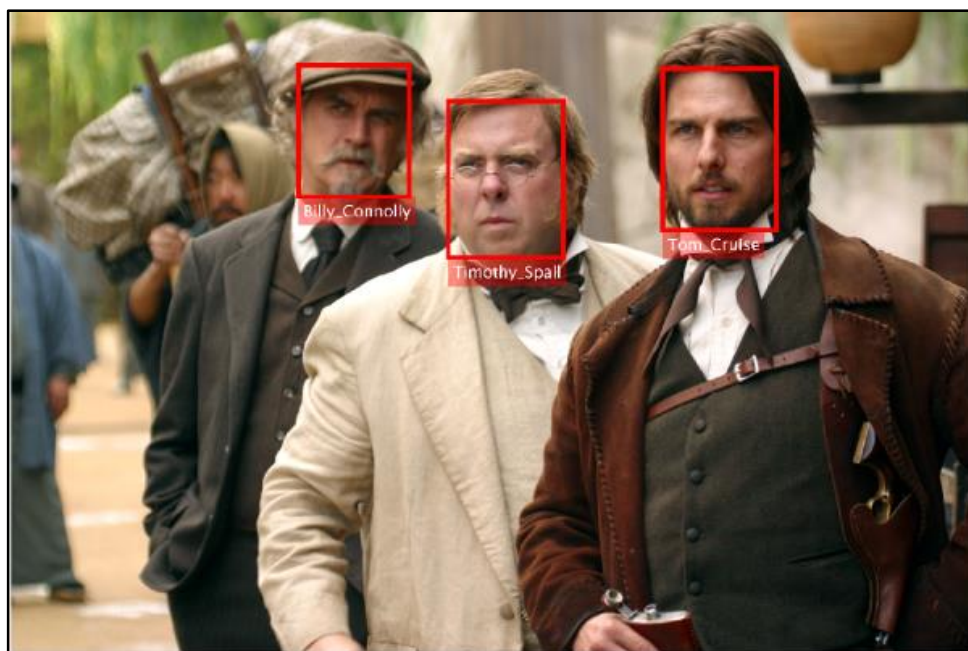


# Automated Tagging of Image and Video Collections using Face Recognition



Andrew Brown

Andrew Zisserman

Ernesto Coto

# The Goal

- Provide TAGs or retrieve frames in videos using face recognition
- Two example datasets

The BFI Untagged dataset



Tags:            ?

Film: "Slumdog Millionaire"

The BBC News dataset



Tags:            ?

Programme: "Newsnight"

# Visual Challenges for Face Recognition

- Performing this task using **Face Recognition** can be challenging

**Age**



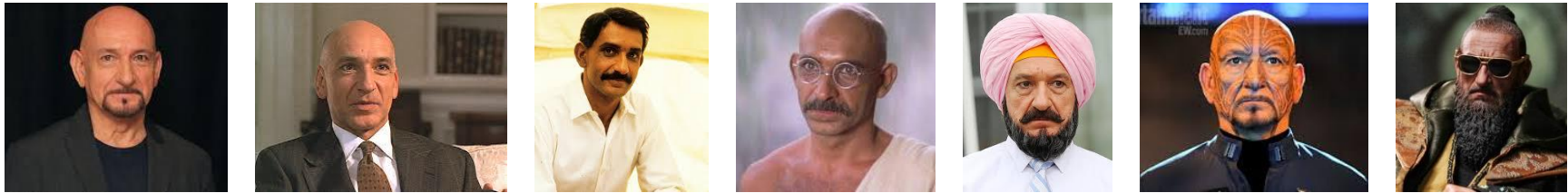
# Visual Challenges for Face Recognition

- Performing this task using **Face Recognition** can be challenging

**Age**



**Appearance**



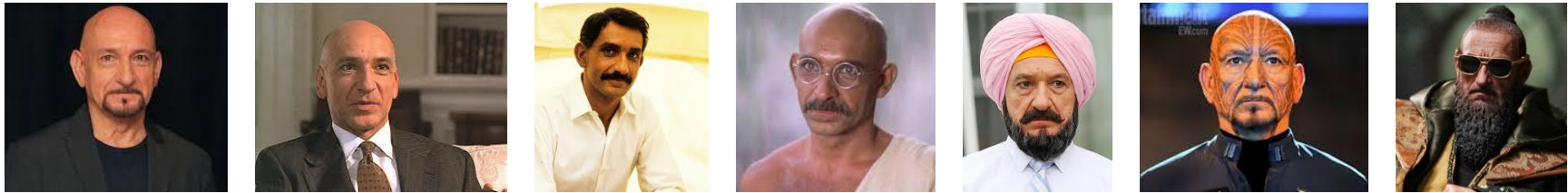
# Visual Challenges for Face Recognition

- Performing this task using **Face Recognition** can be challenging

**Age**



**Appearance**



**Expression**



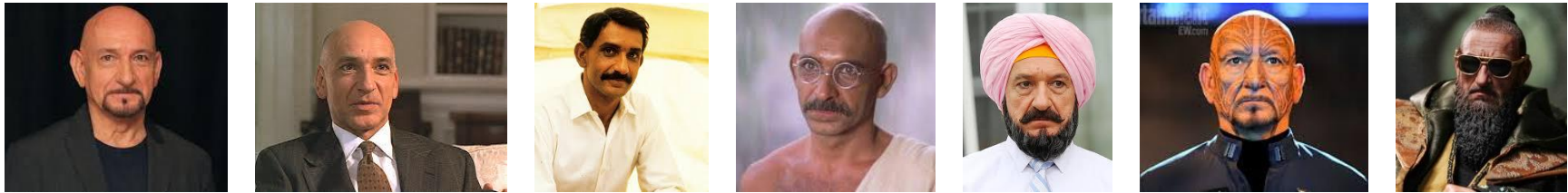
# Visual Challenges for Face Recognition

- Performing this task using **Face Recognition** can be challenging

**Age**



**Appearance**



**Expression**

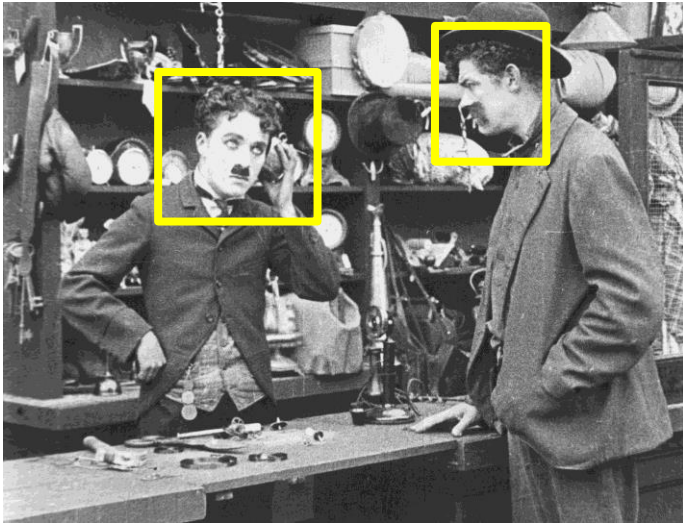


**Viewpoint**



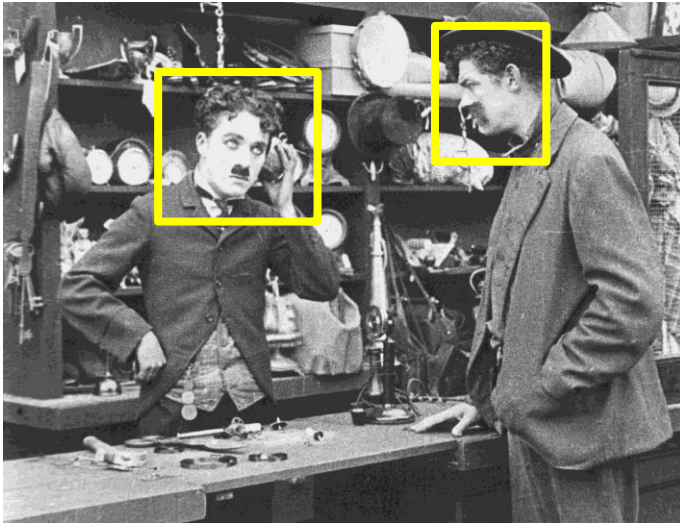
# The Approach

## 1. Detect Faces

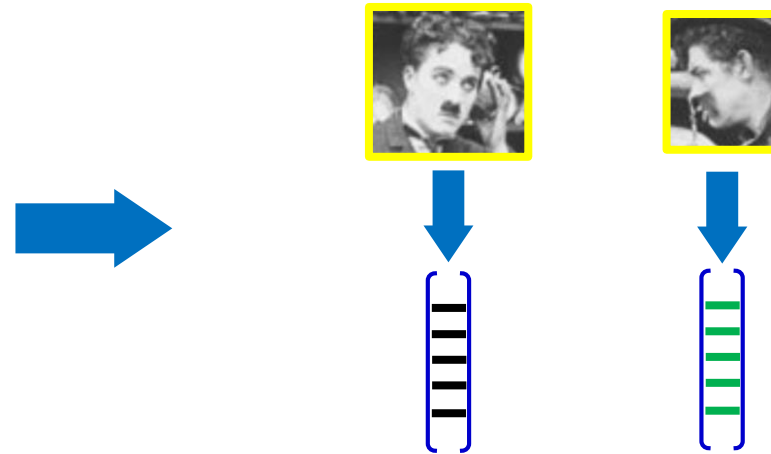


# The Approach

## 1. Detect Faces



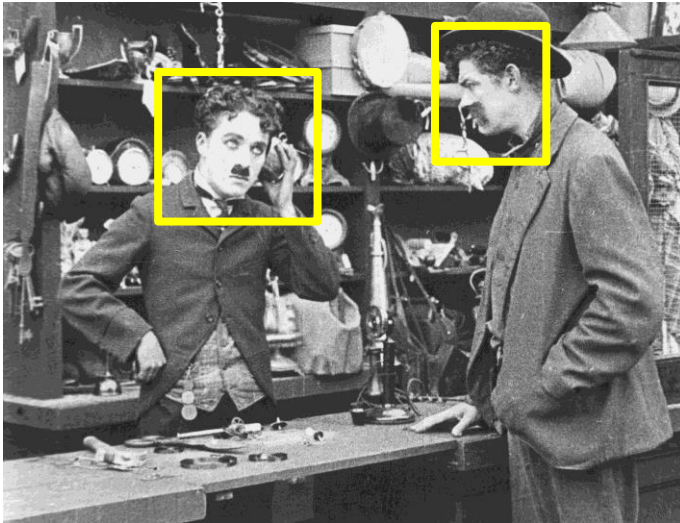
## 2. Represent each face by a vector



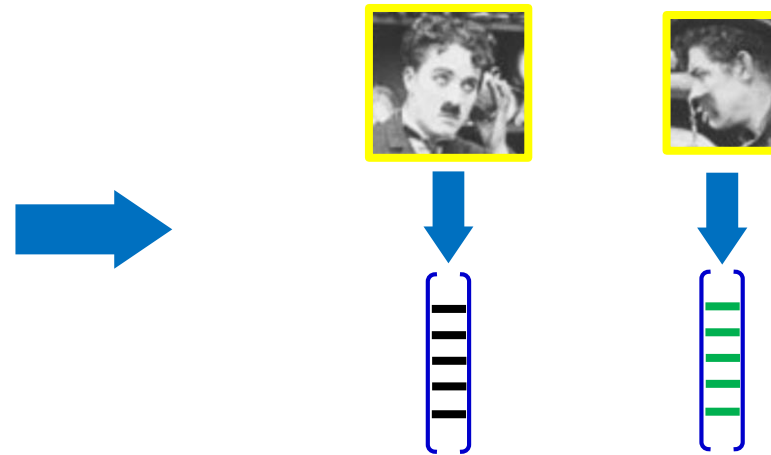


# The Approach

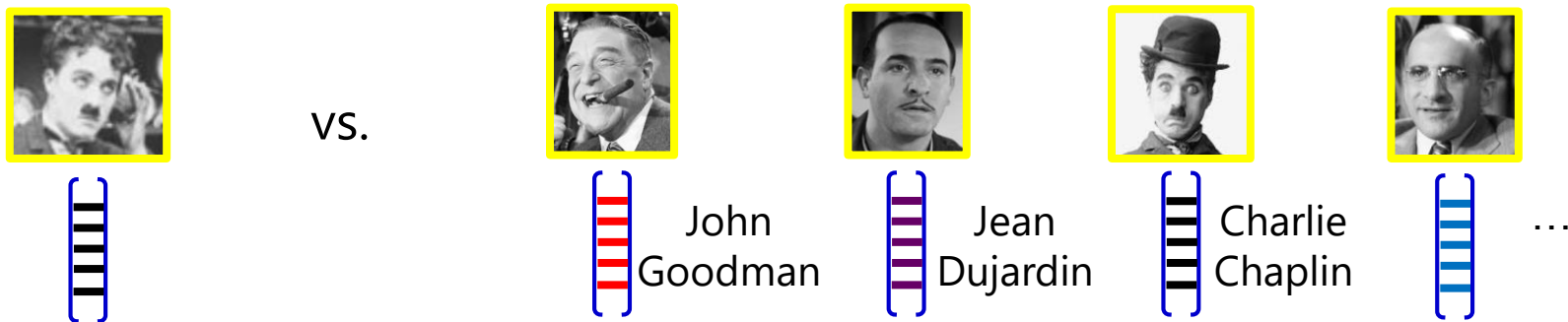
1. Detect Faces



2. Represent each face by a vector

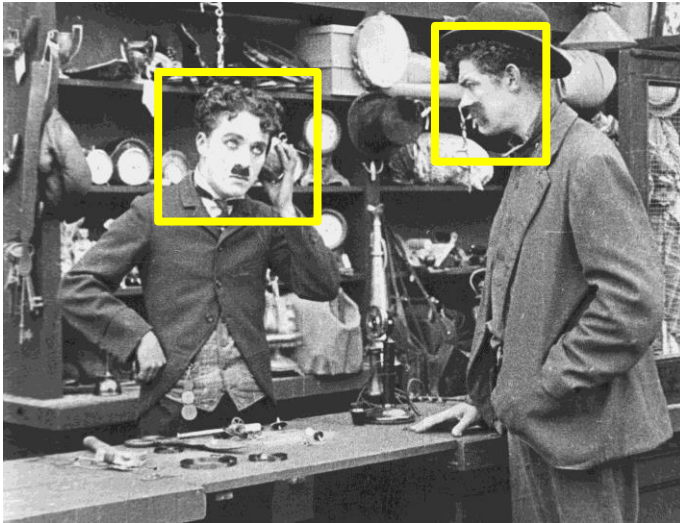


3. Recognize a face from a gallery using closest distance between vectors

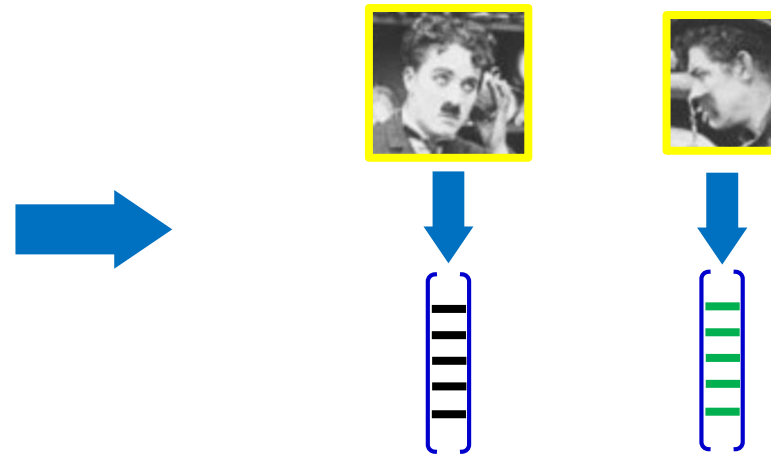


# The Approach

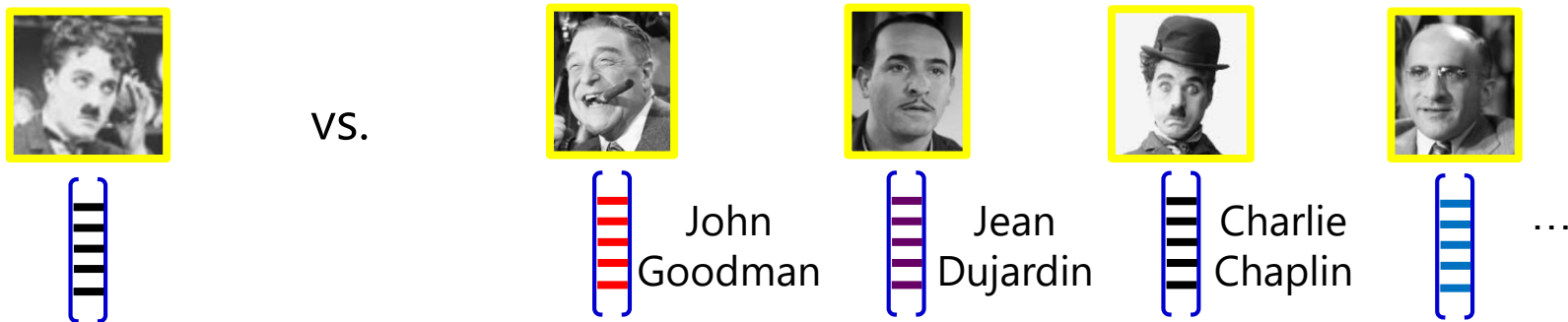
1. Detect Faces



2. Represent each face by a vector



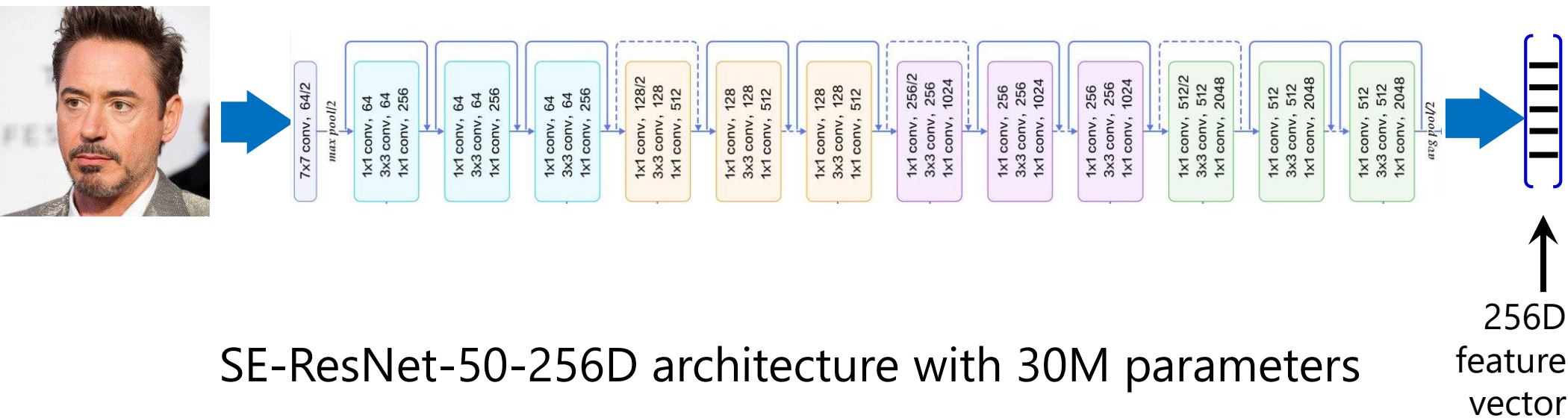
3. Recognize a face from a gallery using closest distance between vectors



For this to work, need vectors to only represent **identity**, and not be affected by expression, pose, lighting, age, etc

# Basic workhorse - face to vector

- Convolutional Neural Network (CNN) deep architecture



SE-ResNet-50-256D architecture with 30M parameters

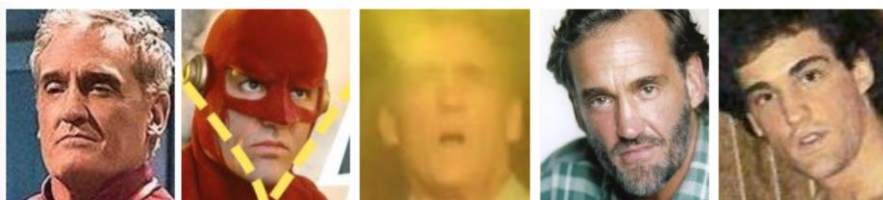
J. Hu, L. Shen, G. Sun, "Squeeze-and-Excitation Networks", IEEE Conference on Computer Vision and Pattern Recognition, 2018

Trained on the VGGFace2 dataset

# Face Dataset (VGGFace2)

- **3M+** face images, **300+** images for each of **9000** identities

Q. Cao, L. Shen, W. Xie, O. M. Parkhi, A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age", International Conference on Automatic Face and Gesture Recognition, 2018



(a) John Wesley Shipp



(e) Roy Jones Jr.



(b) Leymah Gbowee



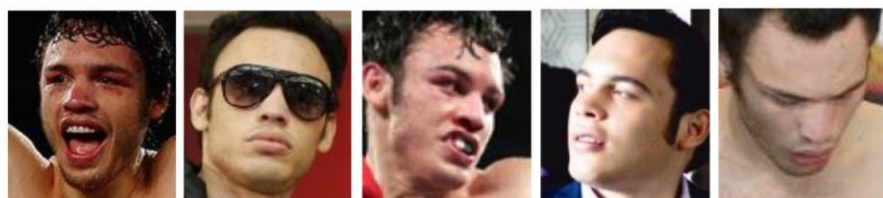
(f) Ruby Lin



(c) Princess Haya Bint Al Hussein



(g) Additi Gupta



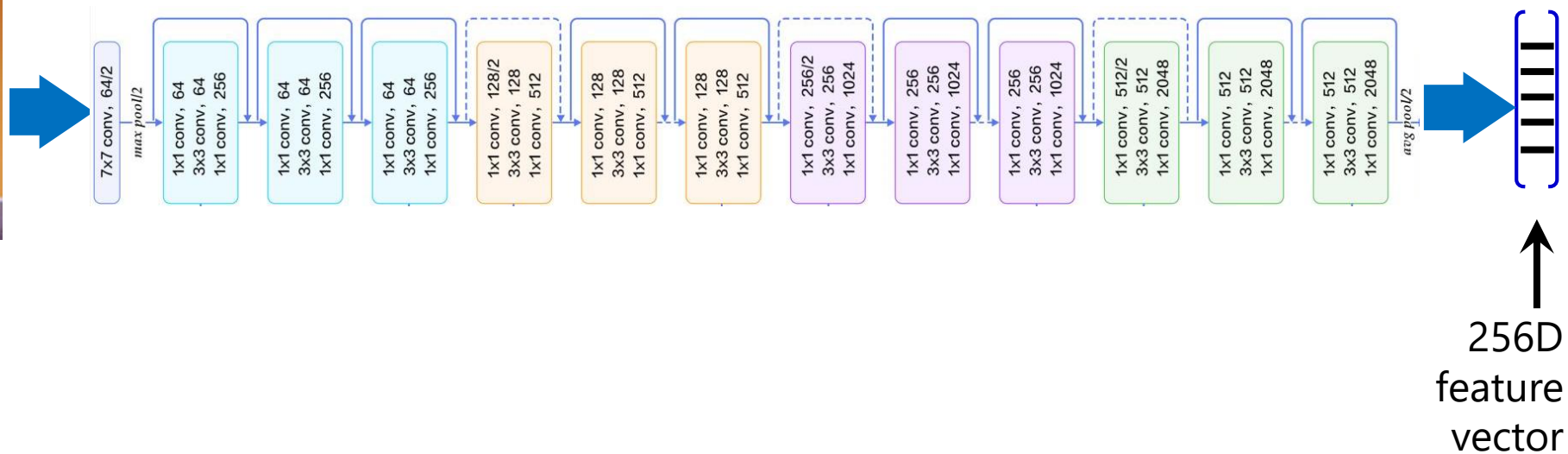
(d) Julio Cesar Chavez Jr.



(h) Lee Jun-gi

# Basic workhorse - face to vector

- Convolutional Neural Network (CNN) deep architecture



Once the network is trained, it can be used to generate feature vectors for previously unseen people

# Automatic Tagging Example

- The British Film Institute provided us with
  - Names of **11K people** of interest
  - Metadata for movies/tv-shows (title, year) and their cast

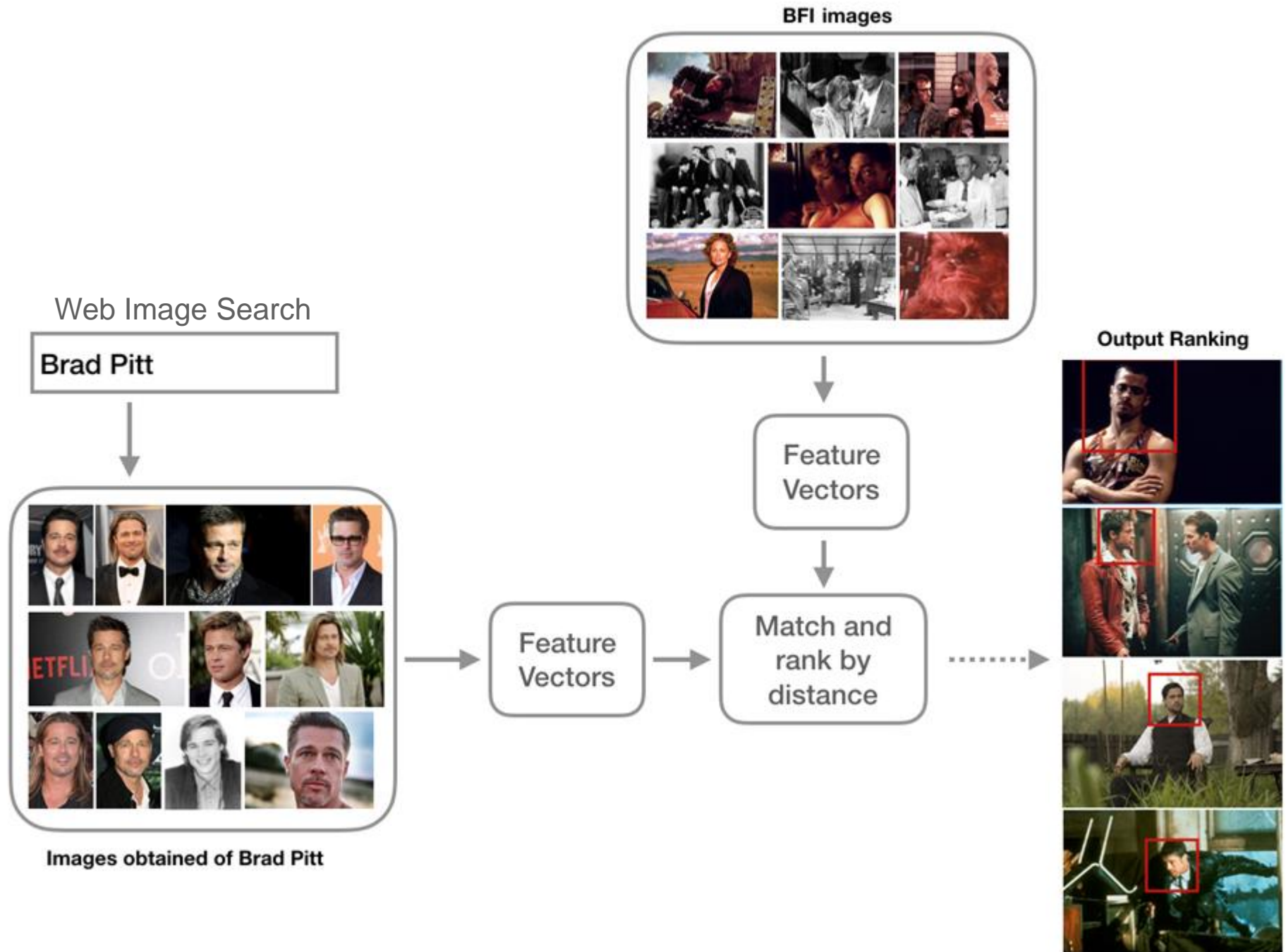


Tags: ?

Film: "Four Weddings and a Funeral"

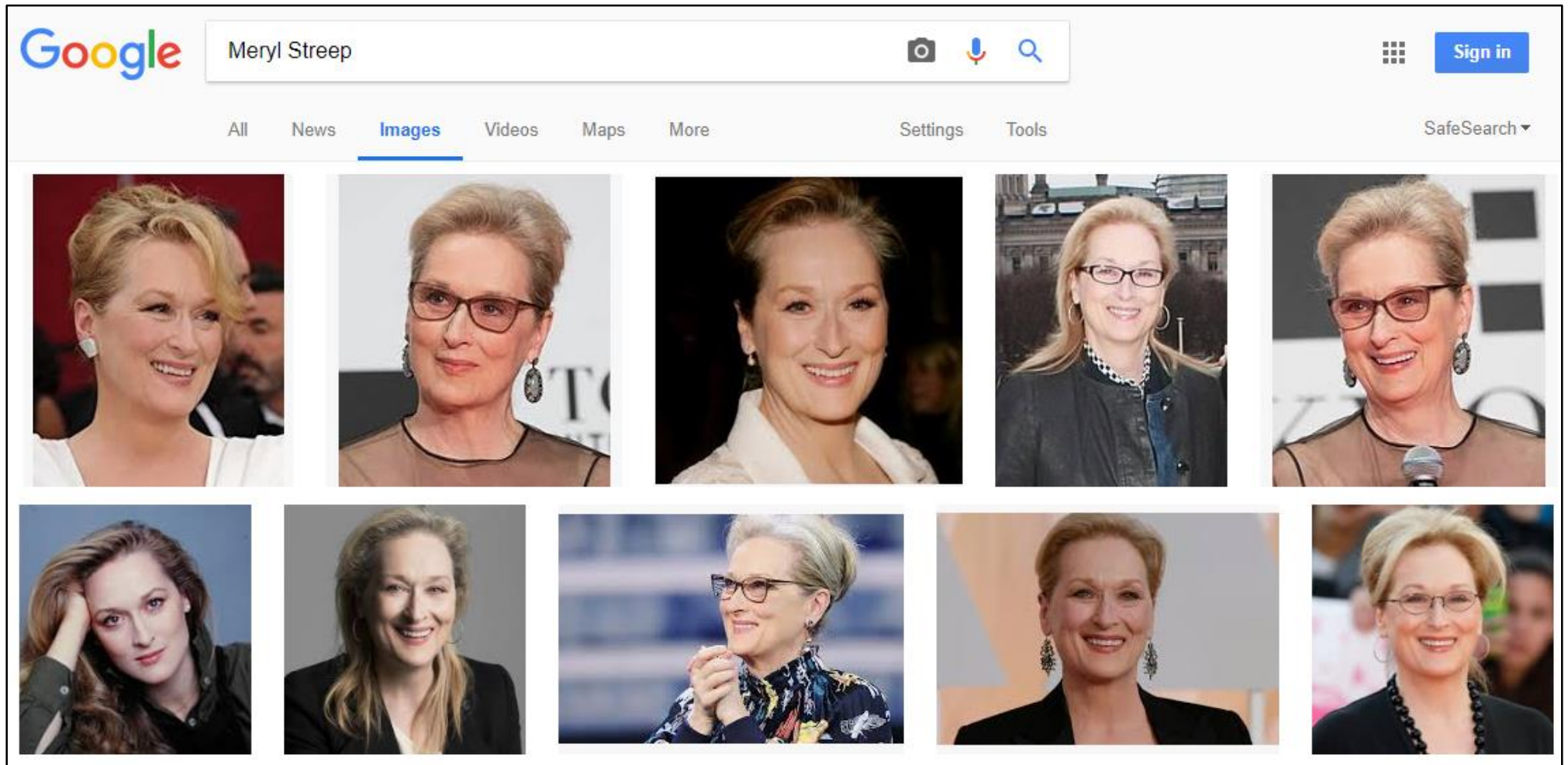
- **GOAL:** Provide tags for these 11K people on dataset with **46K images**

# Tagging the images



# Obtaining identity images

- Use web image search engines (e.g. Google)
  - Works well for “famous” people (~**6.5K** in the dataset)
  - Can automatically determine if famous or not using clustering





# Displaying the image ranking

- Results are ordered starting from the best at the top-left corner
- Location of the detected face can be highlighted

**Text query:** Meryl Streep

Rank:

1



2



3



4

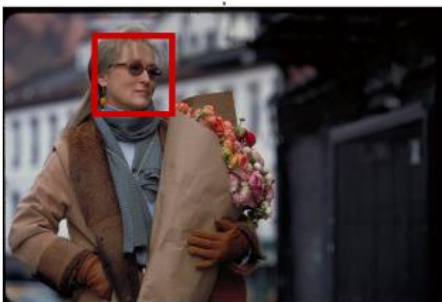


5



Rank:

6



7



8

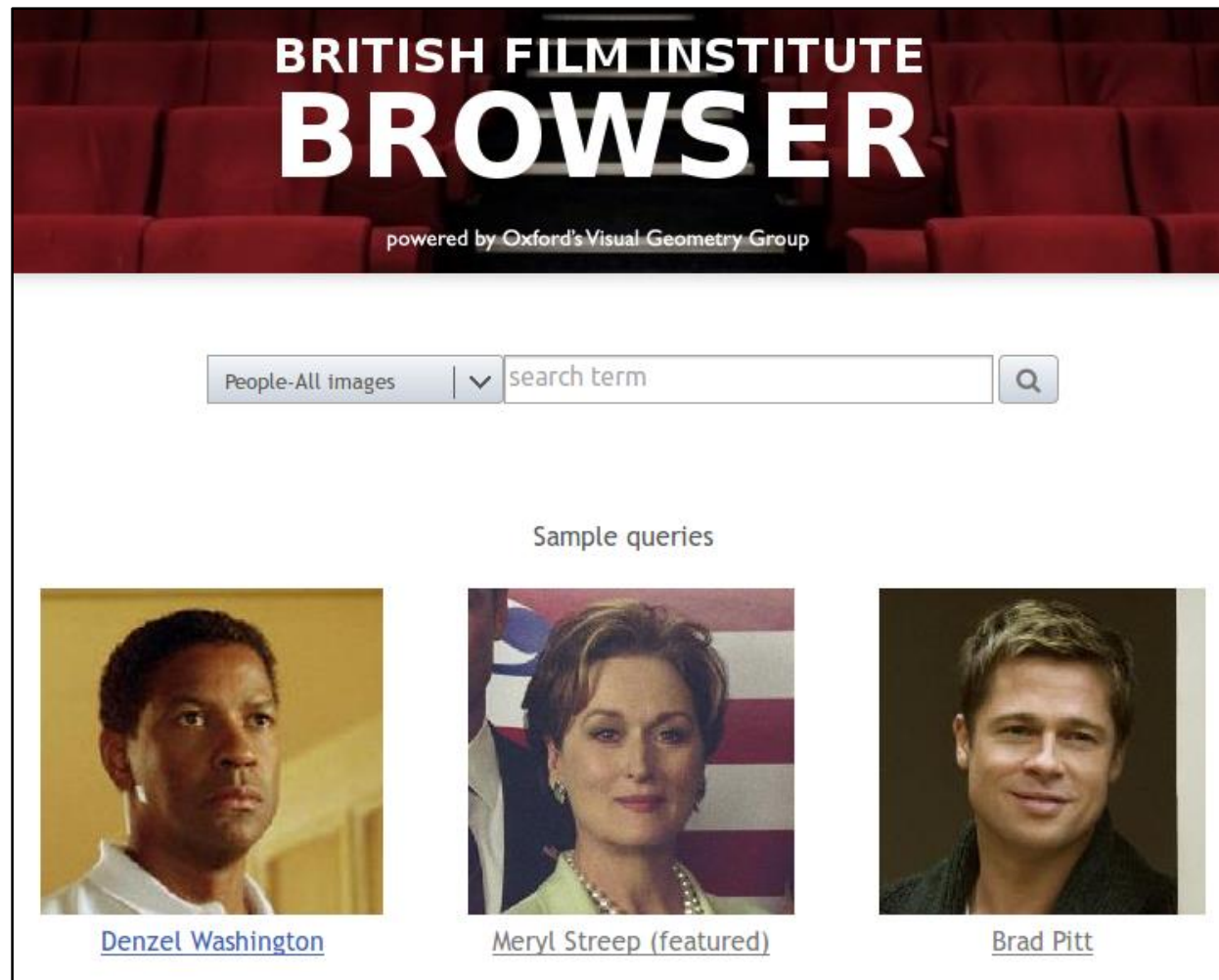


9



# The BFI Browser

- Provides functionality to search for people by name or by movie/tv-show title
- Filter results by “featured” or “non-featured” in the credits



# Anomalies Found

## 1. Images labelled with the wrong film

e.g. all of the images from "War Horse" were labelled as being from "Anterior and Posterior Plaster Beds"



**The Labelled Film Name:** "Anterior and Posterior Plaster Beds" – 1936

**Correct Film Name:** "War Horse" - 2011



**Steven Spielberg** is tagged in this image, but is not in the cast/crew

# Anomalies Found

## 2. Actors/actresses left out of the cast/crew



**Geraldine Somerville** left out of the cast of "Cracker" (TV series)



**Donald Sutherland** left out of the cast of "Commander in Chief" (TV series)

# Anomalies Found

## 3. People appearing on sets when they are not in the cast/crew



**Marilyn Monroe** appears in  
"Ritz" (Film)



**Sharon Stone** appears on  
"Richard & Judy" (TV Series)



**Tony Blair** appears on  
"Richard & Judy" (TV Series)

**1,821 anomalies identified** in total across the dataset

The decision of whether an anomaly is a mislabelling or a surprise appearance **can be done automatically**

# Face Recognition on Video Data

- The BBC News Search system performs visual searches over a large video dataset (~10K hours of video, ~5M keyframes, ~1.5 TB)
- Four search categories are available, including **People**



# Video Demo

The screenshot shows a web browser window with the address bar containing "127.0.0.1/bbc\_search1/". The page has a dark red header with the text "BBC NEWS SEARCH" in large white letters, and "powered by Oxford's Visual Geometry Group" below it. A "Sign in" link is in the top right. Below the header is a navigation bar with buttons for "Instances", "Objects/Scenes", "Text", and "People". A search input field contains "search term/image" and a dropdown menu is set to "BBC News". Below this is a "Why not try?" section with suggestions: "mona lisa", "buckingham palace", and "coca cola". Three small image thumbnails are shown below the suggestions, each with a yellow bounding box. The right side of the page features the "Research project of:" text, the University of Oxford logo, and the text "Visual Geometry Group". A "Getting Started" link is at the bottom right.

127.0.0.1/bbc\_search1/ Search

Sign in

## BBC NEWS SEARCH

powered by Oxford's Visual Geometry Group

Instances Objects/Scenes Text People

search term/image + BBC News

Why not try? mona lisa buckingham palace coca cola

Research project of:  
UNIVERSITY OF OXFORD  
Visual Geometry Group

Getting Started

# Automatic Tagging in News videos

- Similarly to the BFI browser, we could pre-generate lists of results where specific people appears



The image shows a screenshot of a video search interface. On the left, a large video frame displays a man in a suit and tie, likely a news anchor, sitting at a desk. A white arrow points from this frame to a search bar in the top right corner of the interface. The search bar contains the text "03487.jpg" and a search icon. To the right of the search bar are icons for "Admin Tools" and "Sign out". Below the search bar, there are navigation tabs for "Instances", "Objects/Scenes", "Text", and "People", with "People" being the active tab. A "Next >" button is located to the right of these tabs. Below the navigation tabs, the text "Search results page 1 of 100 (5,000 results)" is displayed. The main area of the interface shows a grid of 15 video thumbnails, each featuring the same man in a suit and tie, likely representing different instances of his appearance in news videos. Each thumbnail is labeled "Newsnight" below it.

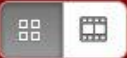


# Extensions: Compound Queries

- Retrieve frames containing both a target **face** and a target **scene**

Query: "Barack Obama in the Office"

**BBC NEWS**  
**SEARCH**



Next >

Search results page 1 of 4 (200 results)



World News Today



BBC Weekend News



BBC News



BBC News



BBC News



Newsnight



Obama The Interview



BBC News



BBC News



BBC News



BBC News



BBC News



BBC News at Six



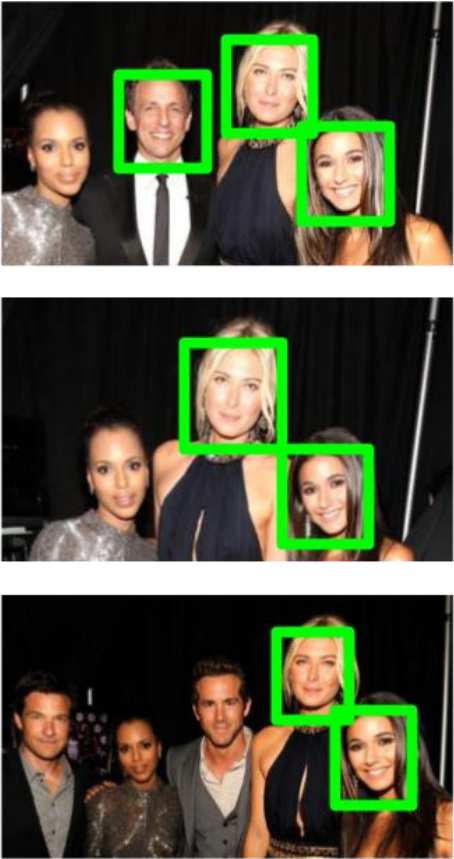
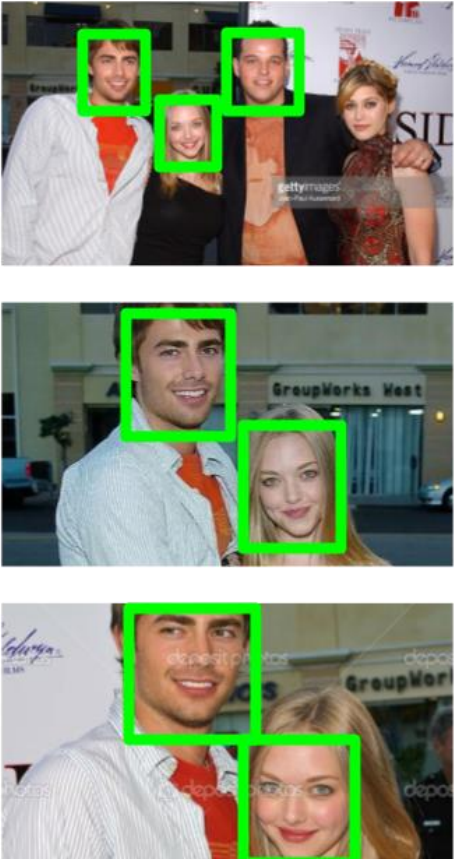
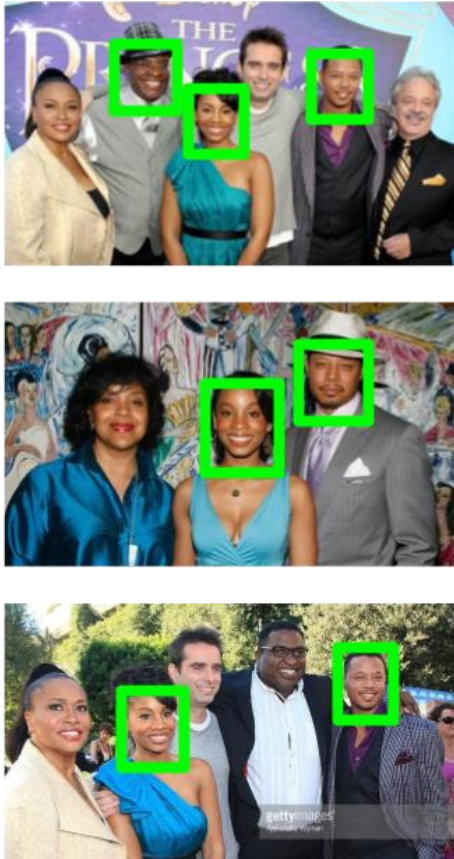
Newsnight



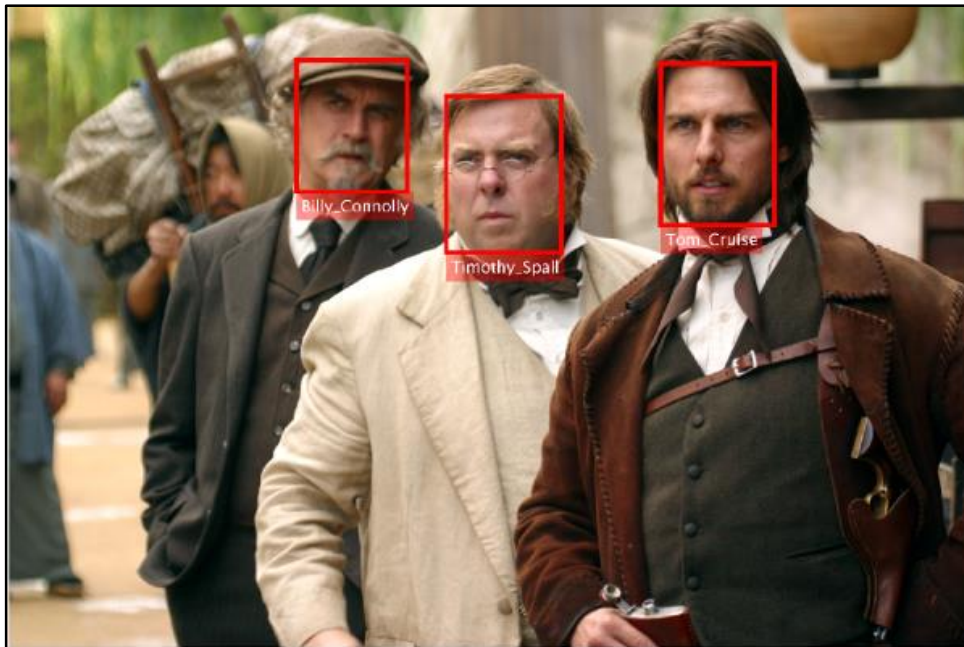
BBC News at Ten

# Extensions: Compound Queries

- Retrieve frames containing **multiple people**

Query:	Query:	Query:
<p data-bbox="472 657 578 686">Top 5:</p> 	<p data-bbox="1025 657 1132 686">Top 5:</p> 	<p data-bbox="1557 657 1664 686">Top 5:</p> 

More information and demos at  
<http://www.robots.ox.ac.uk/~vgg/>



Andrew Brown

Andrew Zisserman

Ernesto Coto