Accio: A Data Set for Face Track Retrieval in Movies Across Age

Esam Ghaleb[1], Makarand Tapaswi[2], Ziad Al-Halah[2]
Hazım Kemal Ekenel[1], Rainer Stiefelhagen[2]
[1]Istanbul Technical University, Istanbul, Turkey
[2]Karlsruhe Institute of Technology, Karlsruhe, Germany
{ghalebe, ekenel}@itu.edu.tr, {tapaswi, ziad.al-halah, rainer.stiefelhagen}@kit.edu

ABSTRACT

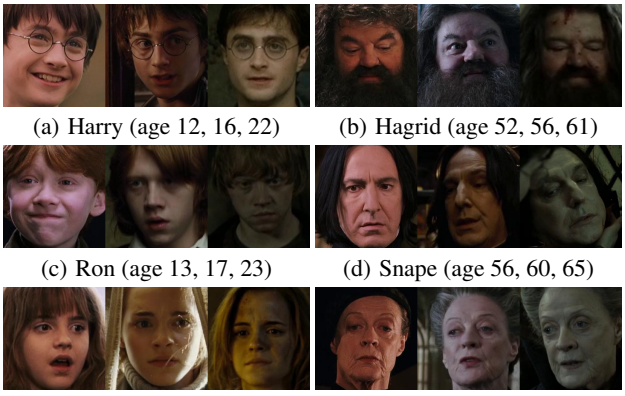
Video face recognition is a very popular task and has come a long way. The primary challenges such as illumination, resolution and pose are well studied through multiple data sets. However there are no video-based data sets dedicated to study the effects of aging on facial appearance. We present a challenging face track data set, Harry Potter Movies Aging Data set (Accio[1]), to study and develop age invariant face recognition methods for videos. Our data set not only has strong challenges of pose, illumination and distractors, but also spans a period of ten years providing substantial variation in facial appearance. We propose two primary tasks: within and across movie face track retrieval; and two protocols which differ in their freedom to use external data. We present baseline results for the retrieval performance using a state-of-the-art face track descriptor. Our experiments show clear trends of reduction in performance as the age gap between the query and database increases. We will make the data set publicly available for further exploration in age-invariant video face recognition.

1. INTRODUCTION

Face recognition in videos has received lots of attention in recent years, and is a challenging topic in computer vision [2, 5, 18]. It has many practical applications: ranging from identifying suspects in surveillance footage [6] to identifying TV characters [5]. There have been significant developments to tackle problems of illumination, resolution, pose and expression [4, 16, 18].

In this paper, we focus our attention to another challenge which is often ignored – changes in face appearance due to aging. While there are multiple still image data sets which analyze age variation (FGNET [1], MORPH [20], CACD [3]), to the best of our knowledge, there is no video face recognition data set with age variation. Often, face track recognition proves to be harder than still image face recognition due to effects of tracking, motion blur and drastic pose variation. Such a data set will help untangle the challenge of age variation catering specially to the broadcast video domain.

We introduce and make available a large, and challenging face track dataset (see Sec. 3) to further evaluate and develop age invariant face recognition systems, specifically for videos. The data set is gathered from the eight movies of the Harry Potter series. The movies span a period of ten years and present a strong setting to study the effects of age variation along with all other challenges of video face recognition. Furthermore, the data set features a large number of face tracks with young faces (age < 20) where the effects of aging are most prominent. Along with a large number of named characters with varying age groups, the data set also contains many unknown tracks which act as distractors.



(a) Harry (age 12, 16, 22) (b) Hagrid (age 52, 56, 61)
(c) Ron (age 13, 17, 23) (d) Snape (age 56, 60, 65)
(e) Hermione (age 11, 15, 21) (f) McGonagall (age 67, 71, 76)

Figure 1: Change in character appearance across the movies

Fig. 1 shows the changes in facial appearance for six characters taken from movies 1, 4 and 8. Note that these images are deliberately chosen to have frontal pose and decent illumination to highlight the variation in their age.

As a benchmark for the face track recognition, we present two primary scenarios: (i) face track retrieval within movies; and (ii) face track retrieval across movies. Similar to LFW [10] or YouTube Faces [23] we also propose to divide the usage of data in a restricted and unrestricted setting. Sec. 4 presents more details about the challenges, and some baseline results.

2. RELATED WORK

In the following, we first discuss publicly available data sets for improving face recognition across ages. We examine the area related to face track retrieval in multimedia data, and finally discuss some prominent works in age-invariant face recognition.

2.1 Data sets

Previous face recognition data sets which focus explicitly on the problem of aging are all on still images [1, 3, 20, 21]. Data sets

[1]The Summoning Charm *Accio* in the Harry Potter series retrieves an object at a distance, here used to signify face track retrieval.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
ICMR '15, June 23 - 26, 2015, Shanghai, China
Copyright is held by the authors. Publication rights licensed to ACM.
ACM 978-1-4503-3274-3/15/06 $15.00.
http://dx.doi.org/10.1145/2671188.2749296

for video face track recognition are often obtained from one movie or one TV series season [2, 5] and thus do not present challenges of age variation. Popular face (track) verification data sets such as LFW [10] or YouTube Faces [23] do not discuss the problem of matching faces across age. To the best of our knowledge this is the first video face recognition data set which is obtained from challenging settings and shows clear trends of the effects of aging.

2.2 Face track retrieval in videos

Among the first popular works in face track retrieval is [22]. They detect frontal faces, track them using a region-tracking method, and use the SIFT descriptor around 4 facial landmarks for matching. This enables to search a query face track in the entire video and find corresponding shots in which the person appears. In a similar retrieval scenario, [7] includes non-frontal faces, and [13] presents a compact descriptor to match face tracks. However, after the seminal paper by Everingham *et al*. [5], most works focus on assigning a name to all face tracks (*e.g*. [2]).

2.3 Age invariant face recognition

There are multiple studies which consider the problem of aging in face recognition. Most systems can be divided in two categories:

(i) *Generative* approaches [8, 12, 17, 19] simulate the face image at the target age and then perform face recognition. Typically they build 2-D or 3-D generative models to synthesize face aging. For example, [19] uses face landmarks to estimate facial growth parameters and predicts the appearance at a certain age.,

(ii) *Discriminative* approaches such as [3, 11, 14, 15] use learning methods to build age invariant face recognition systems. For example, recently, Chen *et al*. [3] use a reference set representation for each person, his/her age, and the face landmark. To achieve an age-invariant representation, they model the problem as least squares and pool across the age space.

Apart from previous methods, there are examples of modeling the appearance through latent spaces. For example, Gong *et al*. [9] propose to use two latent factors: an age-invariant identity factor, and an age factor affected by aging. The observed face appearance is modeled as a combination of the two, and separating the observation allows to perform recognition using the identity part.

3. DATA SET

We present the Harry Potter Movies Aging Data *Accio* to study and develop age invariant face recognition and retrieval methods.

3.1 Data source

This data set is collected and organized using the eight Harry Potter movies that were released in a period of ten years – 2001 to 2011. Each movie in itself contains multiple interesting challenges (i) *illumination* - large number of bright and dark scenes; (ii) *pose* - many non-frontal tracks, actors typically do not look at cameras; (iii) *resolution* - tracks from 25 to 500 pixels; and (iv) *distractors* - roughly 40% of all face tracks are *unknown* characters which appear in the background making the problem an open set task. Finally, the use of all eight movies acts as a good source of variation in the facial appearance due to *aging*.

To obtain the face tracks, we first detect video shot boundaries. Within each shot, following [2] we perform tracking-by-detection using a particle filter. We use multi-pose detectors which cover the entire range of *pan* and *roll* face angles.

The data set contains a large number of tracks of young actors (age < 20). We believe that this is an important aspect since most of the changes in facial appearance occur early in life. Fig. 2 presents

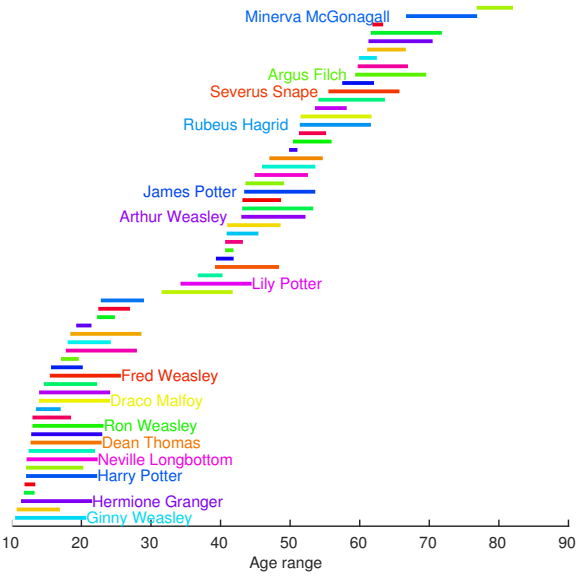


Figure 2: Actors age progression. Each line represents the age span for that actor appearing in the movies.

Table 1: Number of tracks and characters across the movies

	HP-1	HP-2	HP-3	HP-4	HP-5	HP-6	HP-7	HP-8
# characters	36	42	34	44	47	41	56	54
# face tracks	5249	5335	3919	7616	5850	3354	2910	4231
# unknown tracks	2006	1874	1437	4237	2316	1116	623	2025
# named tracks	3243	3461	2482	3379	3534	2238	2287	2206

the age variation for multiple actors. Please note that the figure plots only those actors which appear in at least 2 movies.

3.2 Face track descriptors

Face tracks of the data set are encoded using state-of-the-art Video Fisher Vector Faces (VF^2) descriptor [18]. VF^2 is a face track representation that aggregates a large set of dense SIFT features which are extracted from all the face images of the track. We follow the steps and parameters from [18] and use Root-SIFT with a stride of 1 pixel. The resulting 128 dimensional feature is down-projected via PCA to 64 dimensions and augmented with the spatial location (x, y). A Gaussian Mixture Model (GMM) of 512 Gaussians is trained on all the face tracks and Fisher encoding is performed. This yields a $2 \times 512 \times 66 = 67584$ dimensional descriptor for each track. To compare two tracks we use the standard Euclidean distance between two Fisher vector descriptors for our baseline experiments.

3.3 Statistics

Table 1 presents the number of characters, the number of tracks in each movie and a split betweeen *named* and *unknown* (distractor) tracks. The dataset contains in total of 38,464 face tracks. 22,830 (59.4%) of these face tracks are labeled with one of 121 character identities as they appear in the film series. The rest of the face tracks (40.6%) belong to un-named cast and are treated as distractors for the retrieval evaluation.

Each movie in the Harry Potter film series has varying number of characters and number of tracks. The three main characters (*Harry Potter*, *Hermione Granger*, and *Ron Weasley*) encompass roughly 40-60% of all face tracks in each movie.

The actors span a range of age from 10 to 88 years. Note that we compute the age based on the movie release date and the birth

Table 2: Actors age distribution

	10-19	20-39	40-49	50+	Young	Adults
# characters	33	34	26	44	33	92
# labeled tracks	12,598	4,303	1,728	3,636	12,598	9,667

Table 3: Comparison of age-invariant face recognition data sets

	video?	#images	#people	Age span
FGNET [1]	No	1,002	82	0-45
MORPH [20]	No	55,134	13,618	0-5
CACD [3]	No	163,446	2,000	0-10
Accio (Ours)	Yes	38,464 tracks	121	0-10

year, which can add a slight discrepancy up to a couple years. The largest age difference between face tracks for the same actor is 10 years for actors which appear in both the first and last movie.

Table 2 shows the distribution of tracks and number of characters across various age spans. We see an equitable distribution of characters among the various age groups. However, most of the tracks belong to the first 10-19 (*Young*) age range.

Finally, we present in Table 3 a comparison between existing data sets. Note that unlike other data sets, ours works with face tracks, which on average are 50 frames (2 seconds) long thus yielding over 1.9 million face images.

4. EXPERIMENTS

We first define the evaluation protocol, followed by baseline experiments for face track retrieval. The retrieval experiments are evaluated using two popular measures: (i) mean average precision (MAP), and (ii) precision @ k (number of correct results in top k).

4.1 Evaluation protocol

Tasks The face track retrieval is performed as two tasks.

1. *Within* movies: In this setting, we look at individual movies one at a time. Each named track in the movie is used as a query, while the remainder (other named tracks + unknown tracks) from the same movie form the database. This setting evaluates the retrieval performance subject to typical face recognition challenges such as pose and illumination.

2. *Across* movies: In this setting we analyze the effect of age variation among the actors by evaluating face track retrieval across movies. We consider each named track within one movie (*e.g.* HP-1) as a query, while all tracks of a different movie (*e.g.* HP-2) form the database.

Benchmarks We suggest two protocols which involve both the above tasks, however vary in the extent to which data can be used to learn supervised or unsupervised models.

(i) *Restricted*: In this protocol the experiments should not use any external data. That is, only the query track is available as a positive training sample and retrieval models should not be trained using other *Accio* face tracks, or any external face images/tracks. This can be used for example to compare different face track descriptors, or perform automatic and unsupervised methods such as query expansion and domain adaptation.

(ii) *Unrestricted*: In this protocol the researchers are allowed to use *external* data for training models, for example to model the changes in facial appearance through aging. However, no other tracks from the *Accio* data set (except the query track) are to be used for training actor models.

Table 4: Within movies evaluation

	HP-1	HP-2	HP-3	HP-4	HP-5	HP-6	HP-7	HP-8
p @ 1	90.8	89.3	90.9	85.1	89.4	91.1	91.3	89.4
p @ 5	81.5	79.0	79.1	72.3	79.3	80.4	82.1	78.6
p @ 10	75.9	71.6	70.2	64.5	72.0	72.2	75.3	69.7
p @ 20	71.9	64.0	61.5	57.3	63.7	64.3	70.7	63.0
p @ 50	67.9	57.6	53.0	54.7	59.4	55.8	68.0	57.2
p @ 100	70.9	54.8	53.1	58.5	55.5	51.1	62.8	53.0
MAP	42.40	31.73	31.19	28.36	32.09	30.38	38.55	33.21

Table 5: Across movies face track retrieval performance (MAP)

	HP-1	HP-2	HP-3	HP-4	HP-5	HP-6	HP-7	HP-8
HP-1	42.4	33.1	26.8	24.5	21.5	22.2	27.4	20.0
HP-2	32.1	31.7	22.9	20.1	18.6	19.8	23.2	17.9
HP-3	23.0	20.8	31.2	19.9	18.1	21.0	23.4	16.2
HP-4	27.2	23.3	27.1	28.4	20.9	23.0	24.4	20.5
HP-5	20.3	18.5	22.0	17.4	32.1	22.5	25.7	20.8
HP-6	21.4	18.3	24.3	19.8	23.7	30.4	24.5	17.9
HP-7	24.3	20.8	24.7	19.6	26.4	24.8	38.5	25.5
HP-8	19.9	18.8	20.2	18.1	21.7	20.8	28.4	33.2

Finally, if the data is used in a different manner that does not comply with either of the two previous protocols, for example the face tracks can be used for a classification task [5], the researchers can refer to this as a *free-for-all* scheme.

4.2 Within movies face track retrieval

We now present the results of within movie face track retrieval (Task 1) under the restricted protocol. In each movie, the labeled tracks are used as a query, while all tracks from the same movie form the database. For example, in HP-1, we have 3243 tracks (taken one at a time) as query and 5248 tracks (the number of tracks minus the query) as part of the database.

Table 4 presents the performance of retrieval on the eight movies. Precision @ k (p @ k) values are presented in multiple steps [1, 5, 10, 20, 50, 100], while the MAP is used as the overall performance indicator. As expected the performance for small values of k is much higher and this can be used to perform query expansion.

4.3 Across movies face track retrieval

While the previous section evaluated performance through challenges of pose and illumination, we now look primarily at the problem of aging (Task 2). We expect that the retrieval performance is good to fair when the age difference is small (within same movie, or $\pm$ 1 movie), while it deteriorates as we compare across a large time gap (*e.g.* HP-1 vs. HP-8).

Table 5 presents overall retrieval performance in MAP comparing all eight movies. Rows represent the movie from which query tracks are obtained, while the tracks of the movie in columns serves as the database (*e.g.* query from HP-1 and database from HP-7 shows a MAP of 27.4). As can be seen in the table, the larger the age gap between the query and database the MAP performance reduces. HP-7 appears to be an exception and this could be attributed to the fact that this movie has the least number of tracks, many ($\sim$60%) of which belong to the three primary characters. The diagonal represents within movie performance and is the same as the last row of Table 4.

In Table 6, we consider all queries from the first movie HP-1 and split them into two based on age. Tracks of young actors (age < 20) when used as queries tend to perform worse in the retrieval across

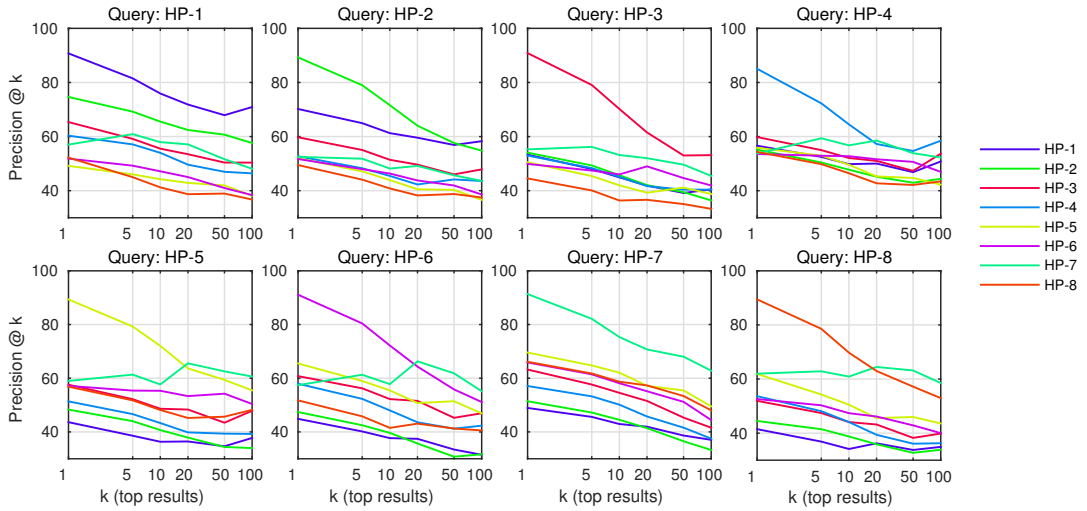


Figure 3: Precision @ k results for across movies retrieval. This figure is best viewed in color.

Table 6: Comparison of MAP scores for adult vs. young actors. The queries are taken here from HP-1 movie.

	HP-1	HP-2	HP-3	HP-4	HP-5	HP-6	HP-7	HP-8
Adult	47.5	40.9	33.6	32.5	30.5	29.8	23.0	21.8
Young	40.9	31.2	25.2	22.9	19.3	20.8	28.4	19.6

movies as compared to tracks of adults. Empirically we see that the face appearance changes more during the youth years and thus has a stronger influence (more errors) on the retrieval performance.

Finally, we present the precision @ k scores for each movie in Fig. 3. Each subplot represents queries from a single movie (mentioned in the title) and the databases are all movies (plotted as separate lines). For any subplot, note how the precision is higher for movies around the movie from which the query tracks are selected. For example, choosing queries from HP-1, the best performance is seen when comparing against tracks from HP-2, HP-3, HP-4. Similarly, for queries from HP-2, the performance against tracks from HP-1 and HP-3 is better than other movies. There are some exceptions, specially at smaller values of k, however the general trend is fairly clear. Together with Table 5, the analysis is a clear indication that age variation is a strongly felt effect in the data set.

5. CONCLUSION

We introduce the *Accio* data set for analyzing the effect of aging in video face recognition methods. *Accio* features 121 characters spread over multiple age groups, and has more than 38,000 face tracks of which a large pool (40%) are distractor tracks. We present two primary tasks for evaluation: within and across movie face track retrieval. Depending on whether the external data is used or not, we present two protocols: unrestricted and restricted. We see a steady decline in the retrieval performance as the age gap between the query and database increases. The data can be downloaded at cvhci.anthropomatik.kit.edu/projects/mma.

Acknowledgments This work was funded by the German Research Foundation under contract no. STI-598/2-1; TUBITAK within the CHIST-ERA project titled CAMOMILE, project no. 112E176; and a Marie Curie FP7 Integration Grant in the 7th EU Framework.

6. REFERENCES

[1] FG-NET Aging Database.

[2] M. Bäuml, M. Tapaswi, and R. Stiefelhagen. Semi-supervised Learning with Constraints for Person Identification in Multimedia Data. In *CVPR*, 2013.

[3] B.-C. Chen, C.-S. Chen, and W. H. Hsu. Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval. In *ECCV*, 2014.

[4] D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of Dimensionality: High-Dimensional Feature and Its Efficient Compression for Face Verification. In *CVPR*, 2013.

[5] M. Everingham, J. Sivic, and A. Zisserman. "Hello ! My name is ... Buffy" – Automatic Naming of Characters in TV Video. In *BMVC*, 2006.

[6] R. Feris, R. Bobbitt, L. Brown, and S. Pankati. Attribute-based People Search: Lessons Learnt from a Practical Surveillance System. In *ICMR*, 2014.

[7] M. Fischer, H. K. Ekenel, and R. Stiefelhagen. Person re-identification in TV series using robust face recognition and user feedback. *Multimedia Tools and Applications*, 2010.

[8] X. Geng, Z.-H. Zhou, and K. Smith-Miles. Automatic Age Estimation Based on Facial Aging Patterns. *T-PAMI*, 29(12):2234–2240, 2007.

[9] D. Gong, Z. Li, D. Lin, J. Liu, and X. Tang. Hidden Factor Analysis for Age Invariant Face Recognition. In *ICCV*, 2013.

[10] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Technical Report 07-49, Univ. of Massachusetts, Amherst, 2007.

[11] B. Klare and A. K. Jain. Face recognition across time lapse: On learning feature subspaces. In *Intl. Joint Conf. on Biometrics*, 2011.

[12] A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Trans Syst Man Cybern: B Cybern*, 34(1):621–628, Feb 2004.

[13] Y. Li, R. Wang, Z. Cui, S. Shan, and X. Chen. Compact Video Code and Its Application to Robust Face Retrieval in TV-Series. In *BMVC*, 2014.

[14] Z. Li, U. Park, and A. K. Jain. A Discriminative Model for Age Invariant Face Recognition. *IEEE Trans on Inf. Forensics Security*, 6(3), 2011.

[15] H. Ling, S. Soatto, N. Ramanathan, and D. Jacobs. Face Verification Across Age Progression Using Discriminative Methods. *IEEE Trans on Inf. Forensics Security*, 5(1):82–91, 2010.

[16] E. G. Ortiz, A. Wright, and M. Shah. Face Recognition in Movie Trailers via Mean Sequence Sparse Representation-based Classification. In *CVPR*, 2013.

[17] U. Park, Y. Tong, and A. Jain. Age-Invariant Face Recognition. *T-PAMI*, 32(5):947–954, May 2010.

[18] O. M. Parkhi, K. Simonyan, A. Vedaldi, and A. Zisserman. A Compact and Discriminative Face Track Descriptor. In *CVPR*, 2014.

[19] N. Ramanathan and R. Chellappa. Modeling Age Progression in Young Faces. In *CVPR*, 2006.

[20] K. Ricanek and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *IEEE FG*, 2006.

[21] S. Setty and et al. Indian Movie Face Database: A Benchmark for Face Recognition Under Wide Variations. In *NCVPRIPG*, 2013.

[22] J. Sivic, M. Everingham, and A. Zisserman. Person spotting: video shot retrieval for face sets. In *CIVR*, 2005.

[23] L. Wolf, T. Hassner, and I. Maoz. Face Recognition in Unconstrained Videos with Matched Background Similarity. In *CVPR*, 2011.