Detecting and Naming Actors in Movies using Generative Appearance Models

Vineet Gandhi, and Remi Ronfard

vineet.gandhi@inria.fr, remi.ronfard@inria.fr

Introduction

We introduce a generative model for learning person and costume specific detectors from labeled examples. We demonstrate the model on the task of localizing and naming actors in one long movie [3], sampled every 10 seconds.

Contributions

- A complete framework to learn view-independent actor models using MSCR [1] features with a novel clustering algorithm
- Two stage detection framework, a search space reduction using the k-nearest neighbours and a sliding window search for the best localization of the actor in position and scale

Generative Model

- Head AND Shoulder representation, each as a constellation of optional MSCR regions
- Arbitrarily complex actor models
- Each Actor associated with visual vocabulary of cluster centers C_i and respective frequencies H_i
- C_i is represented in a 9 dimensional space (position, color, size and shape)

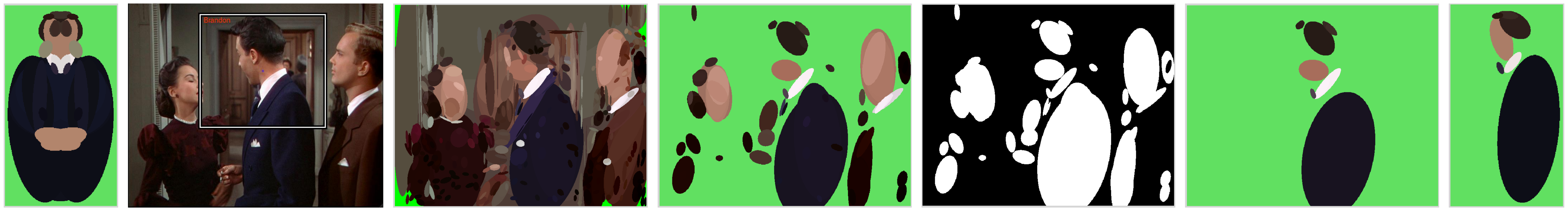
Example of Actor Appearance Models

Formally, our generative model for each actor consists of the following three steps:

1. Choose screen location and window size for the actor on the screen, using the detections in the previous frame as a prior
2. Choose visible features C_i in the "head" and "shoulder" regions independently, each with a probability H_i
3. For all visible features C_i , generate color blob B_i from a gaussian distribution with mean C_i and covariance Σ_i, then translate and scale to the chosen screen location and size

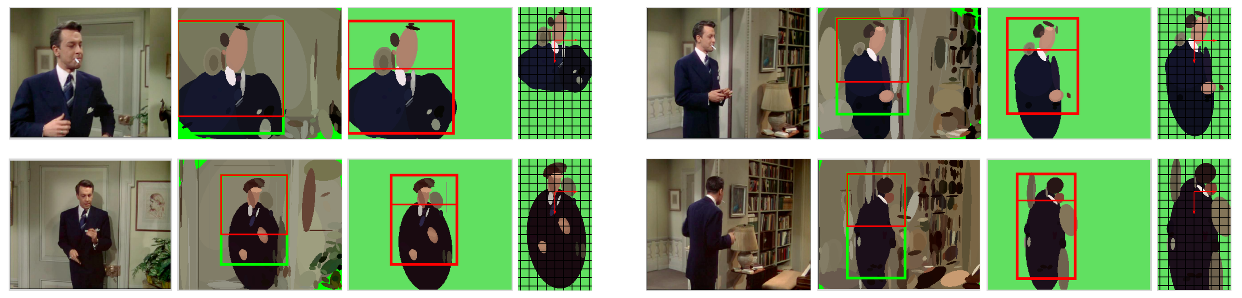
$$P(B_j, m_{ij}, a) = \sum_i H_i m_{ij} \exp\left\{-\left(C_i^a - B_j\right)^T \Sigma_i^{a-1} \left(C_i^a - B_j\right)\right\}$$

Overview of Detection Process



Model | Detected Window | MSCR Features | kNN Refinement | Binary Map | Matched Image Blobs | Clusters

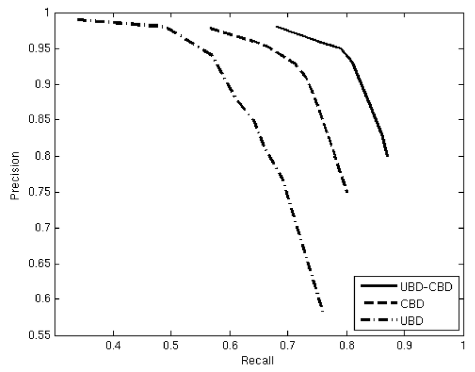
Training Actor Models



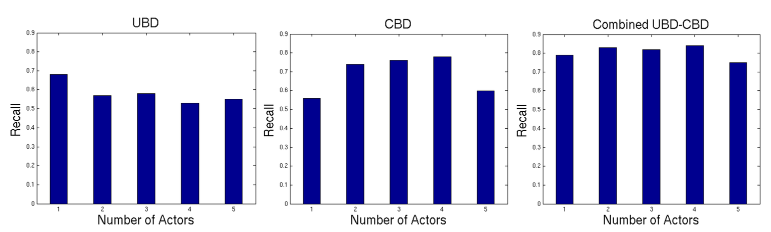
- Constrained agglomerative clustering
- Each blob in the first training image is initialized as a singleton cluster. We then compute pairwise matching with set of blobs from other training images (one by one)
- At each step, for each cluster, we assign at most one more blob. Blobs not assigned to an existing cluster are assigned to their own singleton cluster

Experimental Results

Precision-recall curves compared to a state of the art generic detector [2] on a movie with 8 actors sampled every 10 seconds.



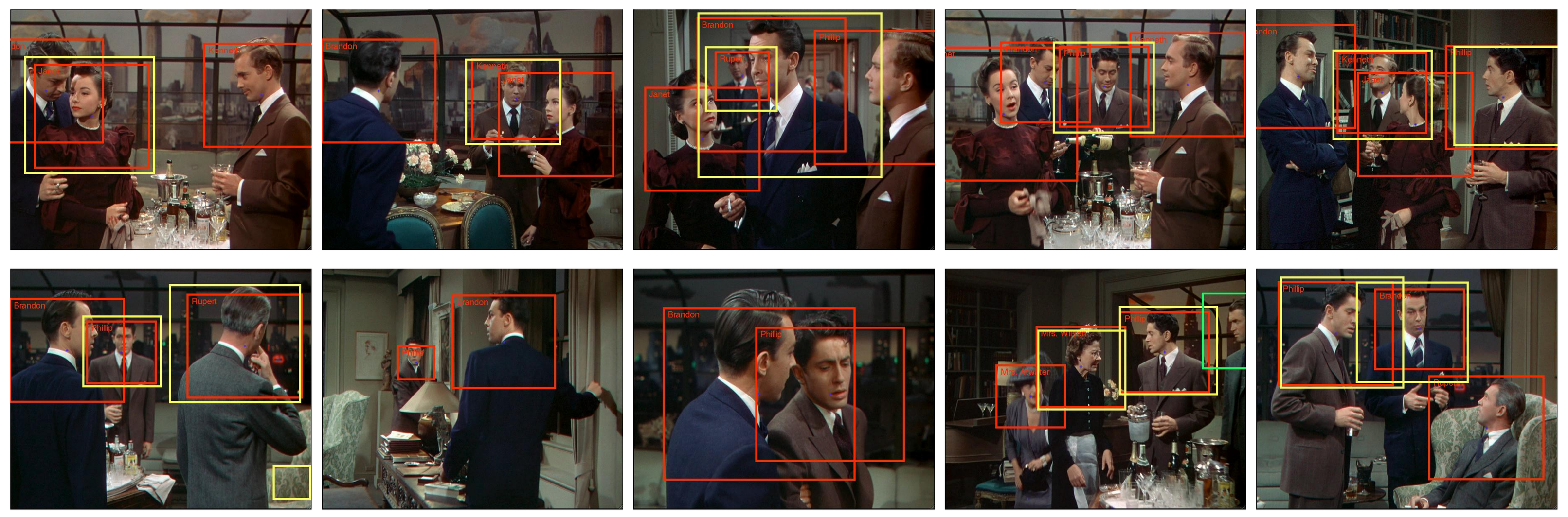
Precision vs Recall



Segmented Recall

	A1	A2	A3	A4	A5	A6	A7	A8
A1	99	0	0	0	1	0	0	0
A2	0	78	0	16	0	2	3	1
A3	0	1	96	0	3	0	0	0
A4	0	25	0	74	0	0	1	0
A5	0	0	5	0	95	0	0	0
A6	0	4	0	0	0	93	0	3
A7	3	4	0	0	1	2	81	9
A8	0	0	0	0	0	6	22	72

Identification



Complete results and dataset:
http://imagine.inrialpes.fr/people/vgandhi/CVPR_2013/

Detection using sliding window search

Each actor detection score is based on the likelihood that the image in the sliding window was generated by the given actors model.

Given Actors Models (C^a, Σ^a, H^a) and image features B
for each actor a **do**
  for each scale s **do**
    Normalize image features w.r.t scale.
    $[IDX, D7] = \text{kNN-SEARCH}(B, C^a, k)$
    Build inverted index i.e. for each unique blob B' in the Knn refined set, store corresponding clusters in C^a and respective distances using IDX and $D7$.
    for each position (x, y) **do**
      Find blob indices J_{head} and $J_{shoulders}$
      Compute m_{ij} using blob indices and inverted indices
      $score(x, y, s, a) = \prod_k (\sum_{j \in J_k} P(B_j, m_{ij}, a))$
    end for
  end for
end for

$[x^*(a), y^*(a), s^*(a)] = \text{argmax} \sum_a (score(x, y, s, a) - t_0)$

Conclusions and Future Work

- View-independent models from few training examples
- Simultaneous detection and identification
- Fast matching using kNN-search
- 80% recall with 90% precision in both detection and identification
- Good candidate for tracking-by-detection and floor-plan view reconstruction

References

[1] Per-Eric Forssen. Maximally Stable Colour Regions for Recognition and Matching. In *CVPR'07*, 2007.

[2] P.F. Felzenszwalb, R.B. Girshick, D. McAllester and D. Ramanan. Object Detection with Discriminatively Trained Part-Based Models In *PAMI'10*, 2010.

[3] Alfred Hitchcock. Rope. *Warner Brothers*, 1948.