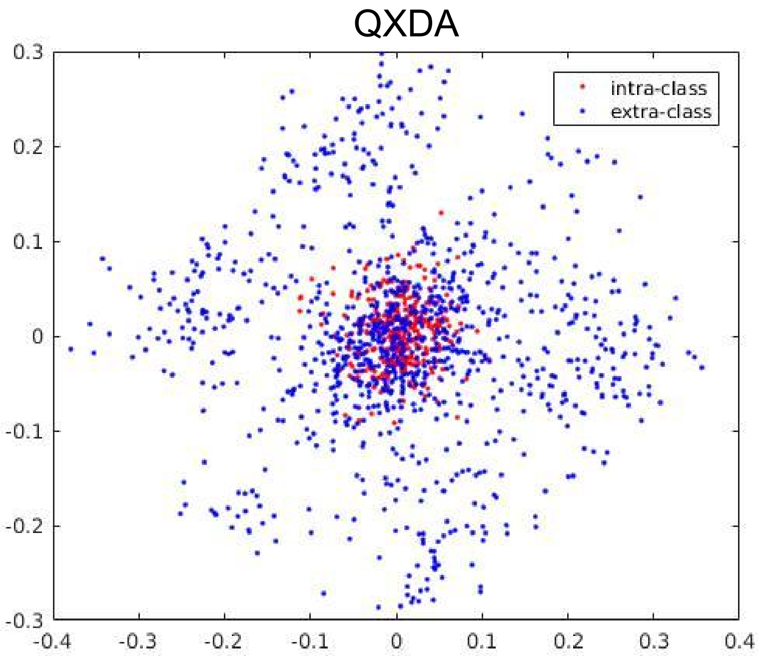


$$\max_{w} \frac{\sigma_E(w)}{\sigma_w(w)}$$



The QXDA chose subspace that maximize the two classes' variance ratio



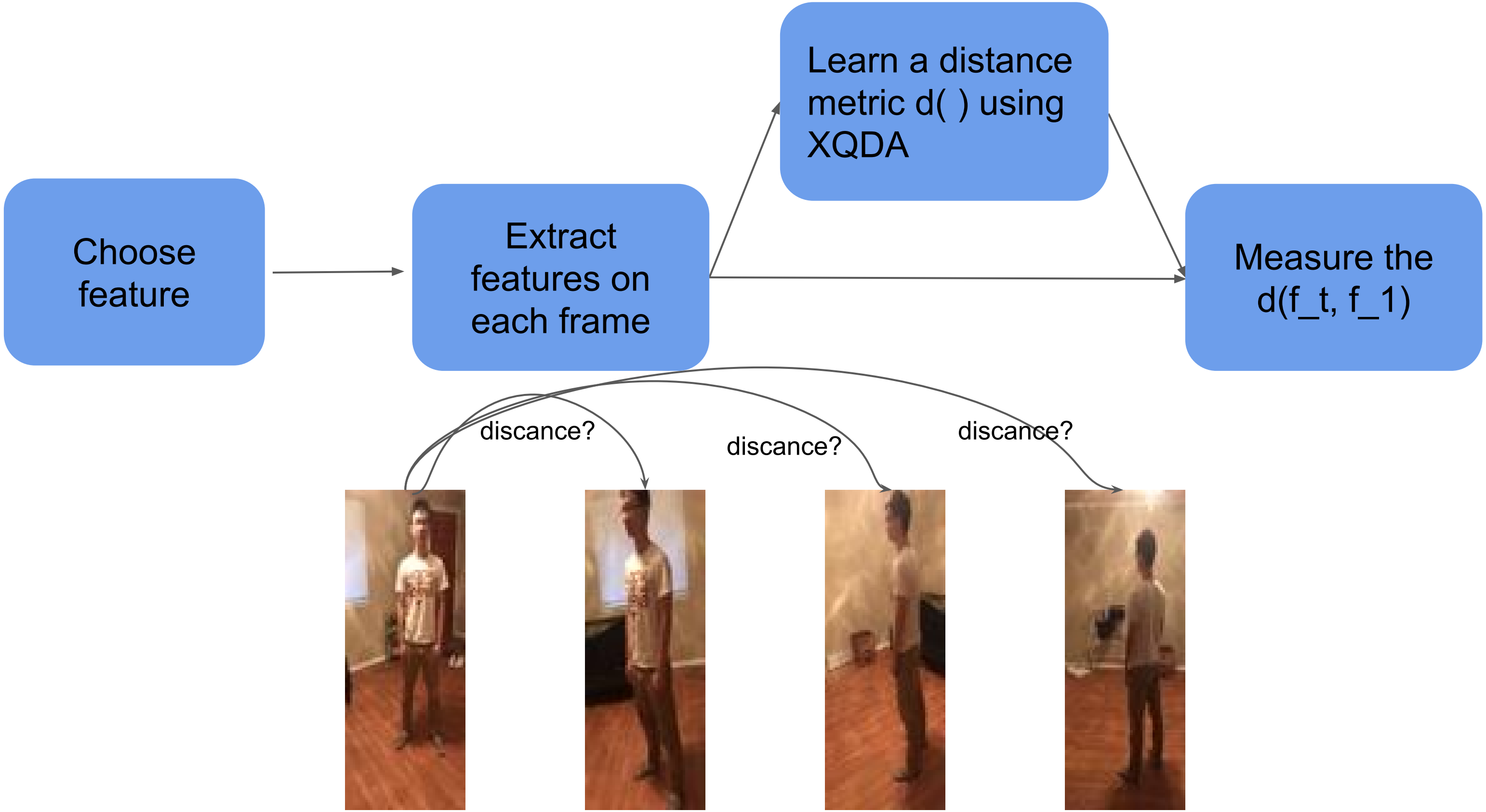


Viewpoint Invariance Analysis

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

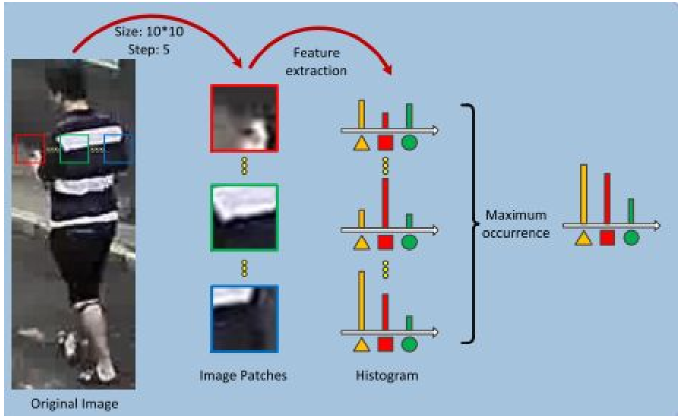
Slides credit: my roomate

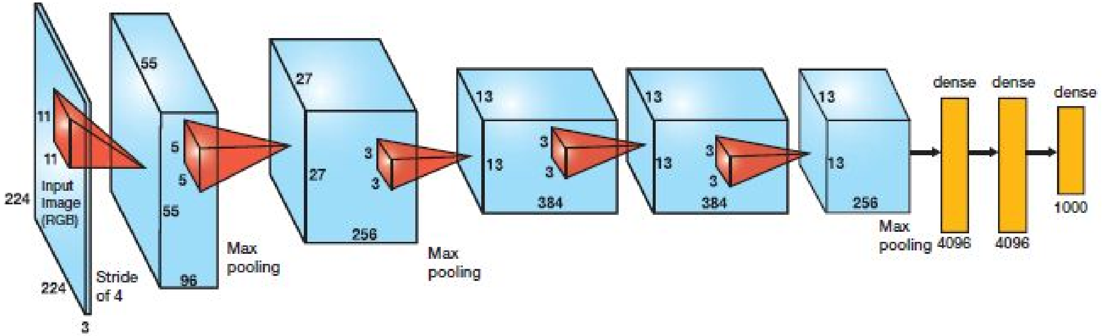
Viewpoint Invariance Analysis



Investigated Features

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

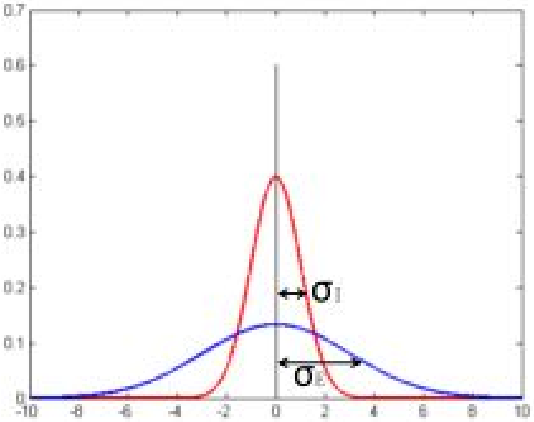




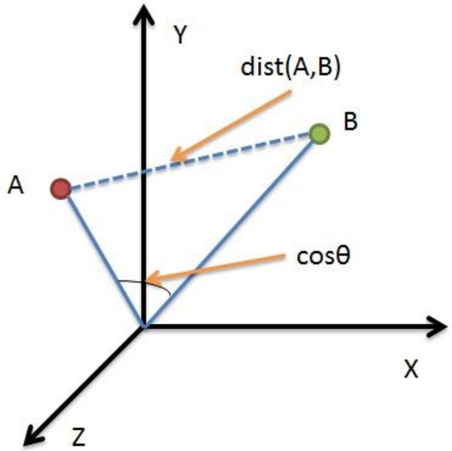
Distance Metric

- Quadratic Discriminant Analysis (XQDA)
- Cosine Similarity

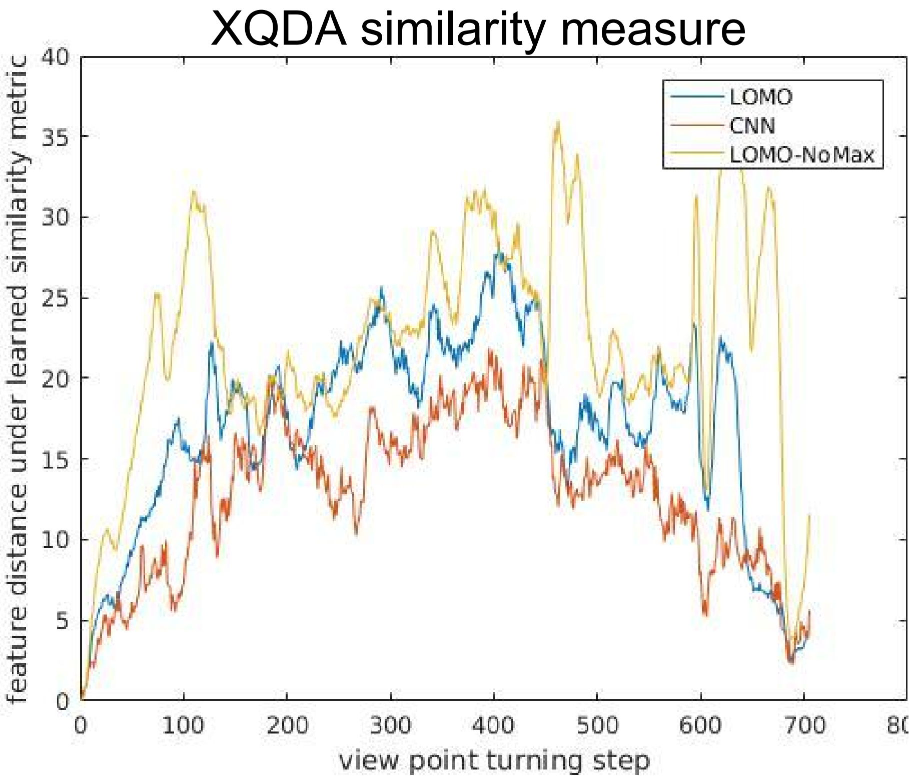
$$\max_{w} \frac{\sigma_E(w)}{\sigma_w(w)}$$

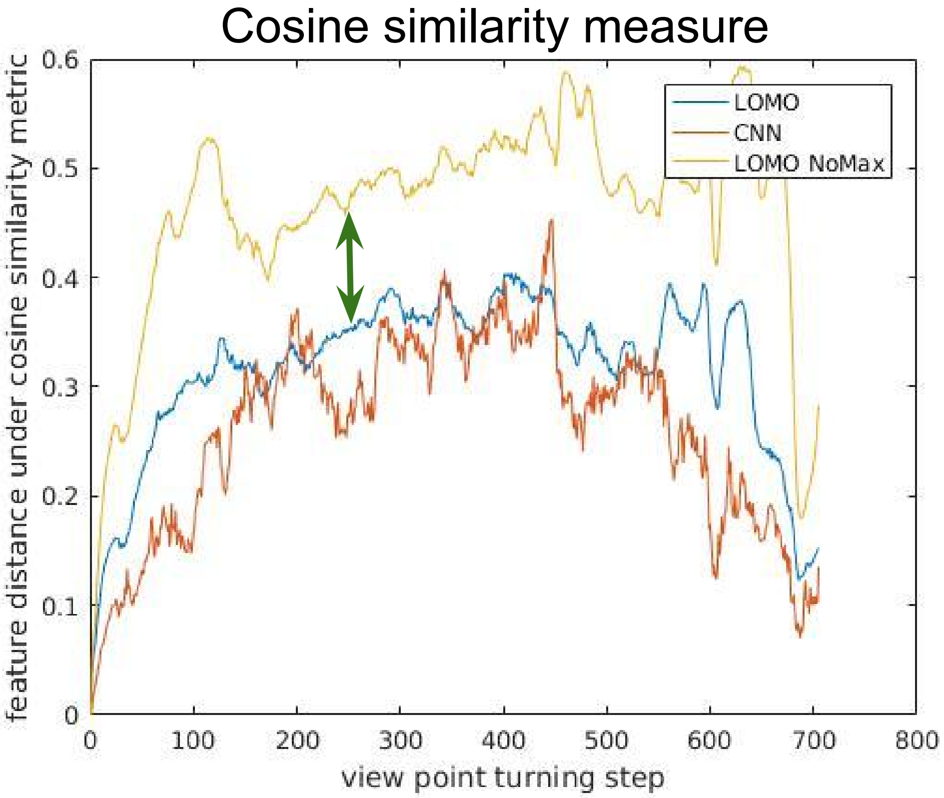


$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$



- ***The max operation in LOMO makes it more robust to viewpoint change***
- ***XQDA can learn more robust metric against viewpoint variation***

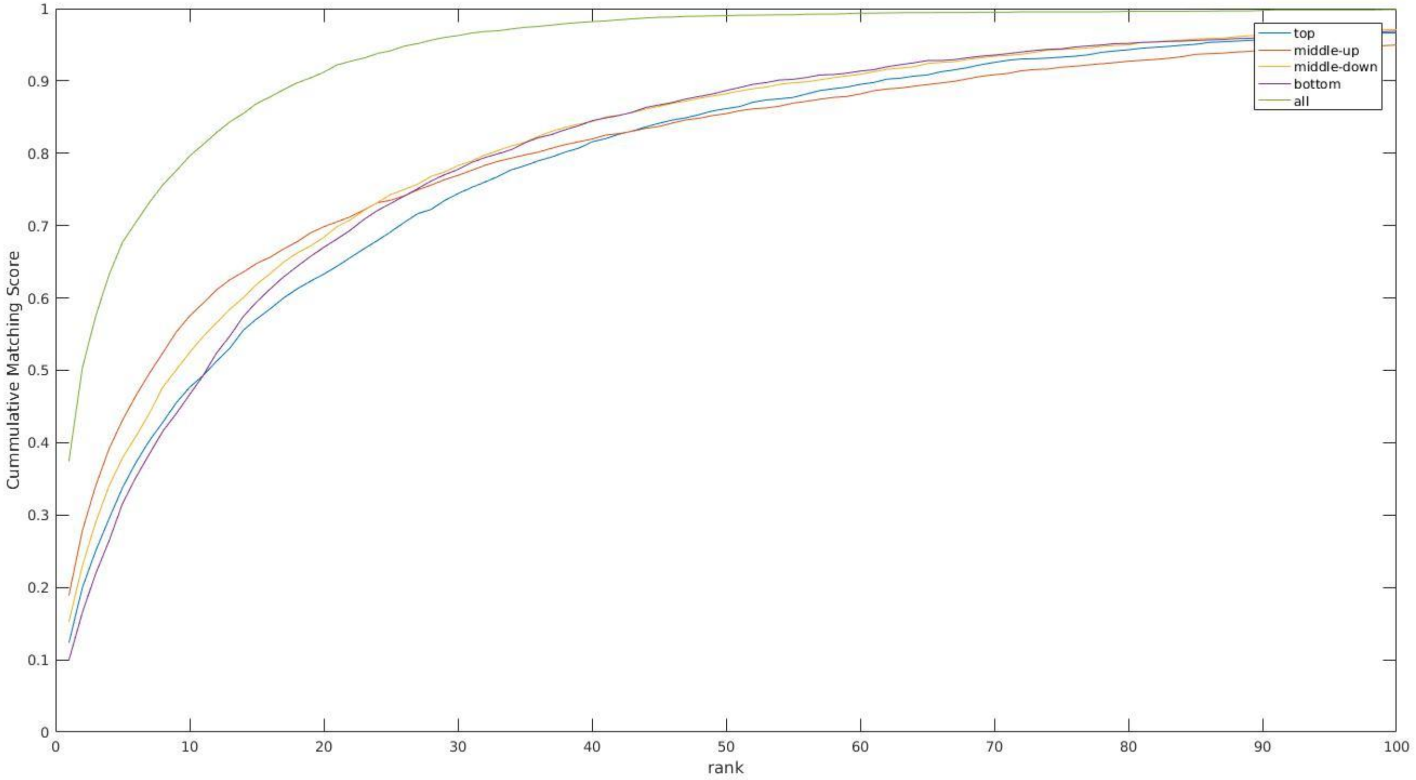




Which region contribute mostly?

- Conduct training on four different body parts
- Compute the matching performance using each body parts

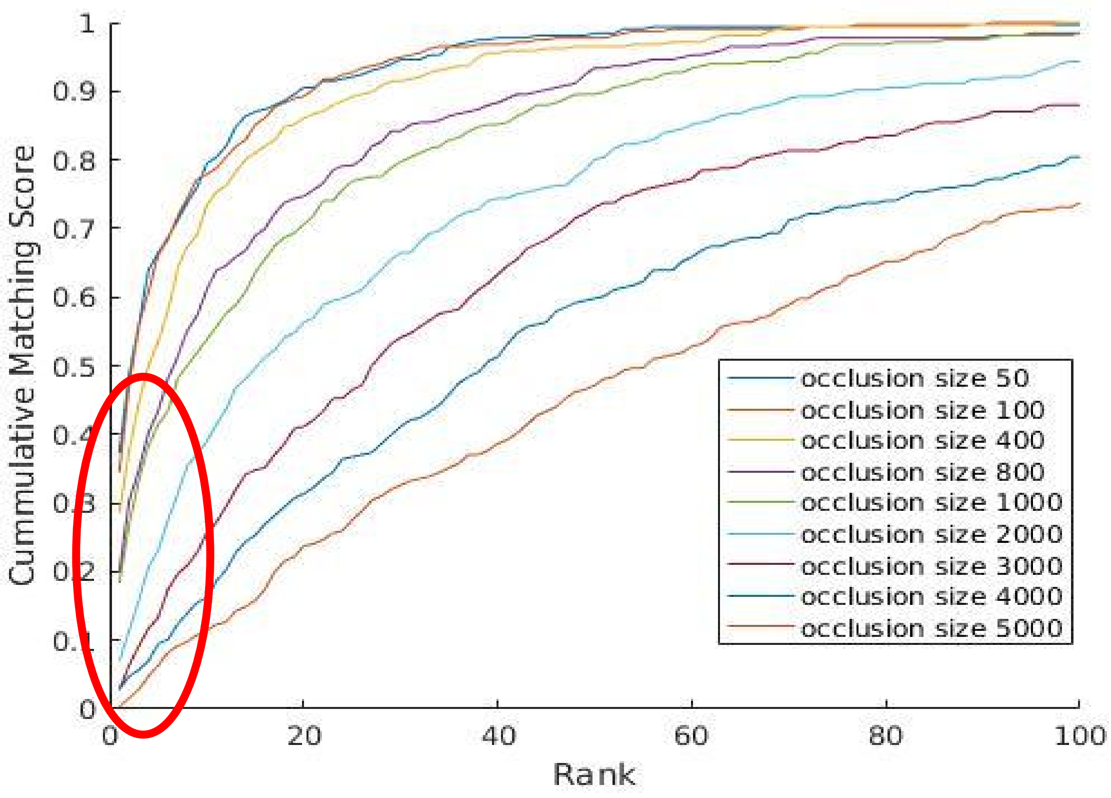
The upper body is the most distinguishable part



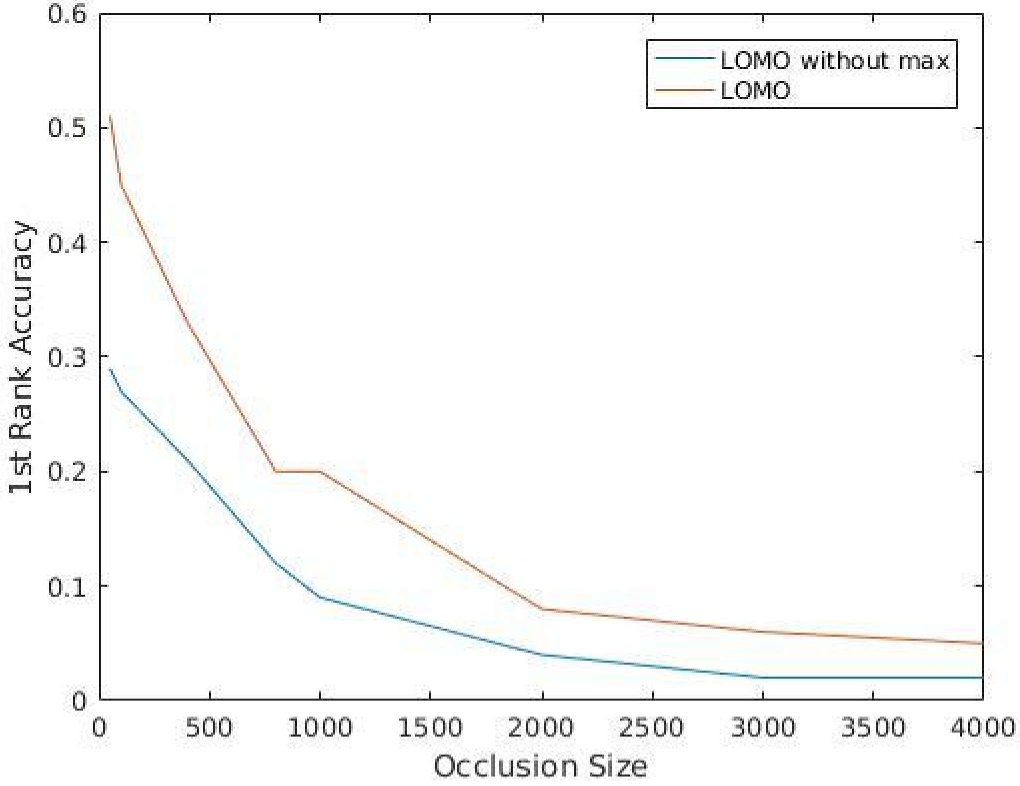
Sensitivity to Occlusion

Parameter: ***the size of occlusion area***

The performance degrades monotonous as occlusion become more severe



1st rank accuracy degrades monotonous 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***