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”
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**

![Images of people of different ages and appearances]
Visual Challenges for Face Recognition

- Performing this task using **Face Recognition** can be challenging.
Visual Challenges for Face Recognition

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

<table>
<thead>
<tr>
<th>Age</th>
<th>Appearance</th>
<th>Expression</th>
<th>Viewpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Age Examples" /></td>
<td><img src="image2.png" alt="Appearance Examples" /></td>
<td><img src="image3.png" alt="Expression Examples" /></td>
<td><img src="image4.png" alt="Viewpoint Examples" /></td>
</tr>
</tbody>
</table>
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

vs.

John Goodman
Jean Dujardin
Charlie Chaplin
...

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

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

Film: “Four Weddings and a Funeral”
Tagging the images

Web Image Search

Brad Pitt

Images obtained of Brad Pitt

BFI images

Feature Vectors

Match and rank by distance

Output Ranking
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
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
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
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

Available at: http://zeus.robots.ox.ac.uk/bbc_search/
Video Demo
Similarly to the BFI browser, we could pre-generate lists of results where specific people appear.
Extensions: Compound Queries

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

Query: “Barack Obama in the Office”
Extensions: Compound Queries

- Retrieve frames containing **multiple people**

<table>
<thead>
<tr>
<th>Query:</th>
<th>Query:</th>
<th>Query:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emmanuelle Chriqui, Maria Sharapova, Seth Meyers</td>
<td>Amanda Seyfried, Daniel Franzese, Jonathan Bennett</td>
<td>Anika Noni Rose, Keith David, Terence Howard</td>
</tr>
</tbody>
</table>

Top 5:
More information and demos at

http://www.robots.ox.ac.uk/~vgg/

Andrew Brown      Andrew Zisserman      Ernesto Coto