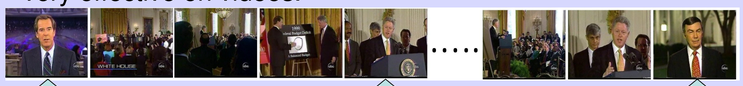
# Person Search Made Easy

Nazlı İkizler, Pınar Duygulu
Bilkent University, Ankara, Turkey

## 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.



(1)…so today it was an energized president **CLINTON** who….. **(name uttered here)**

(5)…*budget marks the hand of an era and ended decades of….* **(Clinton face shows up here)**

(8)….another upward push for mr **CLINTON**'s new sudden… **(name uttered again here)**

- Assumption of people appearing when name is mentioned doesn't always hold:

*Sometimes, they do show up*

*Sometimes, someone else shows up*

*Sometimes, there's no face at all*

**That's why a more automated multi-modal approach for locating people is needed.**

- By using a multi-modal approach, problem of person search can be made easy by reducing the number of results presented to the user:
  - Using face and skin detectors
  - Using textual information
  - Extracting useful features
  - Clustering faces together and forming representative clusters
  - Anchor filtering

## Grouping Similar Faces

**Goal:**

- Cluster images of a specific person in few groups
- Make these clusters as coherent as possible

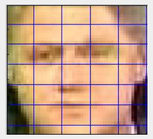
**Method:**

- Using the output of skin-improved face detection method, extract proper features and select the best representative feature for the faces
- Using g-means to cluster the images in few and coherent groups

FEATURE EXTRACTION

COLOR:

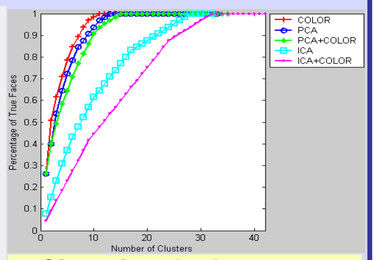
mean, std of 6x5=30 regions in RGB form 30x6=180 features



PCA: First 40 dims in vector quantized image of 256 colors

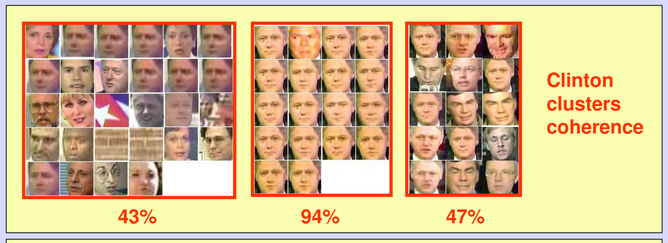
ICA: First 40 dims in vector quantized image of 256 colors learning rate=0.5

*Combination sets of PCA, ICA and COLOR features are also formed*

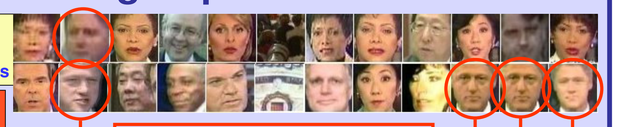



- ***Choosing the best feature set*:** 90% of correct faces distributed to
  - **COLOR: 8 clusters**
  - PCA: 9 clusters
  - PCA+COLOR: 10 clusters
  - ICA: 22 clusters

## Coherence in the clusters




## Retrieval Using Representative Faces




**Instead of looking at 731 images, user just looks at 24 images**

By looking at **5** out of **24 (5/24)** representative faces, **51** out of **65 (51/65)** correct faces are retrieved

| | Shot0 | Shot1 | Shot2 | Shot3 |
|---|---|---|---|---|
| Clinton | (5/24)-(51/65) | (5/44)-(58/138) | (10/72)-(72/158) | (7/66)-(66/170) |
| Sam Donaldson | (9/30)-(35/36) | (8/30)-(76/89) | (8/26)-(98/106) | (8/26)-(101/114) |
| Saddam | (5/22)-(8/8) | (3/26)-(5/13) | (1/30)-(2/14) | (2/30)-(6/14) |

- A single face to represent each cluster is chosen, *representative face*
- When clusters are sufficiently coherent, the user can inspect only representatives instead of all the faces in the cluster

| | Shot0 | Shot1 | Shot2 | Shot3 |
|---|---|---|---|---|
| Clinton | 40% | 39% | 43% | 40% |
| Sam Donaldson | 90% | 81% | 68% | 61% |
| Saddam | 80% | 45% | 100% | 32% |

***Retrieval performance (precision) when representatives are selected***

***Anchor filtering:***

removing clusters that have anchors as representatives

| | Shot0 | Shot1 | Shot2 | Shot3 |
|---|---|---|---|---|
| Clinton | (8/24)-(64/65) | (13/44)-(136/138) | (18/72)-(155/158) | (15/66)-(168/170) |
| Sam Donaldson | (6/30)-(36/36) | (10/30)-(84/89) | (5/26)-(106/106) | (3/26)-(112/114) |
| Saddam | (5/22)-(8/8) | (6/26)-(12/13) | (5/30)-(13/14) | (6/30)-(13/14) |

***Almost all query faces can be found in remaining clusters***

| | Shot0 | Shot1 | Shot2 | Shot3 |
|---|---|---|---|---|
| Clinton | 19% | 14% | 10% | 10% |
| Sam Donaldson | 56% | 56% | 39% | 32% |
| Saddam | 14% | 8% | 6% | 5% |

***However, retrieval performance (precision) is reduced***

***Overall, this system***

- ***reduces the number of images provided to the user extensively***
- ***increases the speed of retrieval by minimal user interaction***
- ***the number of missed images is small***

## Improving Face Detector Accuracy Using Skin Detection

- Gaussian skin model is formed using representative areas of skin from 30 key frames (28376 skin pixels)

$$(x-m_s)^T C_s^{-1}(x-m_s) \le \tau_1$$

- Two methods
  - Average skin pixel value of the face area < $Thr_1$
  - # of pixels < $Thr_2$ (*50 pixels*)

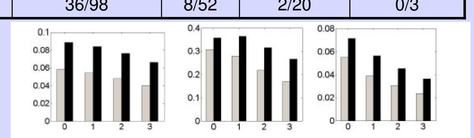
| | Original | Average Skin Color | Number of Skin Pixels |
|---|---|---|---|
| **Precision** | 0.41 | 0.71 | 0.77 |
| **Recall** | 0.40 | 0.38 | 0.38 |

- First prune the videos using transcript information and locate the shots where name is mentioned, then
  - Using only face detection (Mikolajcyzk's face detector)
  - Using skin-improved face detection

| | Clinton | Saddam | Sam Donaldson | Yeltsin | Netanyahu | Henry Hyde |
|---|---|---|---|---|---|---|
| **Text-and-face** | 65/1113 | 8/127 | 36/114 | 8/69 | 4/35 | 1/3 |
| **Text-and-skin** | 65/732 | 8/98 | 36/98 | 8/52 | 2/20 | 0/3 |

**Overall retrieval performance :**
text-and-face: 122/1461 = 8%
text-and-skin: 119/1003 = 12%

Extending the search space to neighboring shots




*Comparison of the retrieval performance when shots corresponding to the text are extended with the neighbors.* **Gray:** *when original face detection is used together with text,* **black:** *when skin color is used to improve the performance. Note that the scales are different. Maximum performances are 9% for Clinton, %36 for Sam Donaldson and %7 for Saddam queries.*