Learning Better Object Models using Video Data

Patrick Li, Inmar Givoni, Brendan Frey

Motivation

Training on a collection of static monocular images is unnatural.

Labelled Training Images are hard to get. And the lack of is becoming a problem.

There is a wealth of video data available.

First Attempt: Learning Bags of Features Models for Image Classification

Goal:

Represent Objects as Bags of SIFT Features

Use unsupervised learning to learn models of objects

Use learned models for image classification

Image Classification

INPUT:

OUTPUT:

“Cow”

TRAINING:

“Boat”

“Car”

“Sofa”

...

Overview of the Technique

Unsupervised Training from Video

PART 1 | PART 2 | PART 3 | • • • | PART 60

Supervised Training on Labelled Images

"Cow"

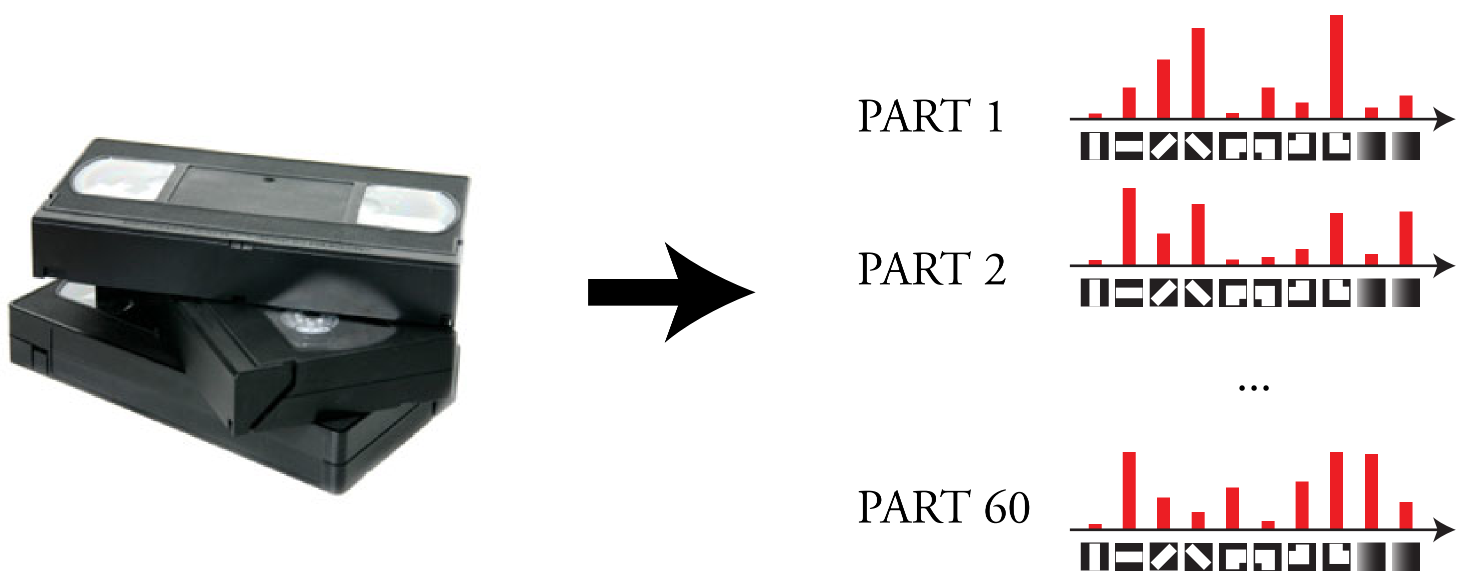
PART 2 | PART 3

Testing

PART 1 | PART 8

"Sofa"

Bags of Features Models



Latent Dirichlet Allocation for Topic Modelling

Convex Clustering with Exemplar-Based Models

Danial Lashkari **Polina Golland**

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139
{danial, polina}@csail.mit.edu

Abstract

Clustering is often formulated as the maximum likelihood estimation of a mixture model that explains the data. The EM algorithm widely used to solve the resulting optimization problem is inherently a gradient-descent method and is sensitive to initialization. The resulting solution is a local optimum in the neighborhood of the initial guess. This sensitivity to initialization presents a significant challenge in clustering large data sets into many clusters. In this paper, we present a different approach to approximate mixture fitting for clustering. We introduce an exemplar-based likelihood function that approximates the exact likelihood. This formulation leads to a convex minimization problem and an efficient algorithm with *guaranteed convergence to the globally optimal solution*. The resulting clustering can be thought of as a probabilistic mapping of the data points to the set of exemplars that minimizes the average distance and the information-theoretic cost of mapping. We present experimental results illustrating the performance of our algorithm and its comparison with the conventional approach to mixture model clustering.

1 Introduction

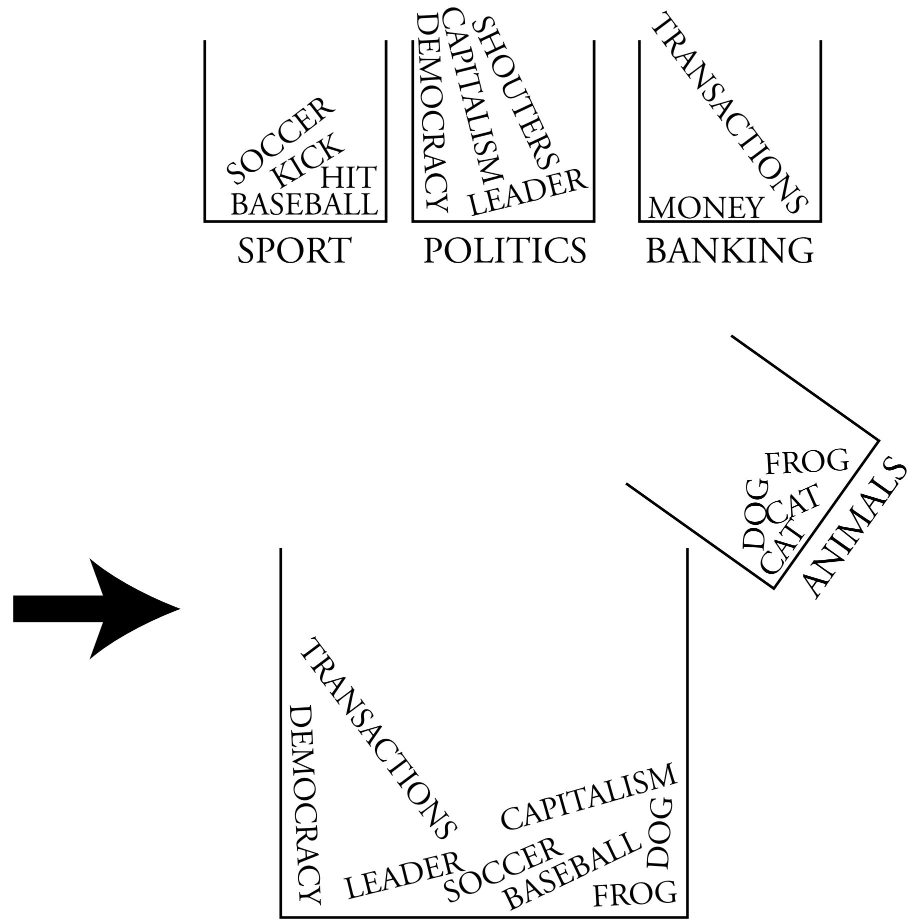
Clustering is one of the most basic problems of unsupervised learning with applications in a wide variety of fields. The input is either *vectorial* data, that is, vectors of data points in the feature space, or *proximity* data, the pairwise similarity or dissimilarity values between the data points. The choice of the clustering cost function and the optimization algorithm employed to solve the problem determines the resulting clustering [1]. Intuitively, most methods seek *compact* clusters of data points, namely, clusters with relatively small intra-cluster and high inter-cluster distances. Other approaches, such as Spectral Clustering [2], look for clusters of more complex shapes lying on some low dimensional manifolds in the feature space. These methods typically transform the data such that the manifold structures get mapped to compact point clouds in a different space. Hence, they do not remove the need for efficient compact-cluster-finding techniques such as k-means.

The widely used Soft k-means method is an instance of maximum likelihood fitting of a mixture model through the EM algorithm. Although this approach yields satisfactory results for problems with a small number of clusters and is relatively fast, its use of a gradient-descent algorithm for minimization of a cost function with many local optima makes it sensitive to initialization. As the search space grows, that is, the number of data points or clusters increases, it becomes harder to find a good initialization. This problem often arises in emerging applications of clustering for large biological data sets such as gene-expression. Typically, one runs the algorithm many times with different random initializations and selects the best solution. More sophisticated initialization methods have been proposed to improve the results but the challenge of finding good initialization for EM algorithm remains [4].

We aim to circumvent the initialization procedure by designing a convex problem whose *global* optimum can be found with a simple algorithm. It has been shown that mixture modeling can

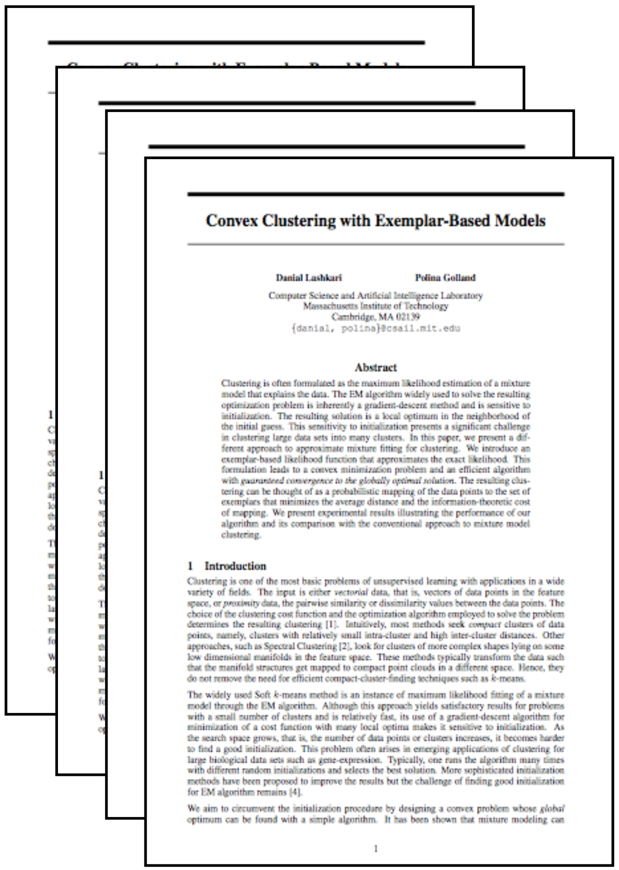
1

Single Document

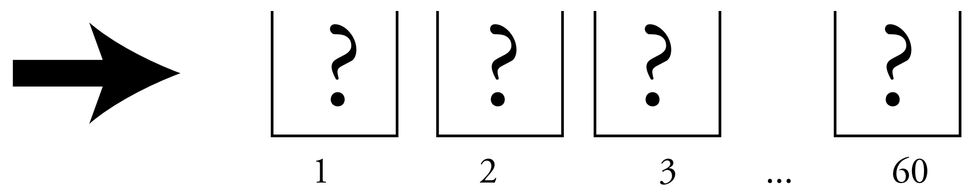


20% ANIMALS
40% POLITICS
39% BANKING
1% SPORTS

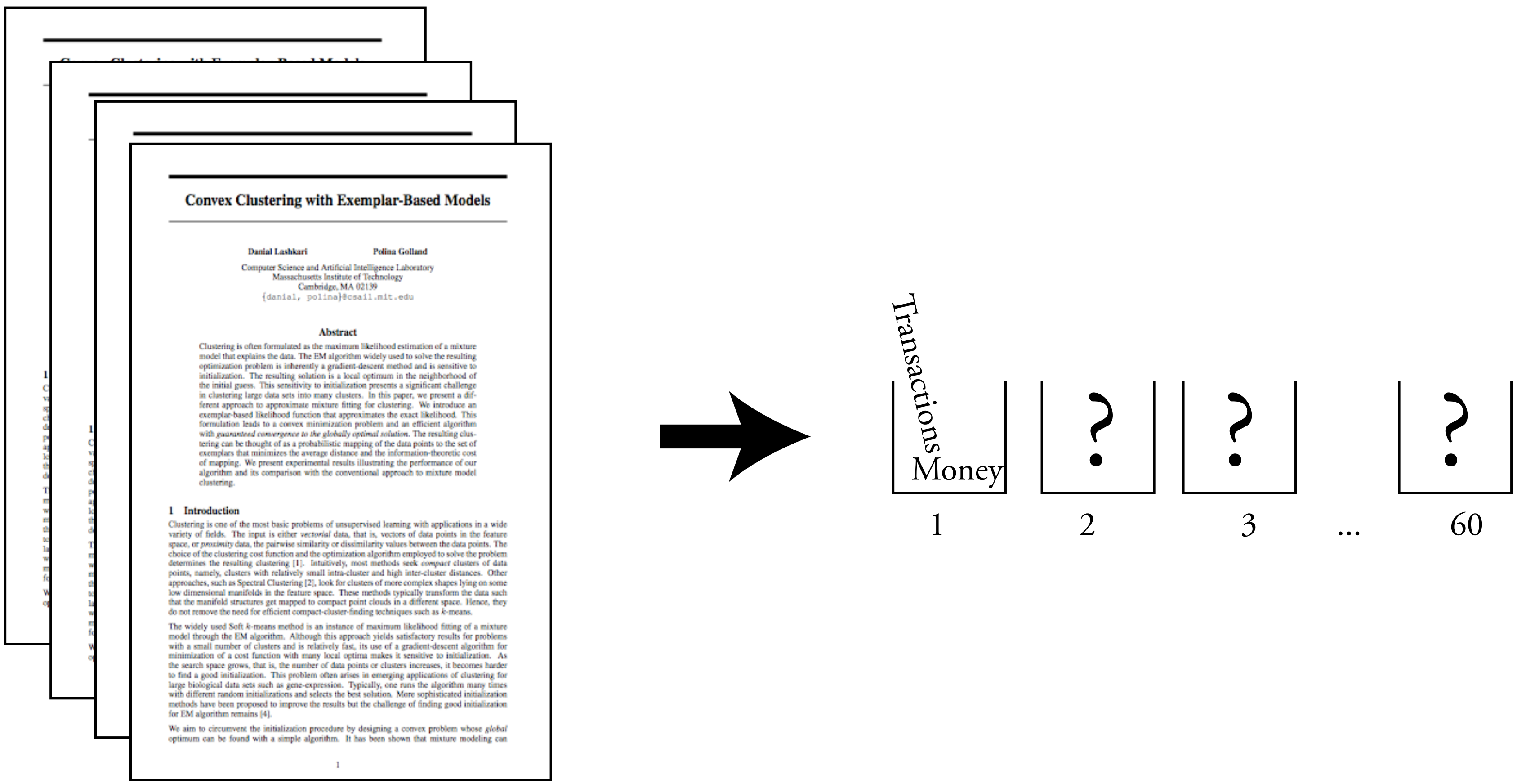
Latent Dirichlet Allocation for Topic Modelling



Corpus of Documents

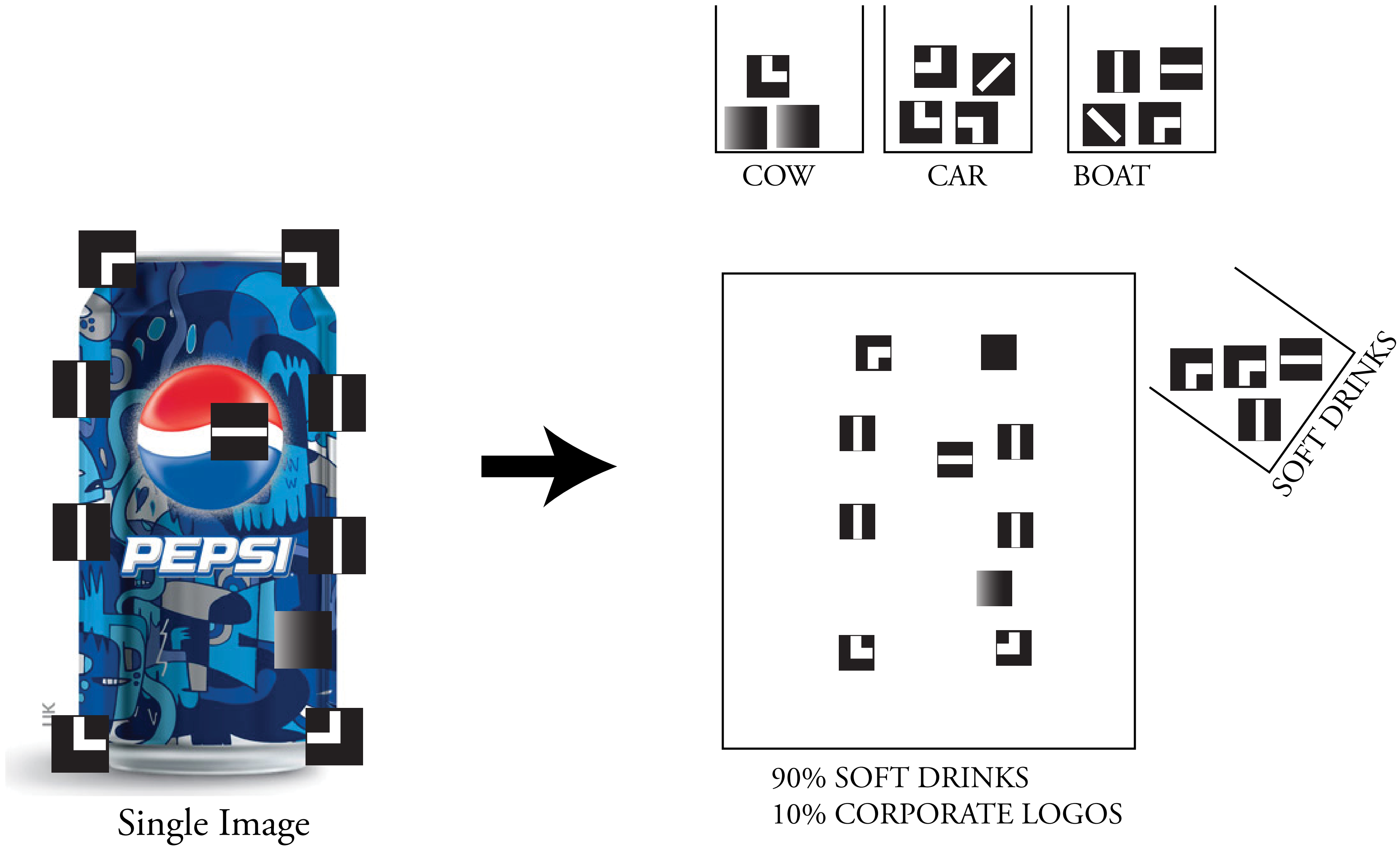


Latent Dirichlet Allocation for Topic Modelling



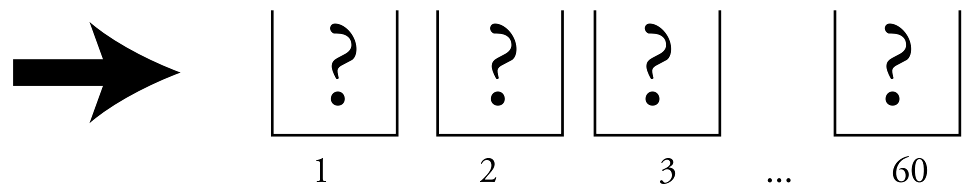
Corpus of Documents

Latent Dirichlet Allocation for Object Modelling

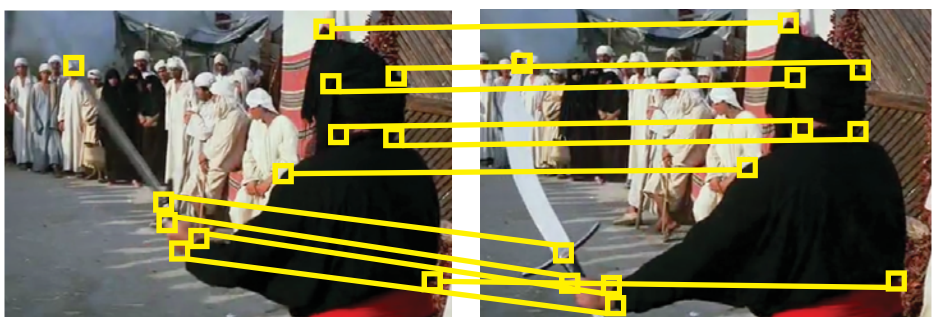


Latent Dirichlet Allocation for Object Modelling

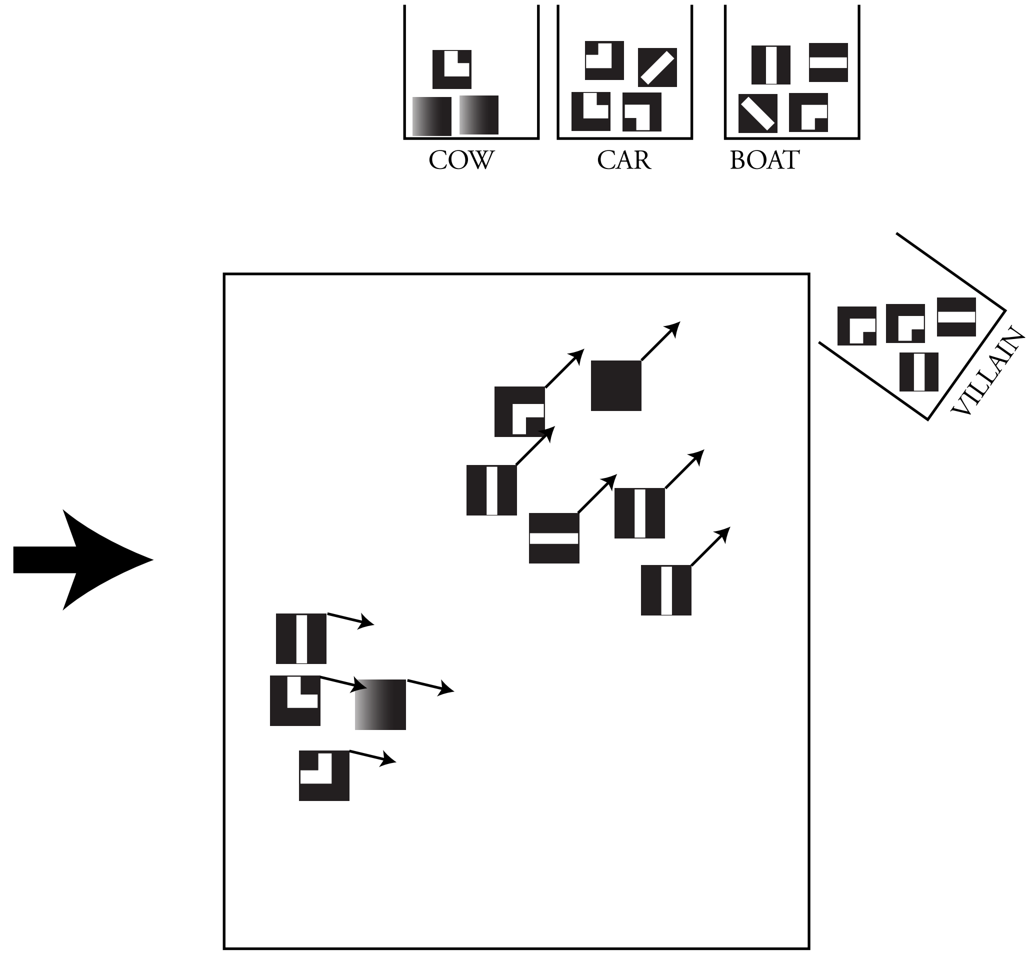
Image Collection



Flow-LDA for Motion Modelling

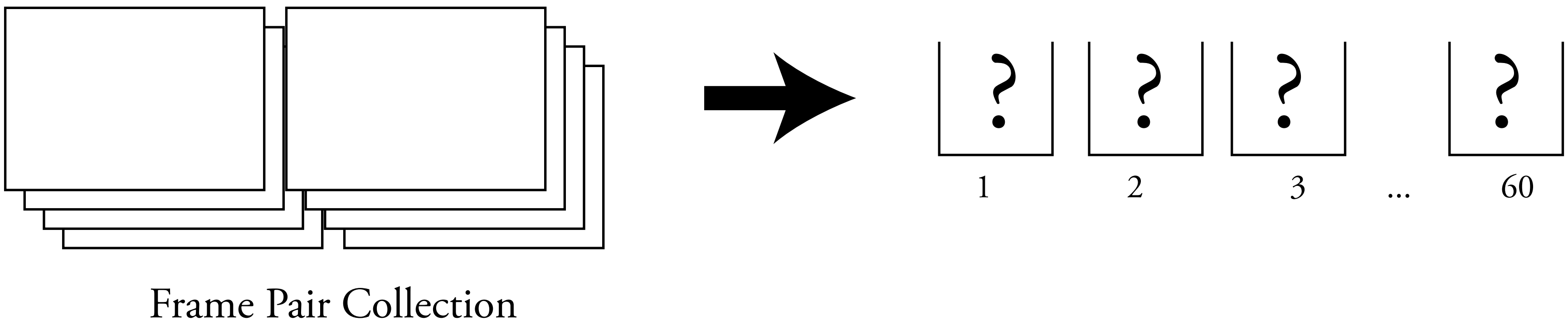


Pair of Consecutive Frame Pairs



50% SWORD
50% VILLAIN

Flow-LDA for Motion Modelling



Flow-LDA for Motion Modelling

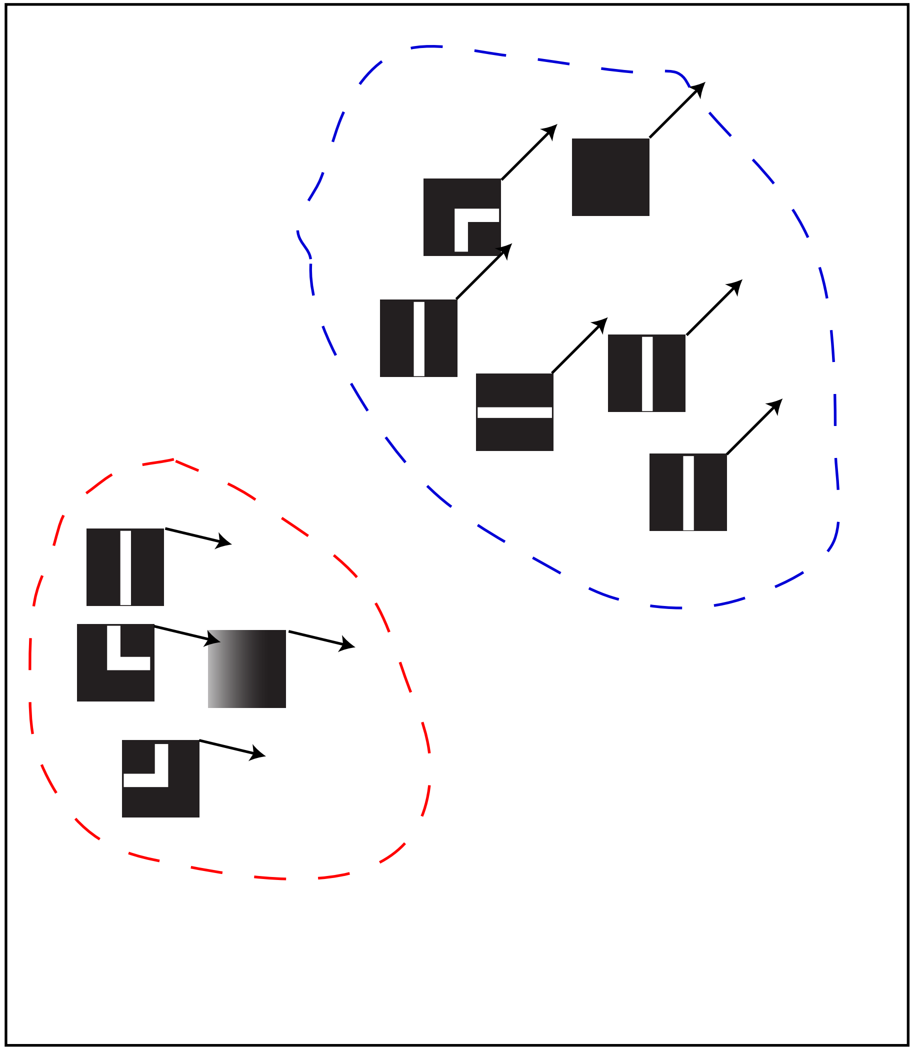
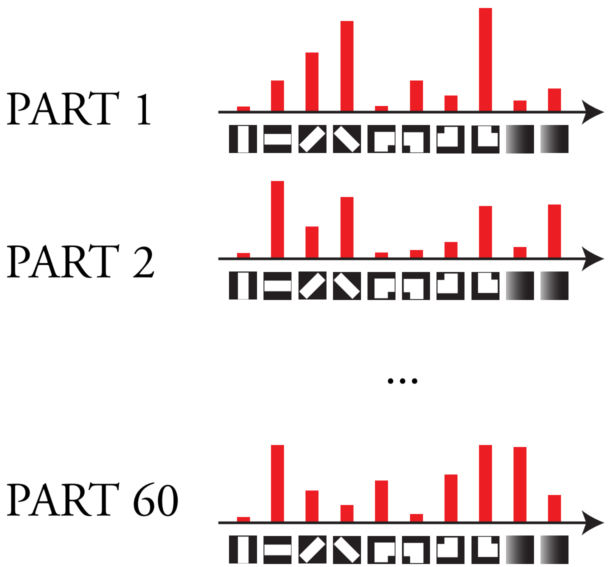
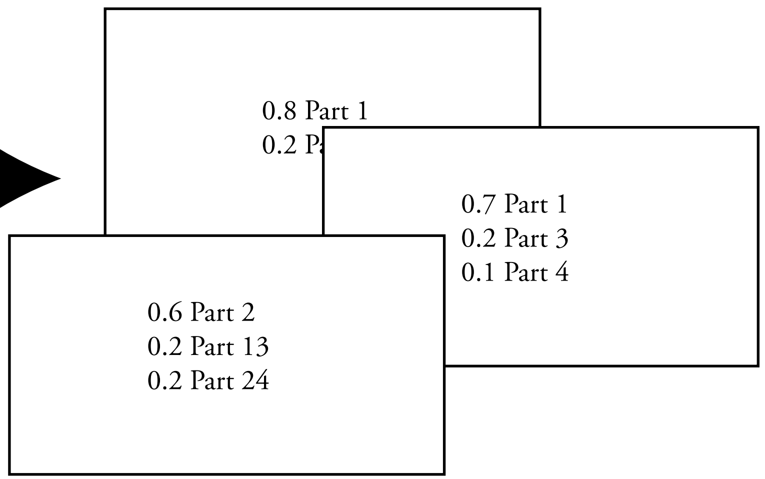


Image Recognition

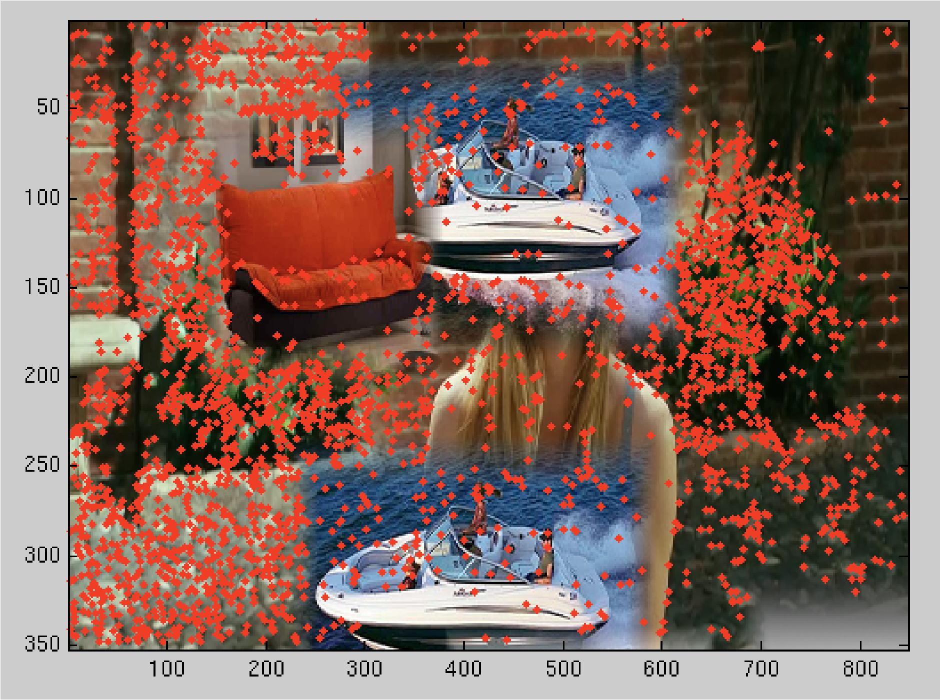
Unsupervised Training from Video using FLDA



Training And Testing Images



Initial Results



Naive Guesser: 8.6% Error
SVM trained on SIFT histograms directly: 8.6% Error
SVM trained using LDA model (no motion): 5.6% Error
SVM trained using FLDA model (motion): 3.7% Error

... to continue

Experiment on Real Dataset

Go beyond Bags of Features models
- -Hierarchical Models
- -Account for Spatial Relations
- -Account for temporal relations between more than 2 frames

Thank you!