Overview: Person Queries in the News

- People tend to be interested mostly in person subjects.
- More queries related to certain people

Current retrieval methods:
- Based mostly on transcript information
- Transcript search locates story, but not necessarily people
- Accurate face recognition is important, current techniques are not very effective on videos.

• People tend to be interested mostly in person subjects.

Goal:
- Accurate face recognition is important, current techniques are not very effective on videos.
- Cluster images of a specific person in few groups
- Make these clusters as coherent as possible

FEATURE EXTRACTION

COLOR:
mean, std of 6x5=30 regions in RGB form
30x6=180 features

PCA: First 40 dims in vector quantized image of 256 colors
ICA: First 40 dims in vector quantized image of 256 colors learning rate=0.5
Combination sets of PCA, ICA and COLOR features are also formed

Choosing the best feature set: 90% of correct faces distributed to
- COLOR: 8 clusters
- PCA: 9 clusters
- PCA+COLOR: 10 clusters
- ICA: 22 clusters

Improving Face Detector Accuracy Using Skin Detection

- Gaussian skin model is formed using representative areas of skin from 30 key frames (28376 skin pixels)
  \[(x - m_3)^T C^{-1}_S (x - m_3) \leq \tau_1\]
- Two methods
  - Average skin pixel value of the face area < Thr1
  - # of pixels < Thr2 (50 pixels)

First prunes the videos using transcript information and locate
the shots where name is mentioned, then
- Using face and skin detectors
- Using textual information
- Extracting useful features
- Clustering faces together and forming representative clusters
- Anchor filtering


together and forming

Anchor filtering: removing clusters that have anchors as representatives

Extending the search space to neighboring shots

Overall retrieval performance:
overall: 122/1461 = 8%
text-and-face: 65/1113 = 6%
text-and-skin: 65/732 = 9%

Comparison of the retrieval performance when shots corresponding to the
shots are retrieved with the neighbors. Notice: when original face detection is used together
with text, locality of skin color is used to improve the performance. Note that the
scores are different. Maximum performance is 58% for Clinton, 53% for Sam Donaldson,
and 32% for Saddam.