Finding People Frequently Appearing in News

Derya Ozkan, Pınar Duygulu

Bilkent University, Department of Computer Engineering
RETINA Vision and Learning Group

http://retina.cs.bilkent.edu.tr

Motivation

- Finding specific people in news videos is important.
- **Common approach:**
  - Search for the names of the people in the speech transcript text.
- **Problem:** likely to produce incorrect results due to:
  - Time variations between visual appearance of face and name.
  - Non-face images extracted by the face detector algorithm.
  - Faces of other people appearing in the same frame.

- Faces should be recognized in order to solve the problem. However, recognizing faces is a long standing and difficult problem.

Approach

- In news, a person visually appears around when his/her name is mentioned.

- If the relation between the **name & visual appearance** of the person is learned, then face recognition and retrieval can be simplified.

Proposed Solution

Step 1: Limit the search space of a particular person by using speech transcript text.

Step 2: Construct a similarity graph among faces in this search space.

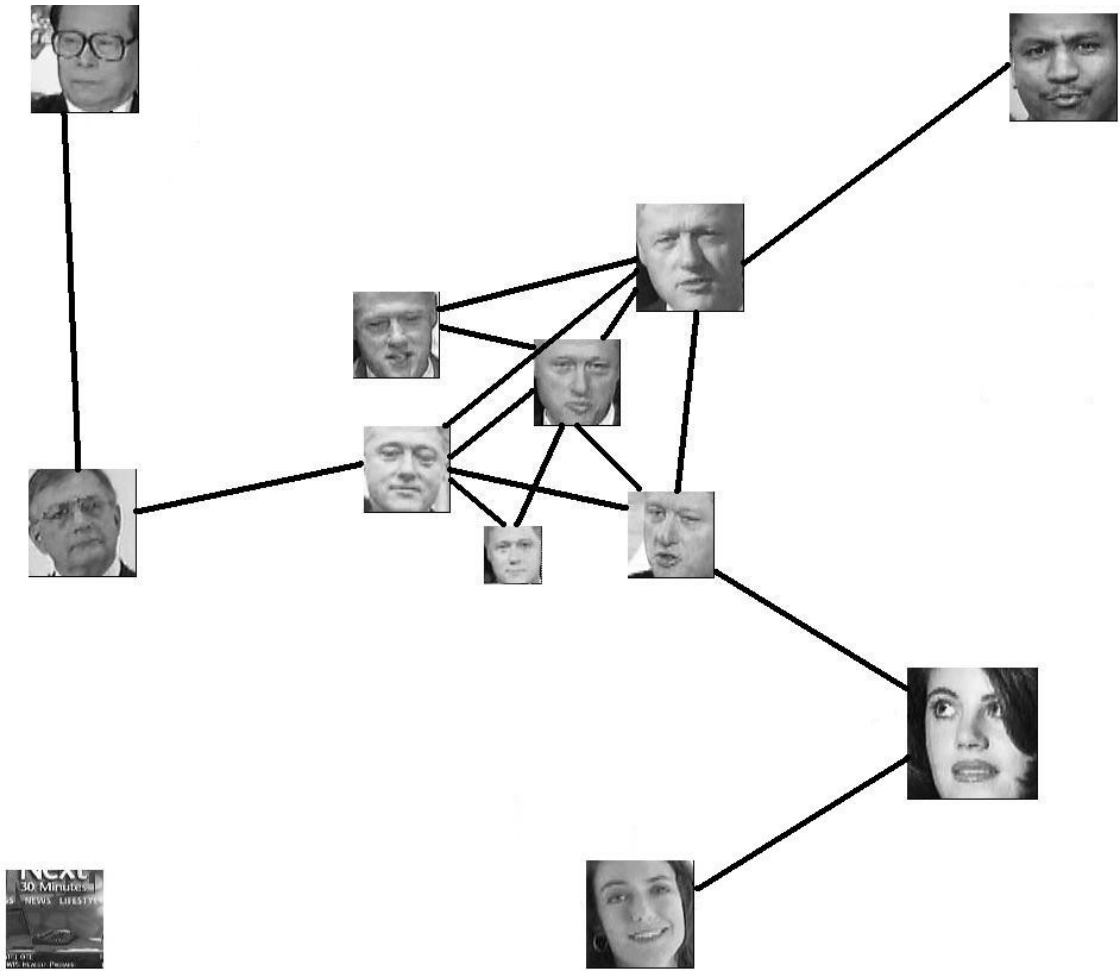
Step 3: Find the densest component in this graph.

Proposed Solution

Step 1: Limit the search space of a particular person by using speech transcript text.

Step 2: Construct a similarity graph among faces in this search space.

Step 3: Find the densest component in this graph.

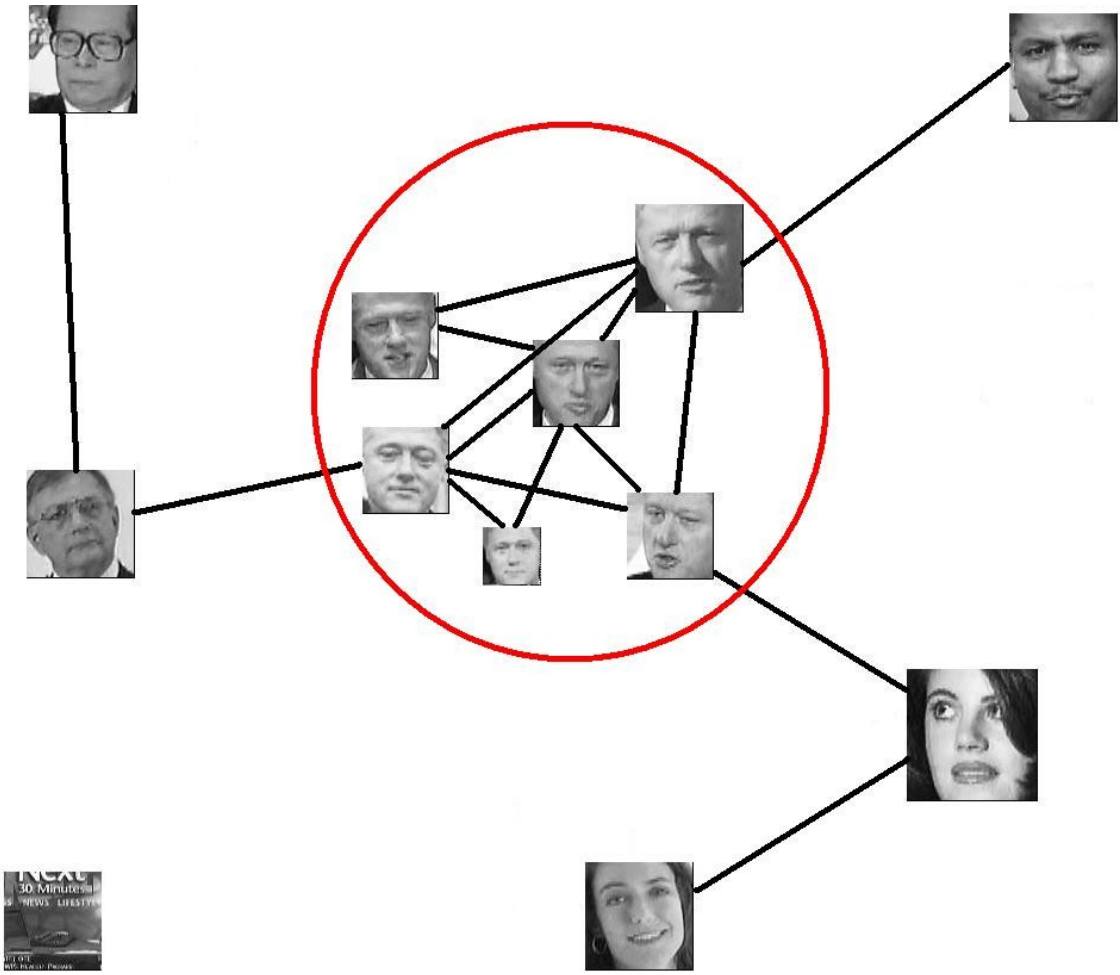


Proposed Solution

Step 1: Limit the search space of a particular person by using speech transcript text.

Step 2: Construct a similarity graph among faces in this search space.

Step 3: Find the densest component in this graph.



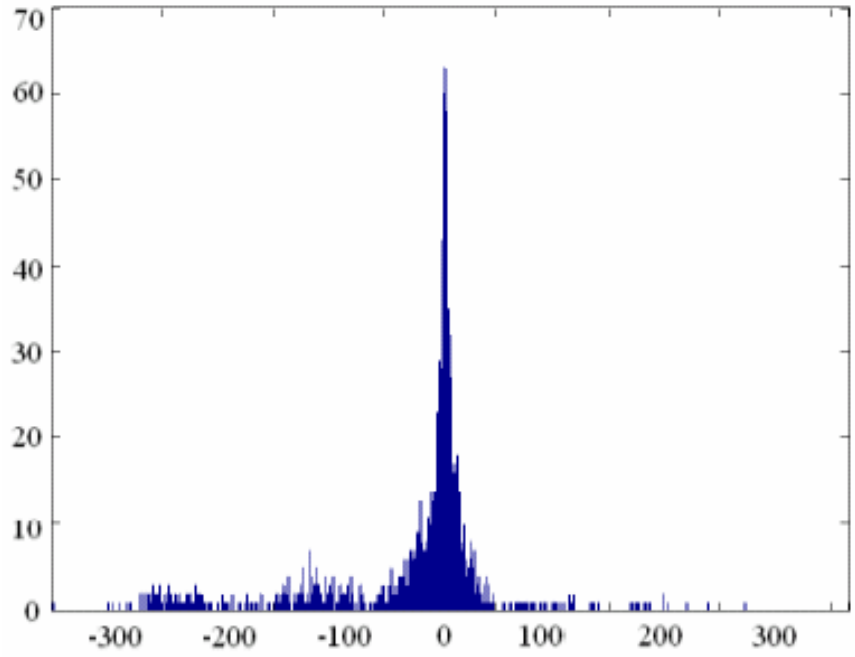
Proposed Solution

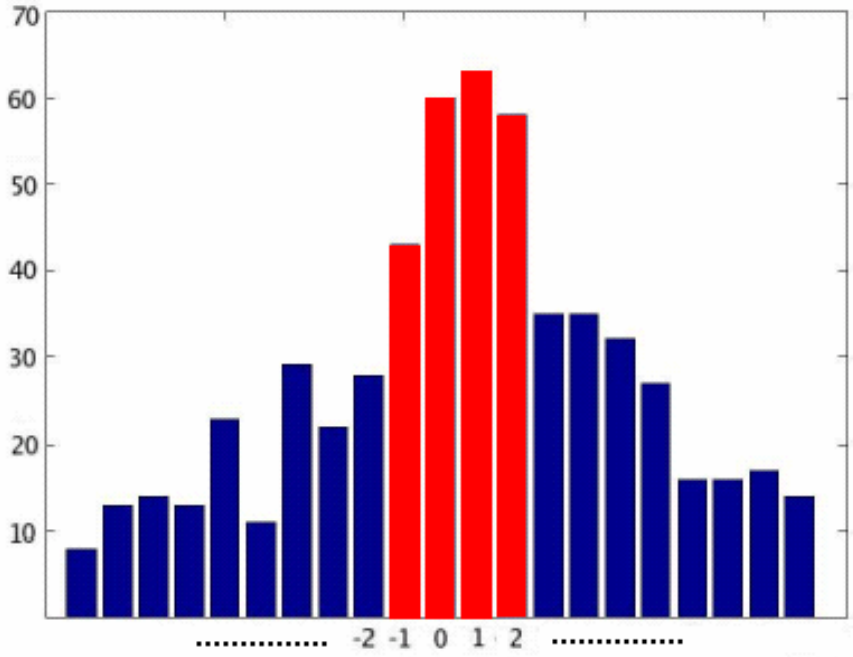
Step 1: Limit the search space of a particular person by using speech transcript text.

Step 2: Construct a similarity graph among faces in this search space.

Step 3: Find the densest component in this graph.

Step 1 : Integrating Names & Faces

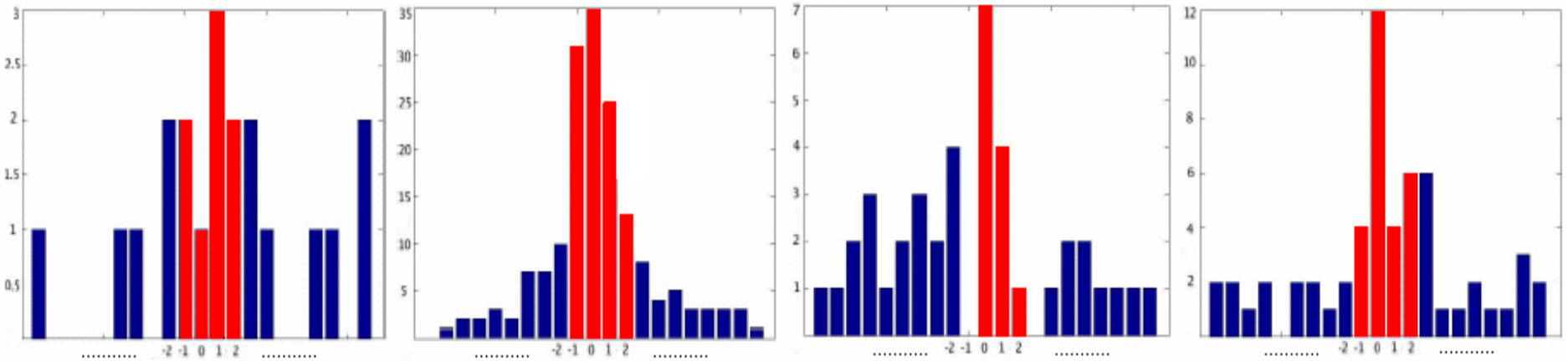




Frequency of Bill Clinton's visual appearance w.r.t the distance to the shot in which his name is mentioned.

Step 1 : Integrating Names & Faces

- The relative position of the faces to the name for Benjamin Netanyahu, Sam Donaldson, Saddam Hussein, and Boris Yeltsin respectively.



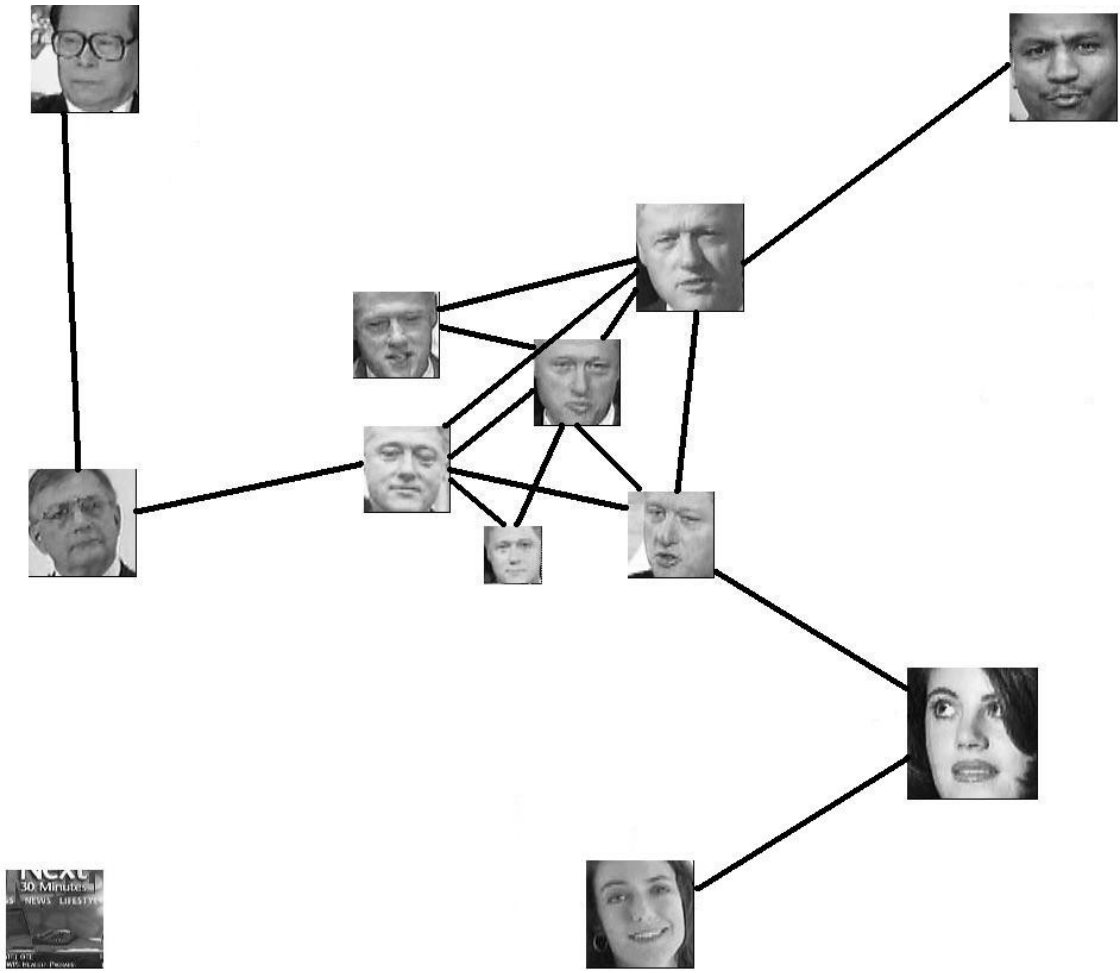
- It is seen that taking only one preceding and two following shots ([-1,2]) is also a good choice.

Proposed Solution

Step 1: Limit the search space of a particular person by using speech transcript text.

Step 2: Construct a similarity graph among faces in this search space.

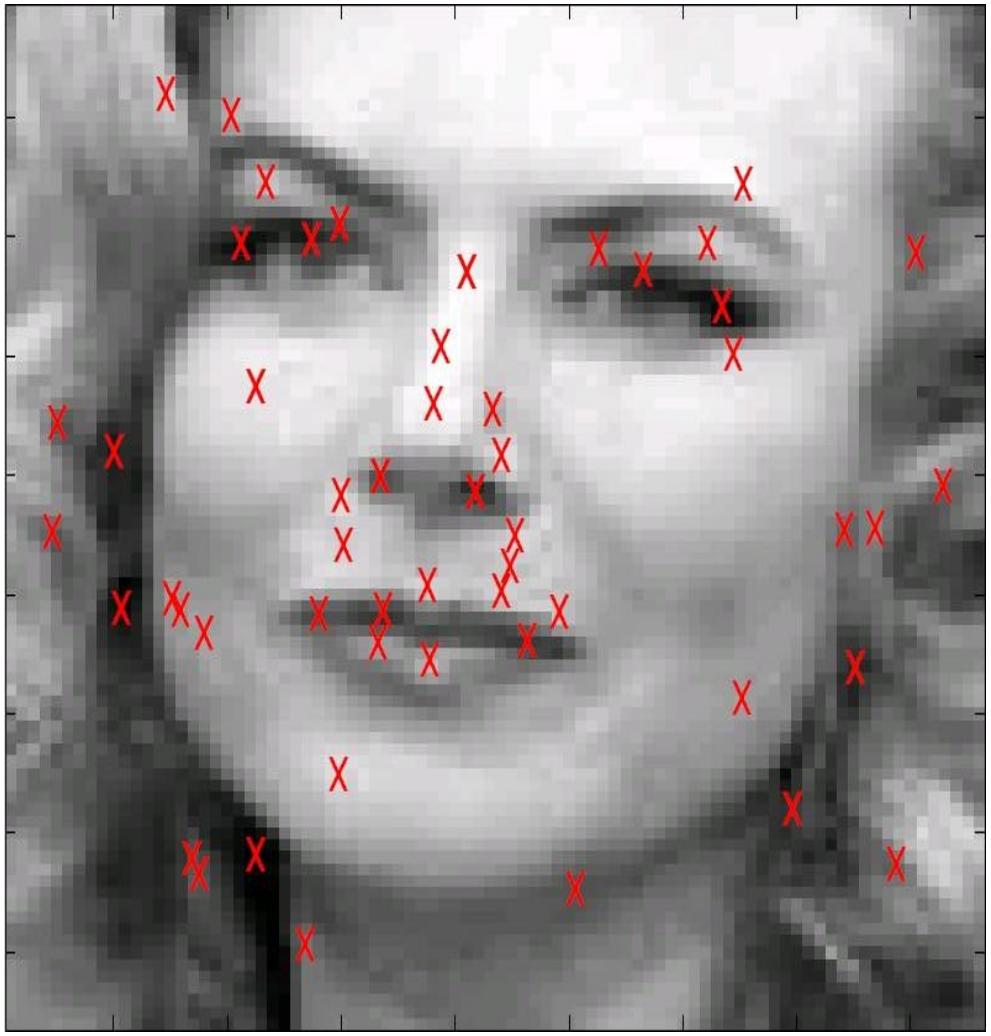
Step 3: Find the densest component in this graph.



Step 2 : Construct Similarity Graph

- The similarity of faces are defined using the interest points extracted from the detected face areas.

- Each face is defined by interest points extracted by using the Lowe's SIFT operator.



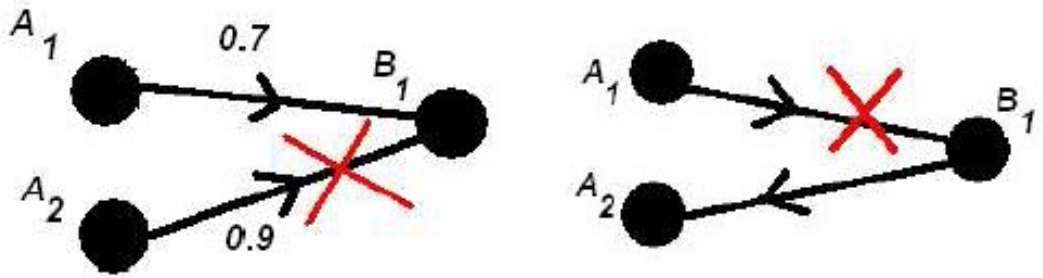
D. G. Lowe. Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision*, 60(2), 2004.

Similarity Between Two Faces

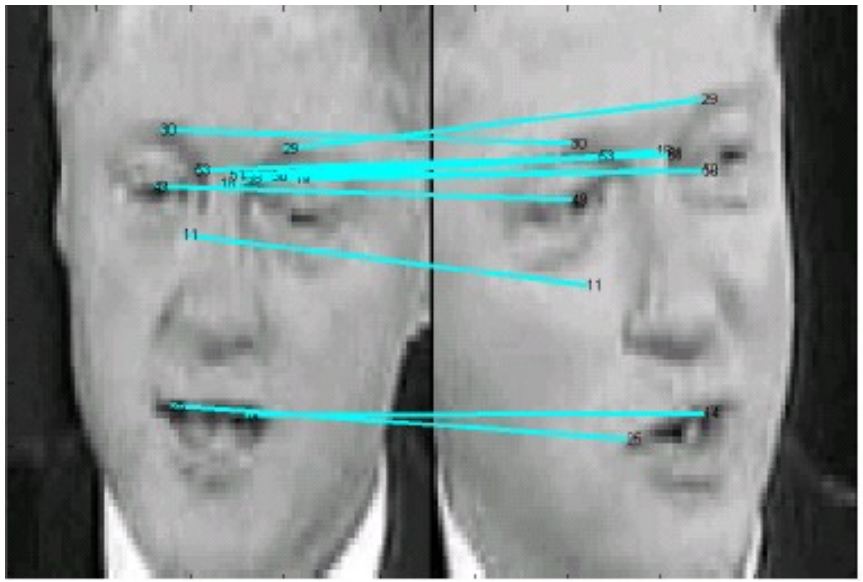
- While comparing two faces, each point on one face is compared with all the points on the other face and the points with the least Euclidean distance are selected as true matches.
- However, this approach returns many false matches along with the true ones.
- Hence, we apply two constraints on those matches:
  - geometrical constraint
  - the unique match constraint.

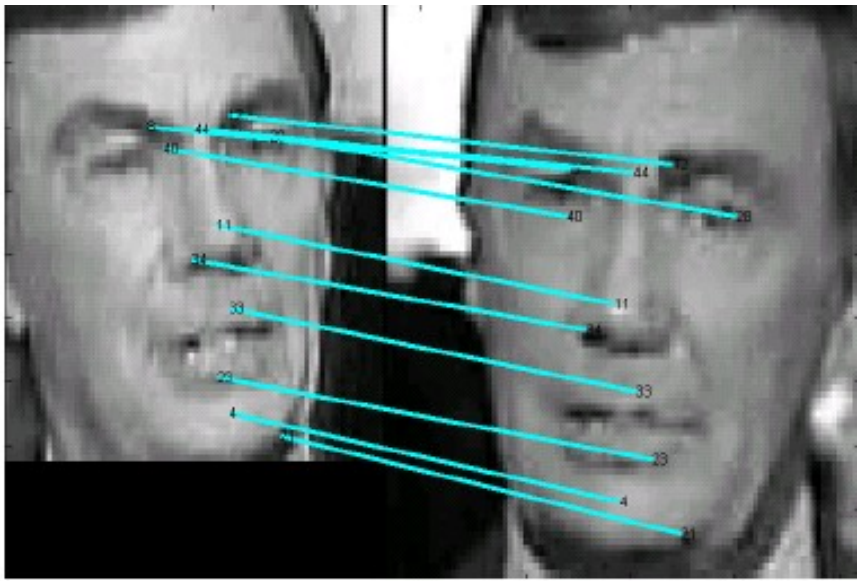
Constraints

- **Geometrical constraint** expects the matching points to appear around similar positions on the face when the normalized positions are considered.
- **Unique match constraint** ensures that each point matches to only a single point by eliminating multiple matches to one point and also by removing one-way matches.



Sample True Matches





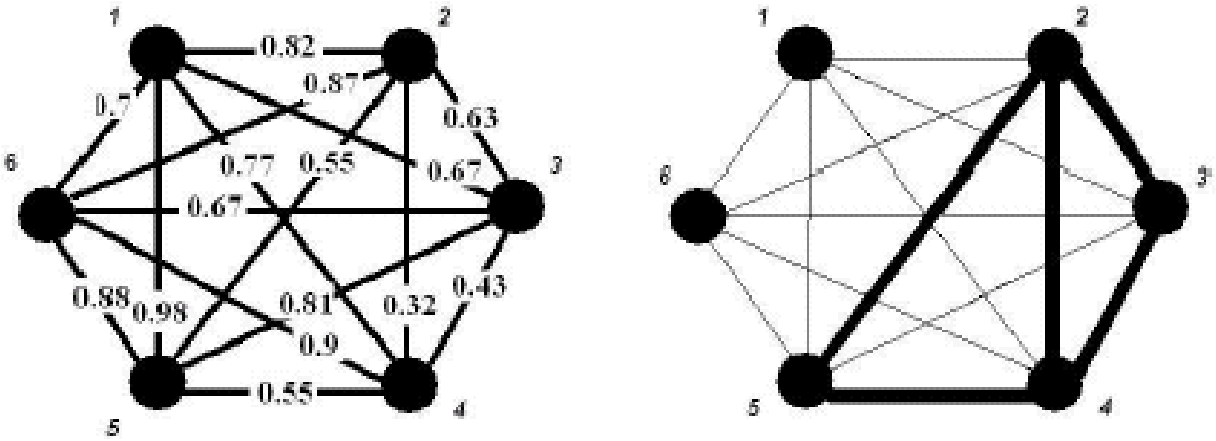
Dissimilarity Between Faces

- Distance between the two faces is defined as the average distance of all matching points between these two faces.

$$dist(A, B) = \frac{\sum_{i=1}^{N} D(i)}{N}$$

The Similarity Graph

- A dissimilarity graph for all the faces in the search space is then constructed using these distances, where nodes are faces and edge weights are distances.
- This graph is converted to a binary graph by using a threshold on the distance between any two nodes:

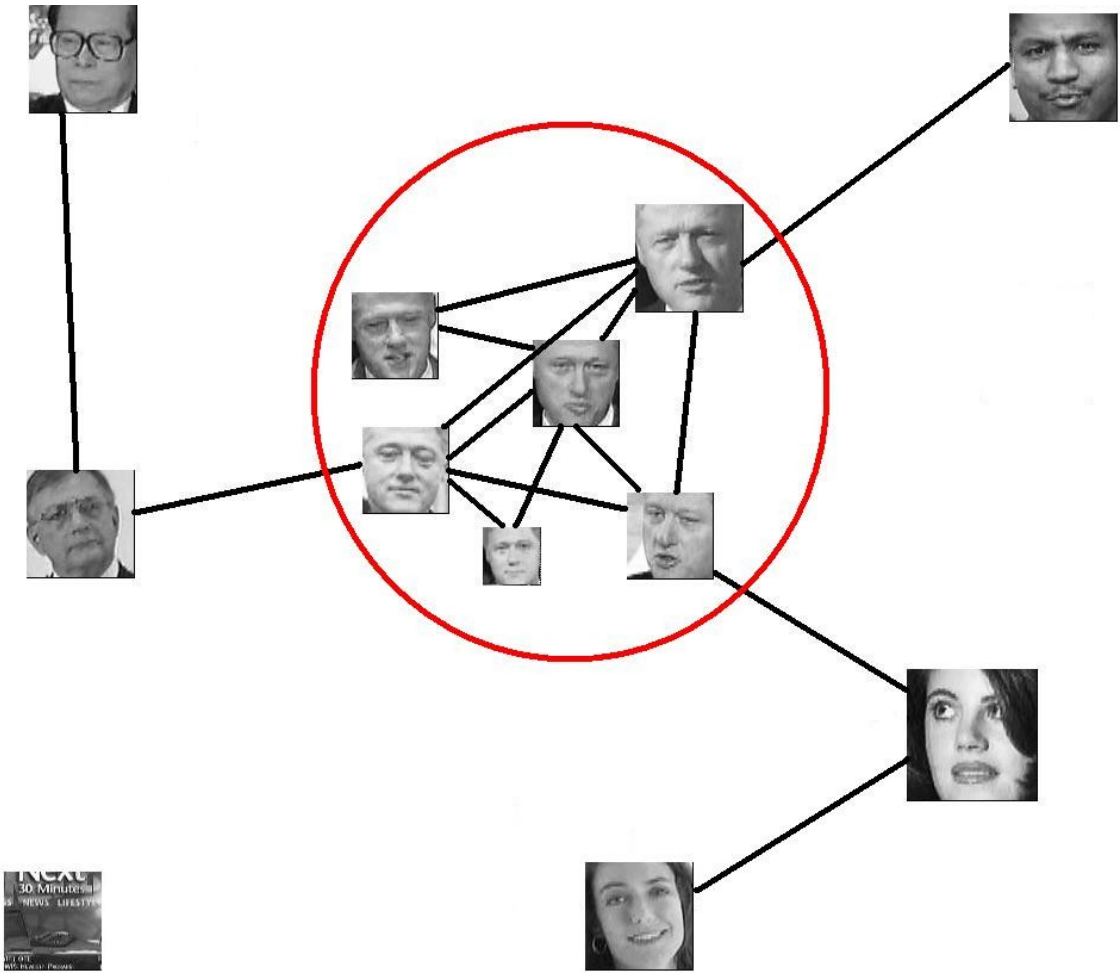


Proposed Solution

Step 1: Limit the search space of a particular person by using speech transcript text.

Step 2: Construct a similarity graph among faces in this search space.

Step 3: Find the densest component in this graph.



Step 3 : Finding Densest Component in the Graph

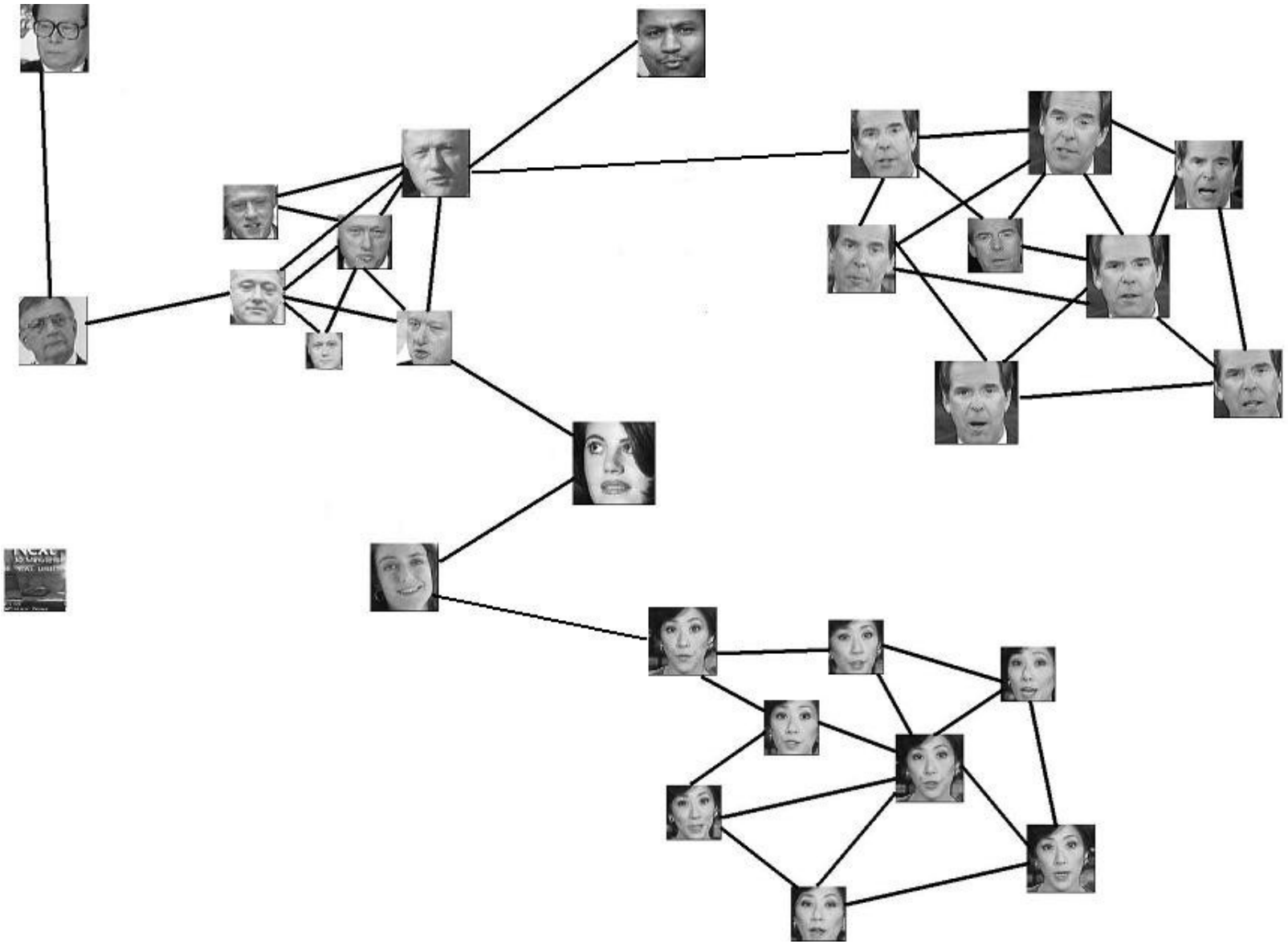
In [1], density of subset S of a graph G is defined as:

$$f(S) = \frac{|E(S)|}{|S|}$$

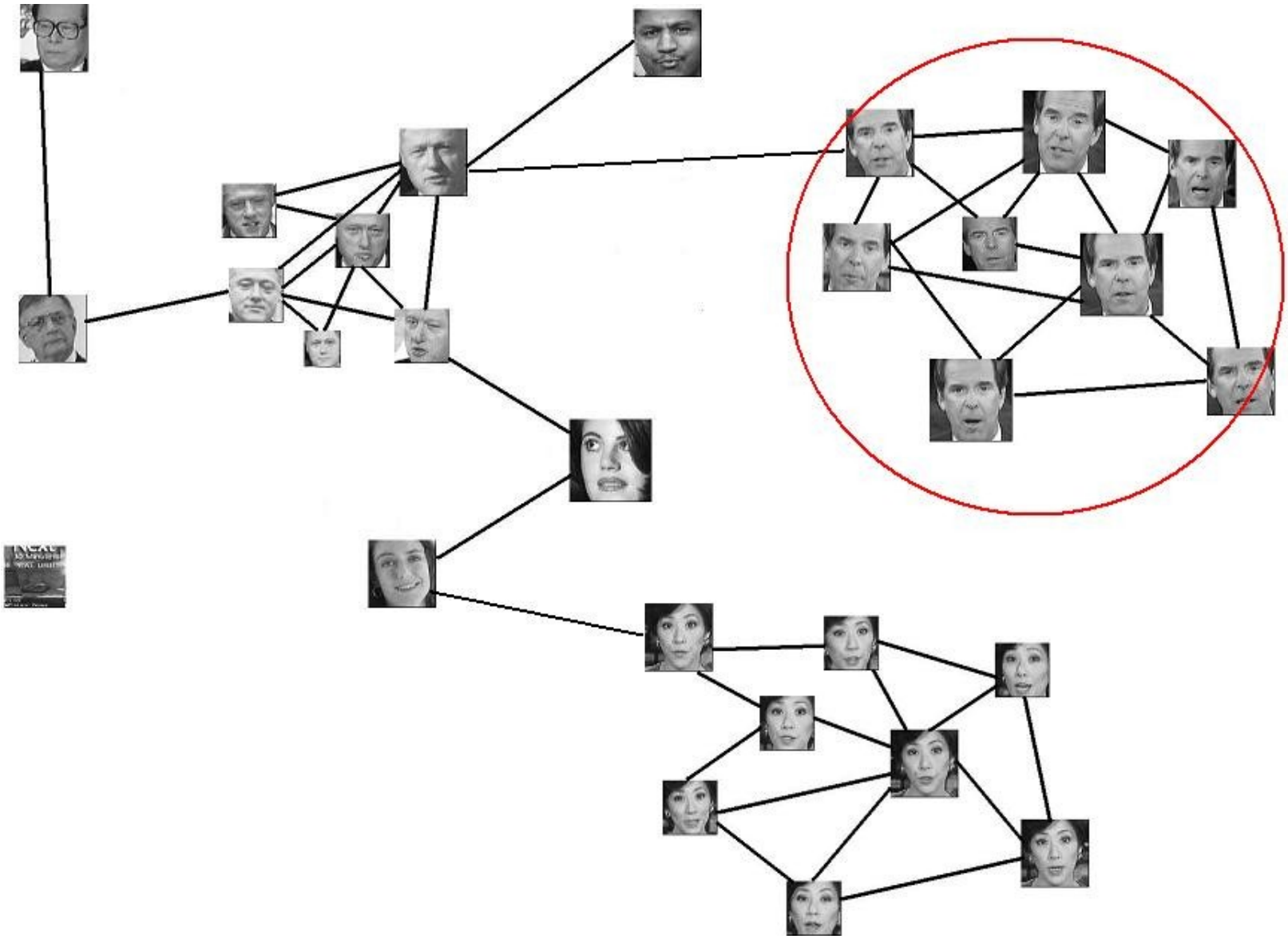
where E is the set of all edges in G and $E(S)$ is the set of edges induced by subset S.

[1] M. Charikar. 'Greedy approximation algorithms for finding dense components in a graph'. In APPROX '00: London, UK, 2000.

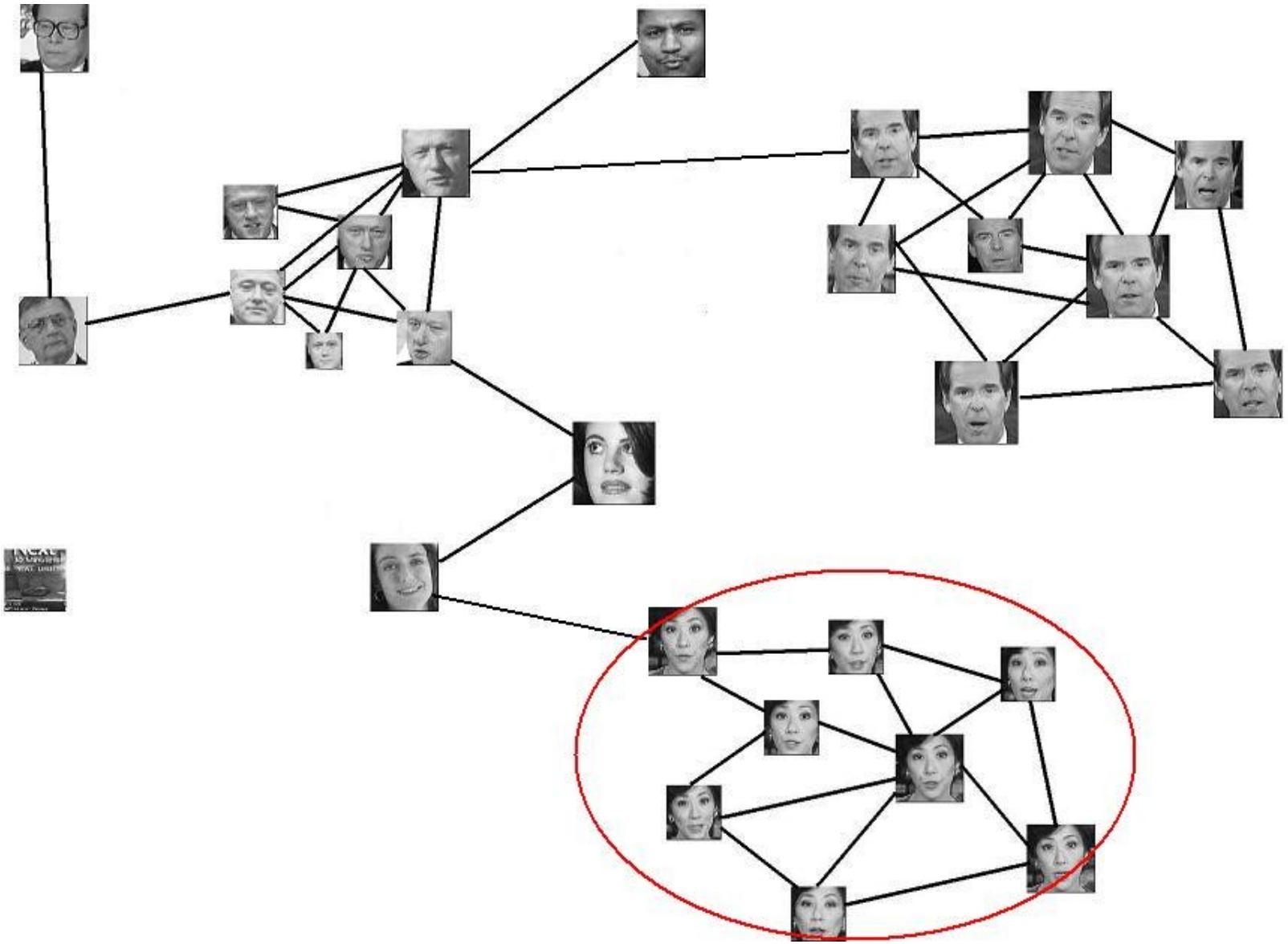
Problem: Anchor persons



Problem: Anchor persons

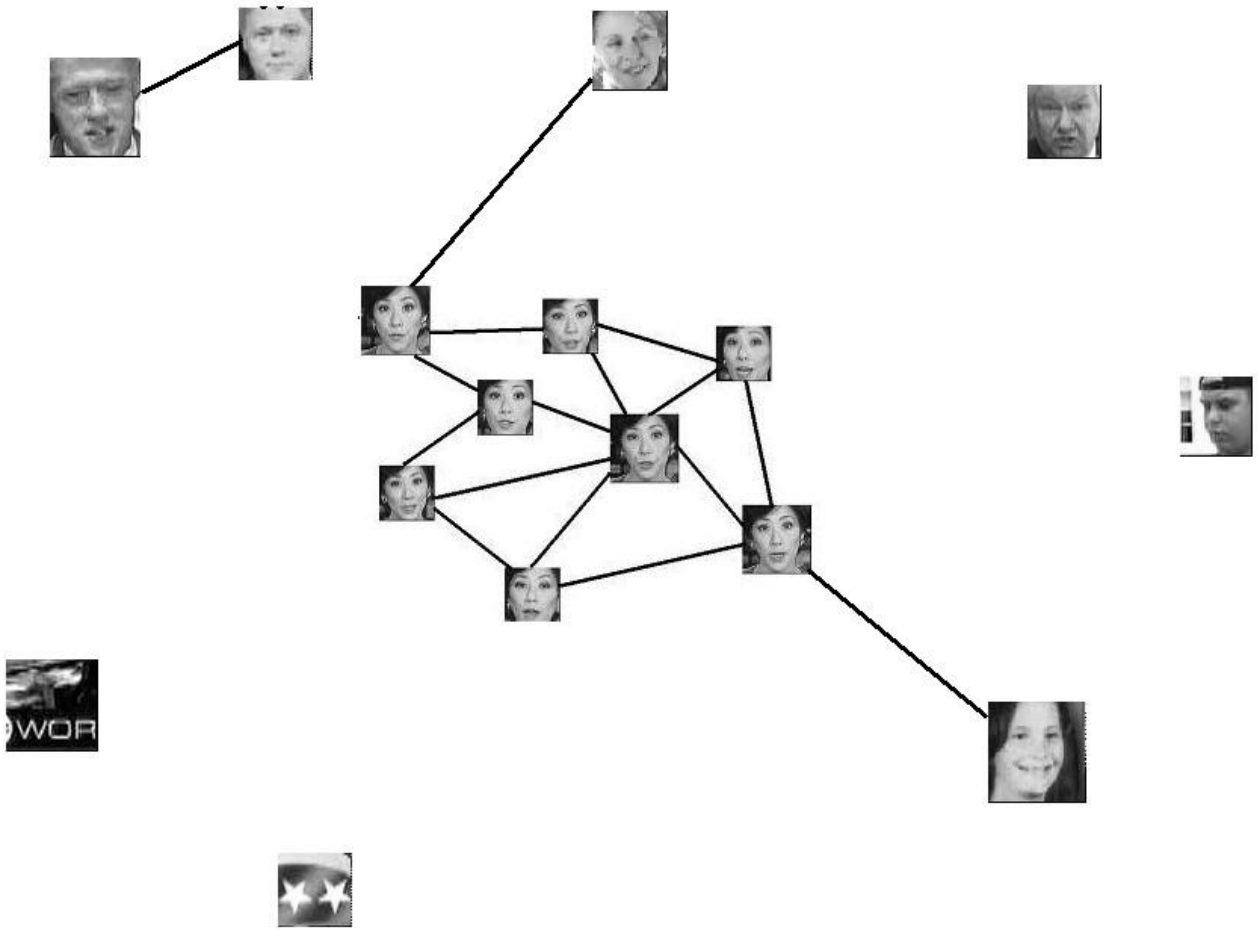


Problem: Anchor persons



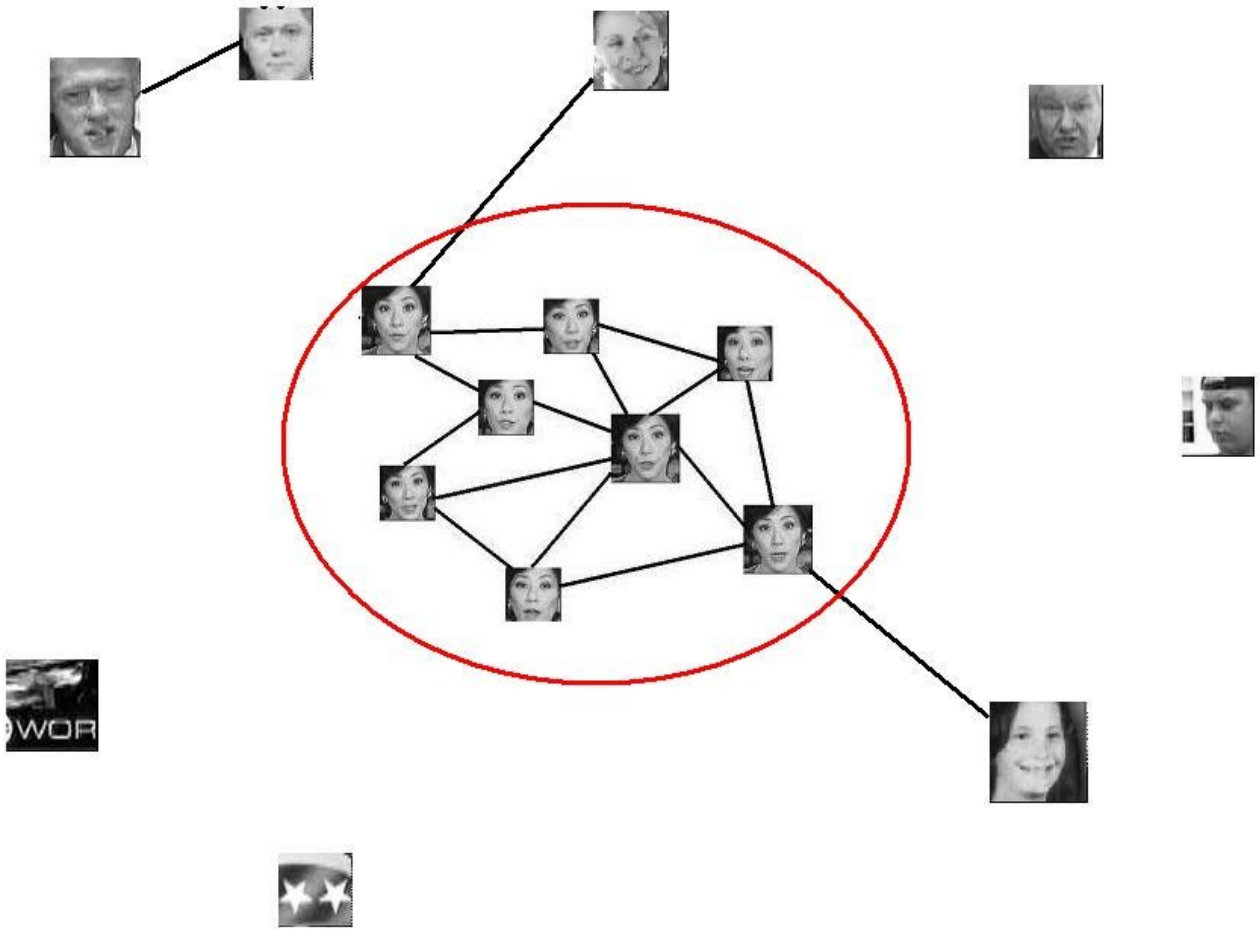
Solution: Anchor person detection

- Construct similarity matrix for each news video.



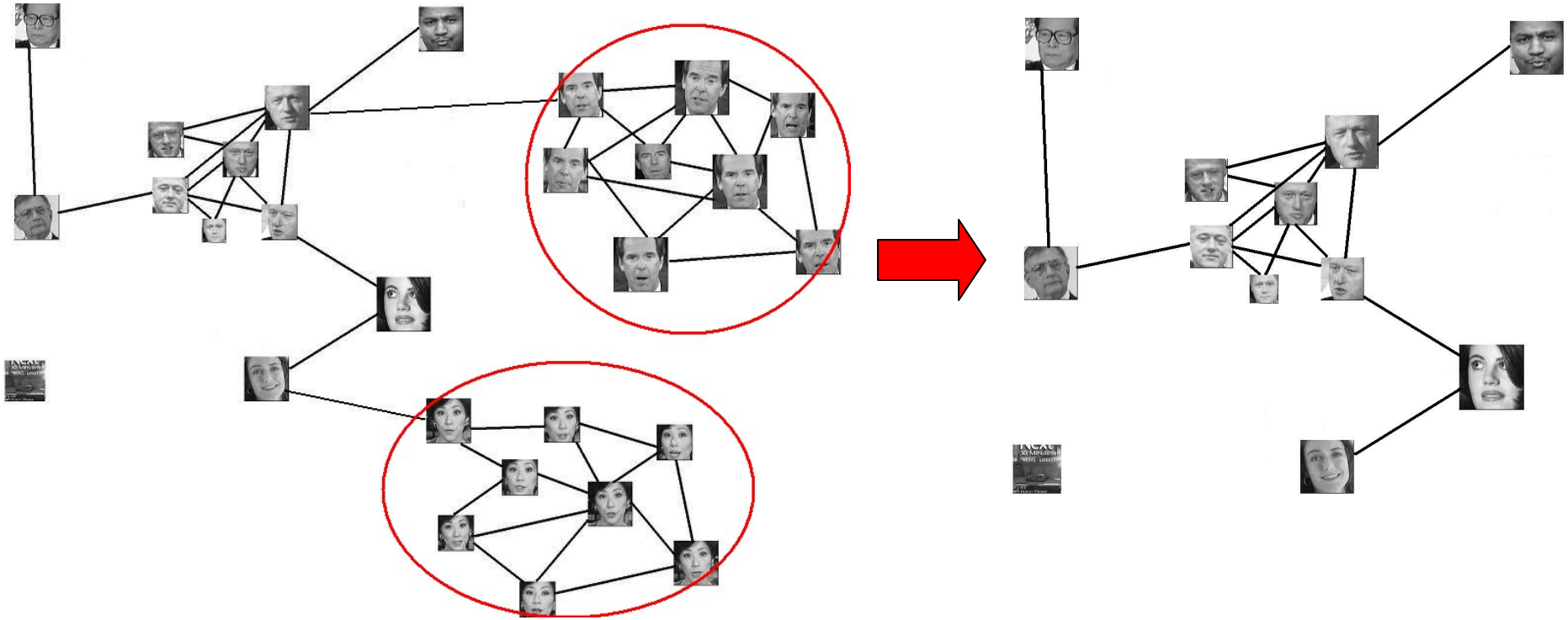
Solution: Anchor person detection

- Find the densest component which corresponds to the anchor person.



Solution: Anchor person detection

- Remove detected anchor persons of all videos from the search space of queried person.



Data Set

- Broadcast news videos provided by NIST for TRECVID video retrieval evaluation competition 2004.
- 229 movies (30 minutes each) from ABC and CNN news.
- The shot boundaries & key-frames are provided by NIST.
- Speech transcripts extracted by LIMSI are used to obtain the associated text for each shot.
- The face detection algorithm provided by Mikolajcyzk is used to extract faces from key-frames. 57% accuracy is recorded for randomly selected 10 videos.

Experiments

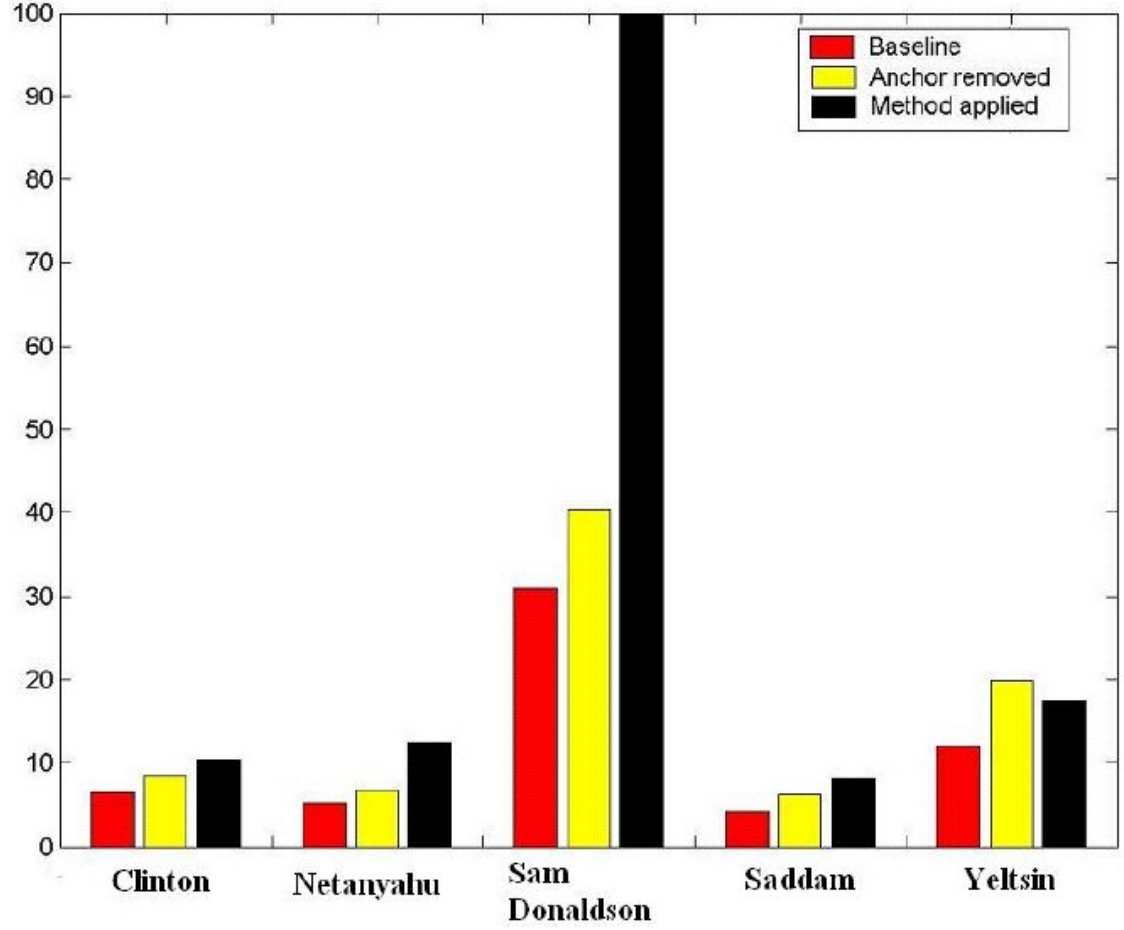
- We choose 5 people,
  - Bill Clinton,
  - Benjamin Netanyahu,
  - Sam Donaldson,
  - Saddam Hussein and
  - Boris Yeltsin.
- In the speech transcript text, their names appear 991, 51, 100, 149 and 78 times respectively.

Anchorperson Detection

- Average recall : 90%
- Average precision : 85%

Person queries

Query name	Clinton	Netanyahu	Sam Donaldson	Saddam	Yeltsin
Text-only	160/2457	6/114	102/330	14/332	19/157
Anchor removed	150/1765	5/74	81/200	14/227	17/122
Method applied	109/1047	4/32	67/67	9/110	10/57



Average precisions

- Baseline: 11.8 %
- After anchor persons removed: 15 %
- Overall: 29 %

29%

152%

Sample Images Retrieved

Conclusions

- We address the problem of finding a specific person in news videos.
- Applying the proposed person finding algorithm on each video separately, we detect the anchorperson in each video.
- Then, we remove detected anchorperson from the search space of the query name and apply the algorithm to the remaining images.

Conclusions

- Experiments are conducted on 229 broadcast news videos archive, which is a difficult set due to large variations in pose, illumination and expressions in data.
- 29% improvement in average precision is achieved after anchor person removal compared to only text based results, and 152% improvement after is achieved after applying the overall algorithm.
- The person finding algorithm also performs well for anchorperson detection without requiring any supervision.

Discussion

- Instead of taking a single face from each shot by only considering the key-frames, face detection can be applied to all frames in a shot to obtain more instances of the same person.

  This approach can help to find better matching interest points and more examples that can be used in the graph algorithm.

- Ozkan, D., Duygulu, P., "A Graph Based Approach for Naming Faces in News Photos". In Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition. (CVPR 2006)

THANK YOU FOR LISTENING !!!