

# Estimating missing features to improve multimedia information retrieval

Abraham Bagherjeiran   Nicole S Love  
Chandrika Kamath

*Lawrence Livermore National Laboratory*

UCRL-PRES-225087

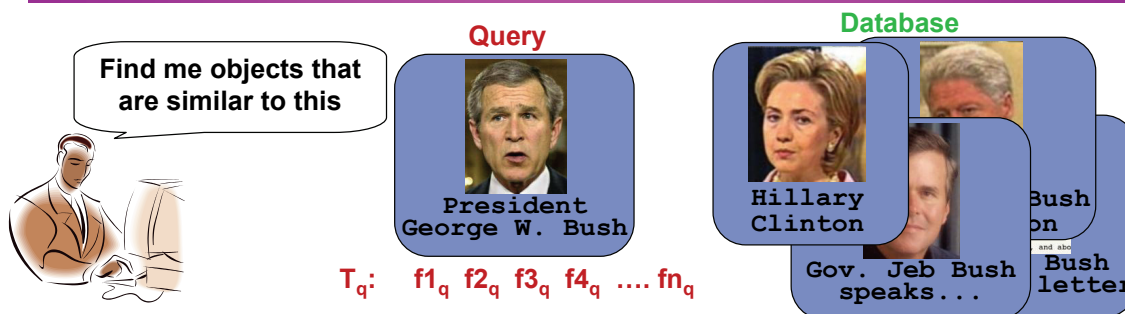
Work performed under the auspices of the U. S. Department of Energy by  
University of California Lawrence Livermore National Laboratory under  
Contract W-7405-Eng-48



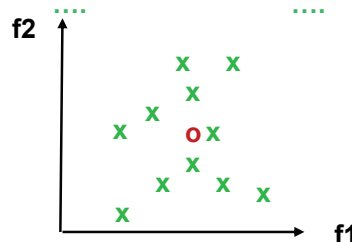
<http://www.llnl.gov/CASC/sapphire/>



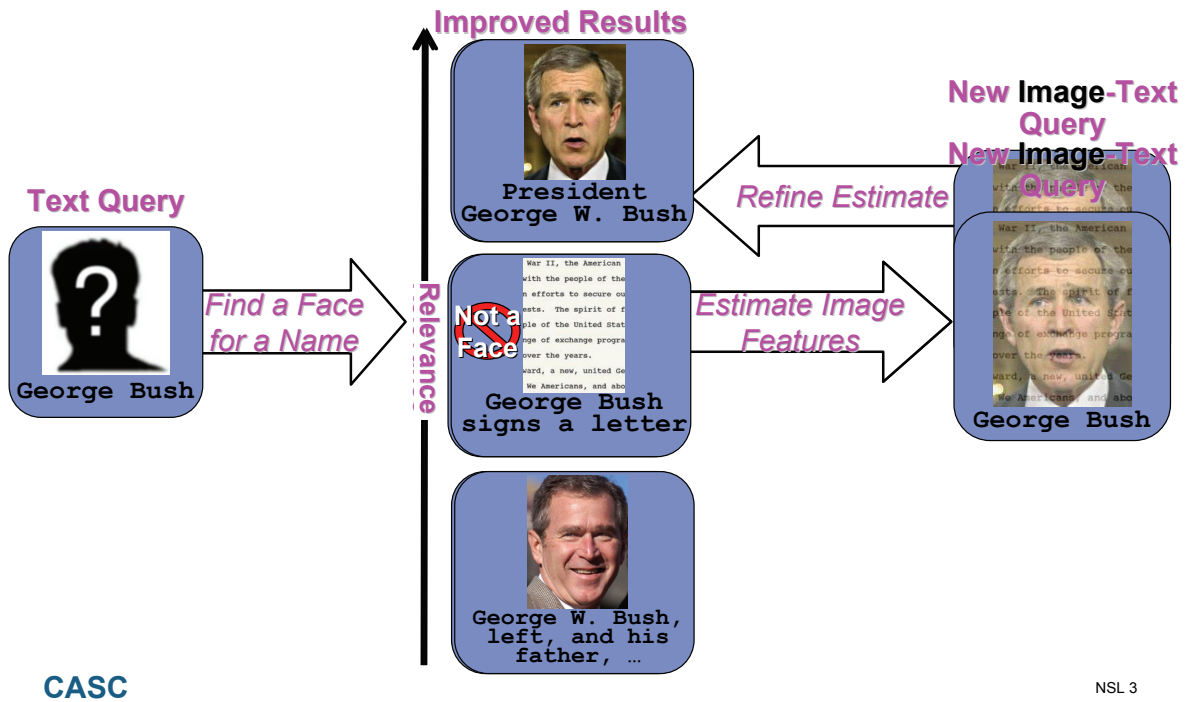
## Multimedia Information Retrieval



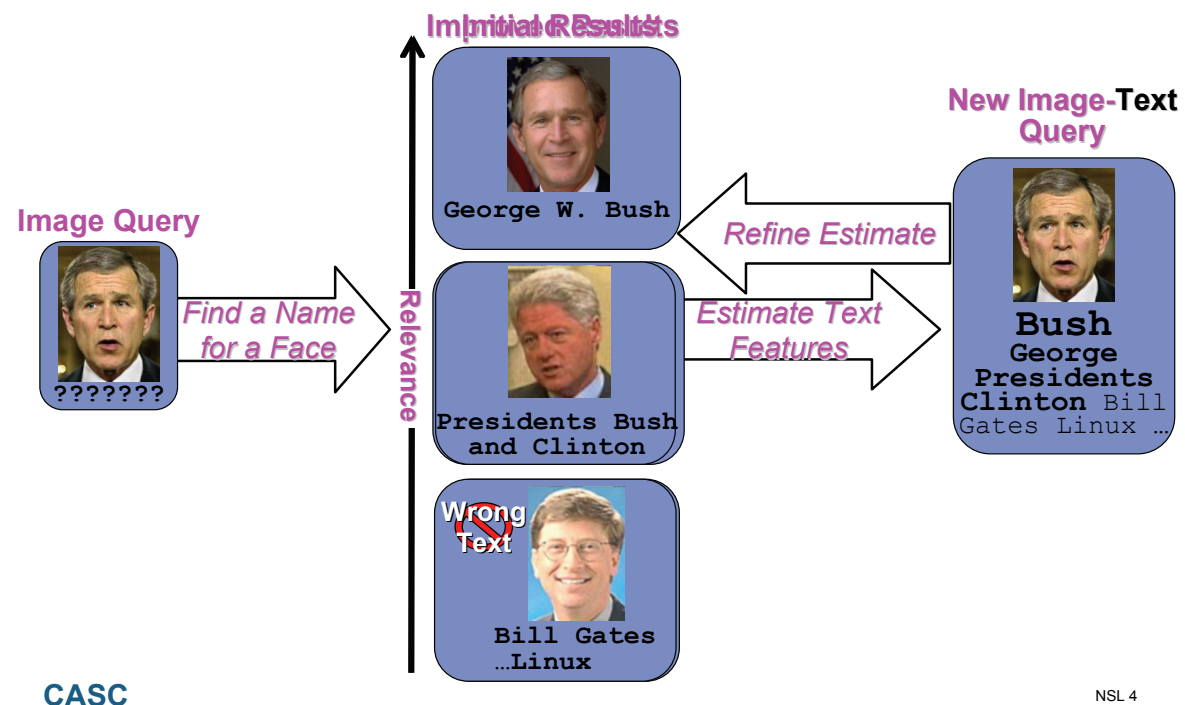
- Extract “faces” and “captions”
- Represent using features
- Similarity search



# Text-only query with image features estimated



# Image-only query with text estimated



# Image and Text Features

- Image Features (187 total)
  - Angular radial transform (71)
  - Normalized histogram (16)
  - Gabor features (60)
  - Gray level co-occurrence matrices (20)
  - Power spectrum (20)
- Text Features
  - TFIDF (text frequency-inverse document frequency)

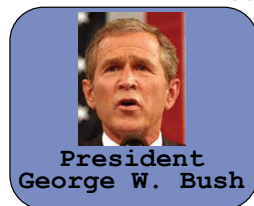
# Completing a partial query: the simple method

## Image Query



$$q = \langle a_q, ? \rangle$$

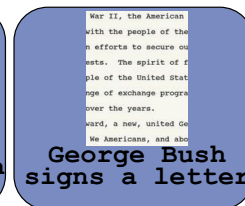
## Retrieved documents



$$r_1 = \langle a_1, m_1 \rangle$$



$$r_i = \langle a_i, m_i \rangle$$



$$r_n = \langle a_n, m_n \rangle$$

$$\hat{m}_q = \frac{\sum_j \gamma_j m_j}{\sum_j \gamma_j}$$

$$\gamma_j = \delta_j e^{-\alpha d(r_j, q)}$$

$$\delta_j = \begin{cases} 1 + \beta & j \leq k^* \\ 1 & j > k^* \end{cases}$$

$$d(r, q) = \begin{cases} d(\langle a_r, \langle a_q \rangle) & \text{if } q = \langle a_q, ? \rangle \\ d(\langle a_r, m_r \rangle, \langle a_q, m_q \rangle) & \text{otherwise} \end{cases}$$

## New Image-Text Query



$$\hat{q} = \langle a_q, \hat{m}_q \rangle$$

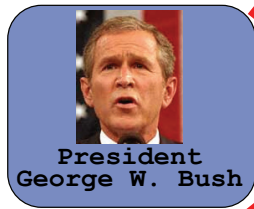
# Completing a partial query: with relevance feedback

## Image Query



$$q = \langle a_q, ? \rangle$$

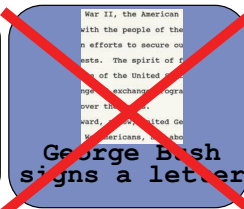
## Retrieved documents



$$r_1 = \langle a_1, m_1 \rangle$$



$$r_i = \langle a_i, m_i \rangle$$



$$r_n = \langle a_n, m_n \rangle$$

$$\hat{m}_q = \frac{\sum_j \gamma_j m_j}{\sum_j \gamma_j}$$

$$\gamma_j = \delta_j e^{-\alpha d(r_j, q)}$$

$$\delta_j = \begin{cases} 1 & j \leq k^* \\ 0 & j > k^* \end{cases}$$

$$d(r, q) = \begin{cases} d(\langle a_r, \rangle, \langle a_q, \rangle) & \text{if } q = \langle a_q, ? \rangle \\ d(\langle a_r, m_r \rangle, \langle a_q, m_q \rangle) & \text{otherwise} \end{cases}$$

## New Image-Text Query



$$\hat{q} = \langle a_q, \hat{m}_q \rangle$$

CASC

NSL 11

## Data Set

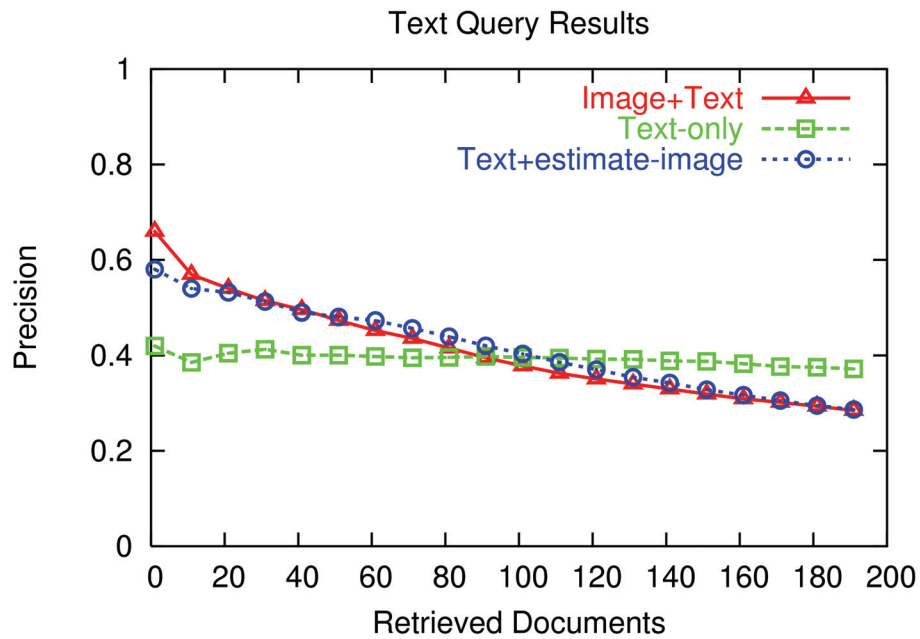
- 5,910 candidate faces and associated captions
  - Krystian Mikolajczyk's face detector
  - faces in the same image have the same caption
  - 619 text features
  - 187 low-level image features
  - 10 known faces (50 queries, 5 per person)
- Collected data set using Google Images Agent\*
- Removed words with document frequency less than 3
  - reduced from 37,000 words to 619 words
- Removed classes with less than 100 documents

\* G. Rouse <http://search.cpan.org/~grouse/>

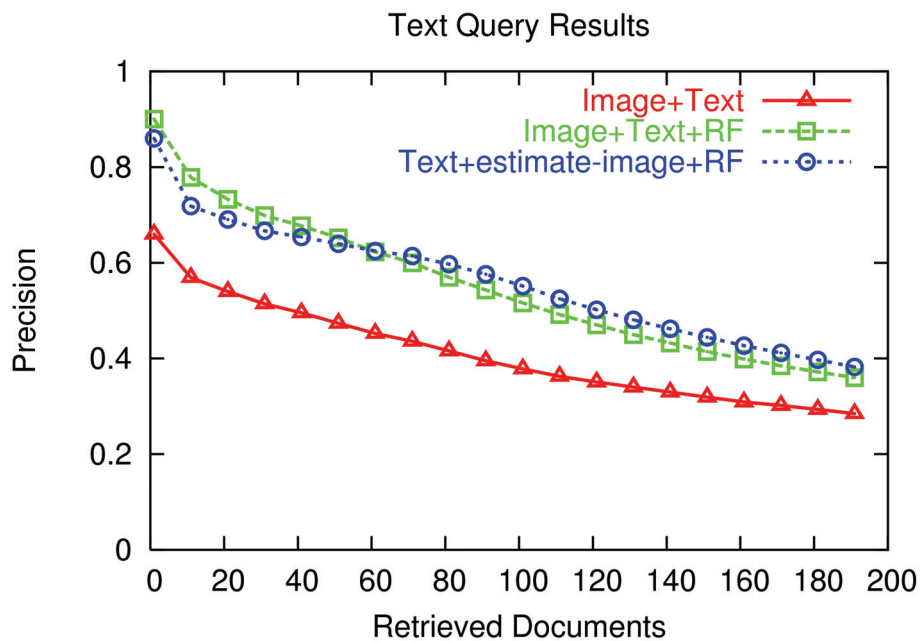
CASC

NSL 12

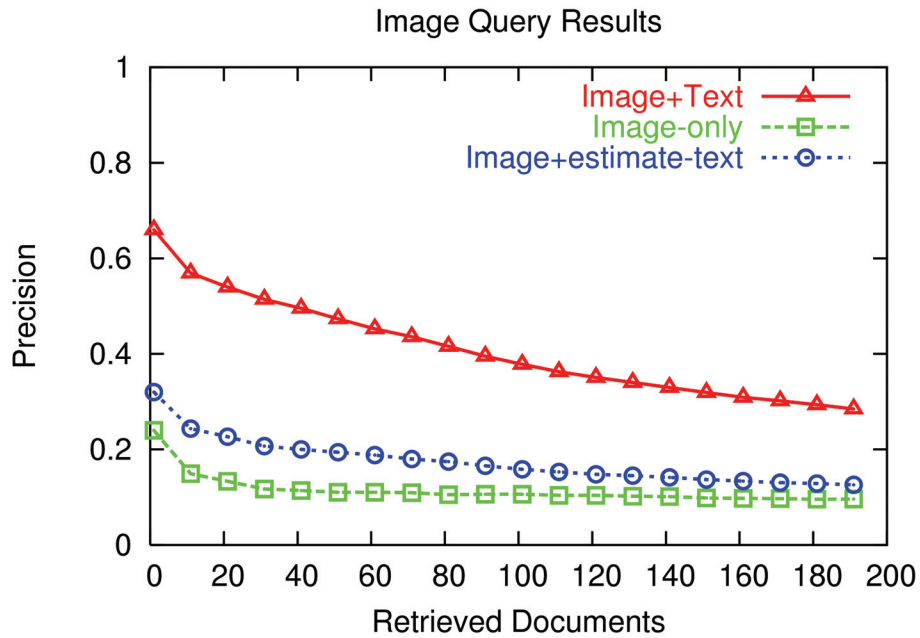
## Text-only, completed with simple method matches full query



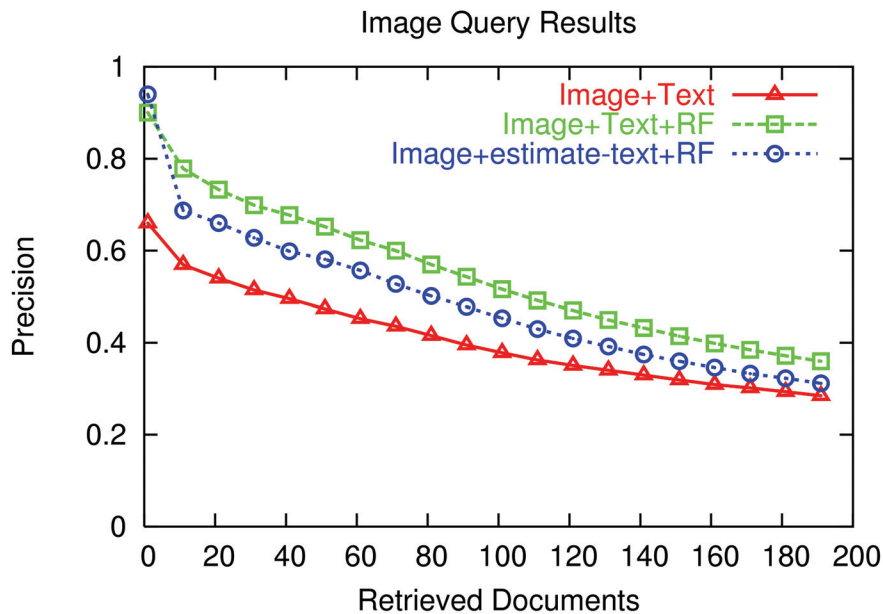
## Text-only query, completed with RF matches full query with RF



## Image-only, completed with simple method improved over image-only



## Image-only query, completed with RF approaches full query with RF



## Conclusions

---

---

- **Text features**
  - Can distinguish individuals
  - Can not distinguish faces from non-faces
- **Image features**
  - Can distinguish faces from non-faces
  - Can not distinguish individuals
- **Query completion showed improvement over partial queries (image-only and text-only)**
- **Text only with estimated image features matches full query**
- **Query completion with relevance feedback outperforms full query**

## Acknowledgments

---

---

- **Dale Slone (LLNL) – script to grab images and captions**
- **Krystian Mikolajczyk – face detector software**
- **Sapphire Team – developed software used**