Names and Faces in the News

Tamara L. Berg, Alexander C. Berg, Jaety Edwards, Michael Maire, Ryan White, Yee-Whye Teh, Erik Learned-Miller and D.A. Forsyth

Yong Jae Lee

March 8th, 2007

Introduction

- Goal: Given an input image and an associated caption, detect the face(s) in the image and label it with the correct name(s) detected in the caption
- Motivation: Build a rich, reasonably accurate collection of labeled faces

General Approach

Detect Names in Captions → Detect + Represent Faces → Associate Names with Faces

President George W. Bush makes a statement in the Rose Garden while Secretary of Defense Donald Rumsfeld looks on, July 23, 2003. Rumsfeld said the United States would release graphic photographs of the dead sons of Saddam Hussein to prove they were killed by American troops. Photo by Larry Downing/Reuters

World number one Lleyton Hewitt of Australia hits a return to Nicolas Mason of Chile at the Japan Open tennis championships in Tokyo October 3, 2002. REUTERS/Eriko Sugita

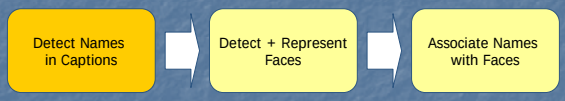
British director Sam Mendes and his partner actress Kate Winslet arrive at the London premiere of 'The Road to Perdition', September 18, 2002. The films stars Tom Hanks as a Chicago hit man who has a separate family life and co-stars Paul Newman and Jude Law. REUTERS/Dan Chung

German supermodel Claudia Schiffer gave birth to a baby boy by Caesarian section January 30, 2003, her spokeswoman said. The baby is the first child for both Schiffer, 32, and her husband, British film producer Matthew Vaughn, who was at her side for the birth. Schiffer is seen on the German television show "Bet It...?!" ("Wetten Dass...?!") in Braunschweig, on January 26, 2002. (Alexandra Winkler/Reuters)

Dataset

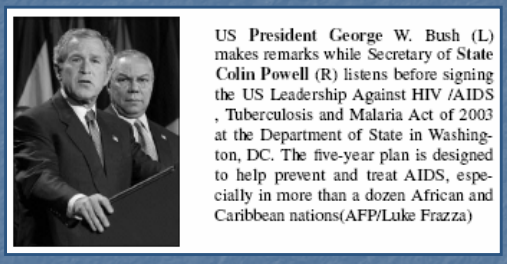
- 0.5 million news pictures with captions
- Various pose, illumination, expression
- Glasses, wigs, mustaches

Names



Names

- Extract names found in the captions
- Identify two or more capitalized words followed by a present tense verb
- Associate every face in an image to every name extracted



US President George W. Bush (L) makes remarks while Secretary of State Colin Powell (R) listens before signing the US Leadership Against HIV /AIDS , Tuberculosis and Malaria Act of 2003 at the Department of State in Washington, DC. The five-year plan is designed to help prevent and treat AIDS, especially in more than a dozen African and Caribbean nations(AFP/Luke Frazza)

Faces

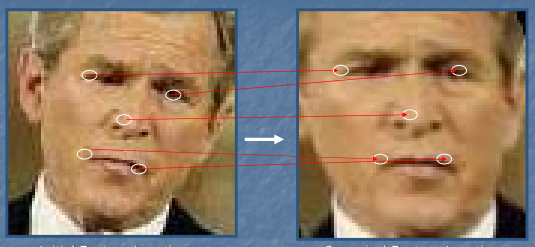
Faces

- "Face Detector" - K. Mikolajczyk
- 44,773 face images of size 86x86
- Biased to frontal views

Rectification of Face Images

- Reduce within-class variance
- 5 SVM
- Geometric blur applied to grayscale patches

- Affine transformation: least squares solution
- Gradient Descent



Initial Feature Locations | Canonical Feature Locations

Full Plane Affine Transformation

- Images with poor rectification scores are removed from the dataset
- 44,773 → 34,623 face images
- Cropped to a region surrounding the eyes, nose and mouth
- RGB pixel values → Vector

Face Representation

- Vectors in a space where same faces are close and different faces are far apart
- Discard components that are similar for all faces
  → Kernel PCA
- Project data into space for discrimination
  → LDA

Kernel PCA

- Compute a kernel matrix, K
- K_{ij} = value of kernel function comparing $image_i$ and $image_j$
- But NxN Kernel Matrix… ~2*10^9 image comparisons

Nystrom Approximation

- Approximation to calculate the eigenvectors of K

$$K = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \qquad A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{(N-n) \times n} \text{ and } C \in \mathbb{R}^{(N-n) \times (N-n)}$$

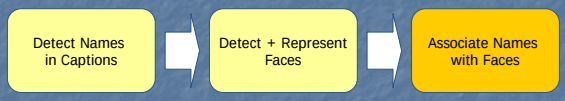
- A: subset of images compared to themselves (1000 random images)
- B: comparison A to rest of images in dataset
- C: Approximated as $\hat{C} = B^T A^{-1} B$

→ $\hat{K} = \begin{bmatrix} A & B \\ B^T & \hat{C} \end{bmatrix}$

LDA

- Form an initial discriminant space – with single face detected in image with single name in caption
- Project all images into this space to discriminate different faces

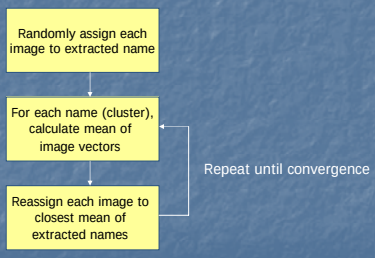
Names and Faces



Faces and Names

- Each image now represented by vector and set of extracted names
- Modified K-means clustering

Modified K-means Clustering



Prune clusters

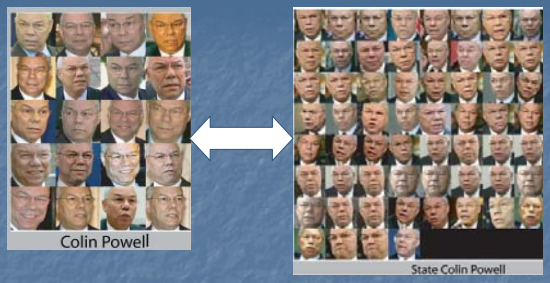
- Nearest neighbor model
- Remove clusters with fewer than three images
- Remove points with low likelihood to get low error rates
- Likelihood

$$= \frac{P(\text{face is from assigned cluster})}{P(\text{face is not from assigned cluster})}$$

#Images	#Clusters	error rate
19355	2357	26%
7901	1510	11%
4545	765	5.2%
3920	725	7.5%
2417	328	6.6%

Merge Clusters

- Different names that correspond to same person
- Similarity of Clusters: distance of their means



Merge into one cluster

Conclusion

- Fairly good assignment of names to faces using simple models for images, names
- Weaknesses:

  - use of RGB pixel values to discriminate between faces of different people

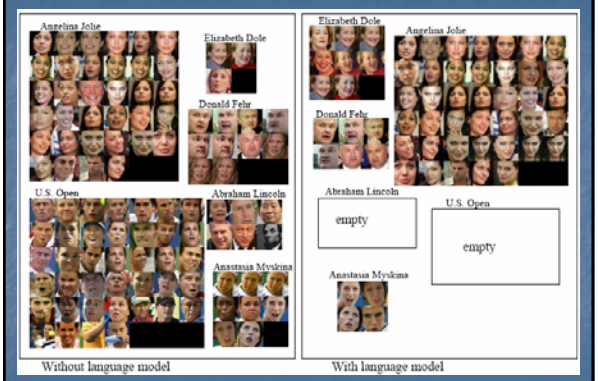
  - random assignment of faces (to one of its extracted names) in clustering process

  - no "NULL" assignment

Follow-up

- "Who's in the Picture?" – language model based on context
- Assign a probability to each name based on context
- Uses cues such as (L), (R), (C), location of the name in the caption, etc.

"The Right Stuff" cast members **IN Pamela Reed IN**, (L) poses with fellow cast member **IN Veronica Cartwright IN** at the 20th anniversary of the film in Hollywood, June 9, 2003. The women played wives of astronauts in the film about early United States test pilots and the space program. The film directed by **OUT Philip Kaufman OUT**, is celebrating its 20th anniversary and is being released on DVD. REUTERS/Fred Prouser



References

- Tamara L. Berg, Alexander C. Berg, Jaety Edwards, Michael Maire, Ryan White, Yee-Whye Teh, Erik Learned-Miller and D.A. Forsyth, “Names and Faces in the News”, Computer Vision and Pattern Recognition (CVPR), 2004
- Tamara L. Berg, Alexander C. Berg, Jaety Edwards, D.A. Forsyth, “Who’s in the Picture?”, Neural Information Processing Systems (NIPS), 2004
- Alexander C. Berg, Jitendra Malik, “Geometric Blur for Template Matching”, Computer Vision and Pattern Recognition (CVPR) 2001, Hawaii, pp I.607-614