Identifying Leading Casts in Movies using Neural Network Models

Submitted by:

Mohd. Musharraf	15BCS0044
Md. Hamid	15BCS0045
Rashid Aziz	15BCS0047

Guided by:

Dr. Tanvir Ahmad

Department of Computer Engineering,

Jamia Millia Islamia,

New Delhi-110025

Abstract

Recognizing human faces in the wild is emerging as a critically important and technically challenging computer vision problem. With a few notable exceptions, most previous works in the last several decades have focused on recognizing faces captured in a laboratory setting. However, with the introduction of databases, face recognition community is gradually shifting its focus on much more challenging unconstrained settings. To further boost the unconstrained face recognition research, we introduce a more challenging approach that has much more variability compared to previous approaches. Our database consists of faces of many known actors collected from their known movies. Unlike old methods which used face detectors to automatically detect the faces from the web collection, images in our database are generated manually from the movies. Manual selection of faces from movies may result in high degree of variability (in scale, pose, expression, illumination, age, occlusion, makeup) which one could ever see in natural world. Our approach will provide a detailed annotation in terms of age, pose, gender, expression, amount of occlusion, for each face which may help other face related applications.

Introduction

There are many popular databases for benchmarking face recognition algorithms. Establishing accuracies on these databases has become a mandatory requirement for reporting progress in face recognition. This resulted in systematic progress in this area, especially for laboratory or co-operative environments. After several decades of research, steady performance has been achieved on these databases. Images in earlier databases were harvested from the Internet thereby capturing large variations that are seen in the wild.

Our main motivation to introduce a new database is to provide a large set of unconstrained face images overcoming some of the earlier limitations. First, since the images in these databases are collected from Internet sources such as Yahoo news, they mostly contain pose, illumination, resolution variations. However, large fractions of images of a particular person have similar public appearance. Second, since the images are collected from Internet through a search query, they may not

cover significant age variations. Third, public figures often retain the identity (appearance, dress patterns, expressions) over sessions reflecting their public behavior. Fourth, older approach was built on images which are detected by Viola-Jones face detector thereby focusing only on end-to-end systems that have automatic face detection followed by recognition. The use of face detector may result in a subset of possible variations of pose, scale, and occlusion. However, we believe that it is equally important to focus on improving the performance of face recognition algorithms in applications such as identifying criminal suspects or image retrieval from videos where manual detection is feasible/affordable and performance is important. The need for more progress in handling pose, occlusion, resolution, etc is recently stressed upon in a recent case study on unconstrained face recognition. We will keep the following design guidelines in mind while building our database.

(i) Capturing images with cameras having different resolutions.

(ii) A database of Indian subjects. Most of the previous databases were built with face images of non-Indians, and the appearance (eg. color and texture) and expressions of the Indian faces could be significantly different from that of these subjects.

(iii) Capturing faces in a natural setting (often referred to as “wild”). This results in face images with wide variations in pose, illumination and partial occlusions unlike many of the existing databases (eg. Yale, AR)

(iv) A database of faces which cover significant age variations for each person. This can help in designing recognition schemes that can robustly recognize individuals independent of the age. This demanded the use of images that were captured over a wide span of years. This is in contrast to the existing databases that are built over a short time period which naturally does not cover age variations.

(v) Capturing rich variety of facial expressions as seen in natural world other than pre-defined expressions like smiling, anger, etc. This is often hard to obtain in a laboratory setting. Also, we would like to have the expressions as natural as possible, and therefore capture these expressions from the experts (popular actors).

(vii) Cover extreme effects of makeups and facial appearance manipulation that a typical face recognition method will have to address in stress-tests.

Proposed Method/ Algorithm

Our method comes with collecting detailed annotation in terms of age, bounding box, movie release, expression, gender, pose, makeup, and possible kind of occlusion. The database is designed through following steps:

1) Selection of movies and actors,

2) Selection of frames from videos,

3) Cropping of faces,

4) Pruning the database

5) Annotation

Selection of movies and actors:

Identification of actors and movies became the critical part in designing the database and optimizing the human labor. First, in order to ensure the diversity in appearance, we will select the movies from wide range. All the movies will be collected from personal collection and YouTube. In the second step, we will select the actors that have a long career span so that we can obtain multiple movies of the actors. For each actor, we will select the movies that give wide variations in age. As far as resolution and quality is concerned, old movies were at poorer resolution while new movies were available at different resolutions. This will result in variation in terms of resolution and quality of images. The number of movies selected for each actor will be varied from $2 - 5$. Since the images will be extracted through a manual process, it is important to minimize the number of movies as much as possible in order to reduce the manual labor. During this stage, we will carefully select the movies in such a way that there is a maximum overlap of actors across movies. For each actor, these movies gave all the variations including significant age variations.

Selection of frames:

Once the movies are collected, we are going to extract the frames from these videos with frame interval of 10. So to build a database that offers serious challenge to recognition algorithms, we will move to manual selection of frames. First, only one frame with signification variation from a shot unless there is another frame with significant difference with the first frame. Second, if there are multiple variations available in a shot, faces with occlusion and pose variation, which offer a serious challenge to recognition algorithms compared to facial expression and illumination, are preferred. Third, those frames will not be considered with small faces and difficult to recognize manually.

Cropping of faces:

After the frames selection, manual cropping of the faces with a tight bounding box will be done. Tool imcrop will be used which gives the cropped region and bounding box information on selecting the face. It is difficult to define a bounding box for a face as it varies from pose to pose.

Pruning the database:

As a post-processing step, through a careful inspection removing of any duplicates or similar images for each subject will be done.

Feeding into Neural Network model:

Finally we will feed our database into our neural network by dividing our dataset in 30-70 ratio. Model will train on 70% and rest 30% will be used for testing.

Technology used

1. Opencv
2. Python
3. Tensorflow
4. Scikit-learn
5. Pandas

References

1. Understading Face Recogntion , Vicki Bruce, Andy Young
   https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2044-8295.1986.tb02199.x
2. Face recognition: A literature survey, W.Zhao, R.Chellappa, P.J Philips, A.Rosenfeld
   https://dl.acm.org/citation.cfm?id=954342
3. Supervised kernel locality preserving projections for face recognition, jian Cheng,Qingshan Liu, hanqing Lu
   https://www.sciencedirect.com/science/article/pii/S0925231204005594