

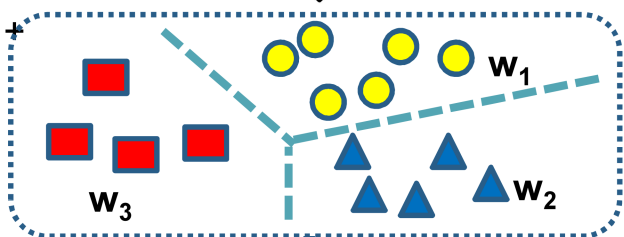
Input
video



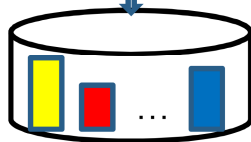
Local
descriptors



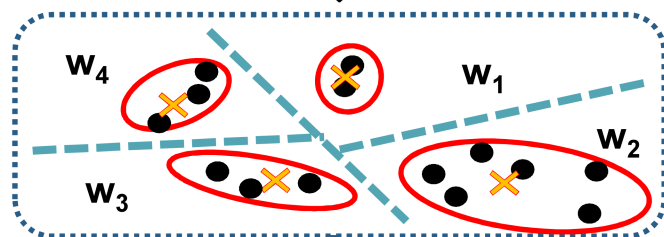
K-Means +
Bag-of-
words
Encoding



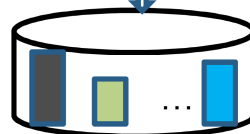
Final feature
representation



GMM +
Fisher
Vector
Encoding



w_i : GMM cluster
X : mean of local
descriptors
O : variance of local
descriptors



Classification
(χ^2 -SVM + Pre-computed Kernel)

Classification
(Linear SVM + Pre-computed Kernel)