Large Scale Unconstrained Open Set Face Database

[1]Archana Sapkota
[1]University of Colorado at Colorado Springs
asapkota@vast.uccs.edu

[1 2]Terrance E. Boult
[2]Securics Inc, Colorado Springs
tboult@vast.uccs.edu

Abstract

This paper addresses large scale, unconstrained, open set face recognition, which exhibits the properties of operational face recognition scenarios. Most of the existing face recognition databases have been designed under controlled conditions or have been constructed from the images collected from the web. Face images collected from the web are less constrained than a mug-shot like collection. However, they lack information about the imaging conditions and have no operational paradigm. In either case, most of the databases and evaluation algorithms have taken the form of "closed set" recognition, in which all testing classes are assumed to be known at training time. A more realistic scenario in face recognition is an "open set," where limited knowledge is available at training time and unknown classes can be present at test time. The database we provide supports the open set paradigm, which more closely mimics actual usage than classic closed set testing. The database also exhibits the natural variability among the face images such as pose, illumination, scale, expressions, occlusion, etc. Our goal is to provide around 100,000 images of more than 1,000 people. Also, with this paper, we release part 1 of the database, which consists of 6,337 images from 308 subjects. The paper discusses the details of the database followed by the challenges and results of baseline algorithms.

1 Introduction

Face recognition is a well studied problem in computer vision, with about 2,520,000 search results on Google Scholar at the time of the writing this paper. There are also almost 50 different databases to evaluate the face recognition algorithm [2]. Is there still a need for a new database? We believe there is. To meet this need, we propose and provide a database which is very close to the operational scenarios.

Most of the evaluation methods of recognition problems in computer vision literature have considered the "closed set" form, in which all classes that appear during testing are present at the training time as well. This is not a realistic scenario in general recognition problem nor in face recognition. Consider a surveillance scenario in which an authority is watching the people coming in front of the camera, but is only interested in a certain list of people. Face recognition in this scenario should be able to "recognize" the people seen before and be able to say "do not recognize" to those not seen before. This kind of form is called "open set" recognition. A recent work [20] formalizes the "open set" recognition and explains how recognition problems should be considered as open set. In this work, we provide a database and a protocol for evaluating open-set algorithms.

Previous works on face recognition have also considered face recognition problem as either a pair matching or a general identification problem. A pair matching problem states that, given a pair of images, the system should correctly recognize whether the images are from the same person or different people. An identification problem is considered more of an enrollment problem where one is interested in knowing whether a subject is already present in the enrolled gallery. If the person is present, the identification system should correctly recognize who the person is. Moving beyond these problems, a successful application of face recognition technology in real world problems in security or personal photo collection demands more practical evaluation protocols. For example, in security purposes, it is not only of interest to identify whether the person is already present in the gallery, but also to see how often someone appears in front of the camera [4]. Similarly, a frequent observer or a traveler in the airport may or may not be a subject of interest. For such applications, it is necessary to do multiple identifications on various days of appearance. With a surveillance camera, a sequence of images of a person at a particular time can be captured and used for identification. This problem can first be solved by clustering the sequence of images during the test time and the sequence of gallery images or the models generated by considering a sequence of images in the gallery. Because of scenarios like this, face identity clustering and sequence-to-sequence matching are interesting and practical problems. Similarly, in a personal photo collection with auto tagging of pictures [21], the identity clusters and suggestion of tags can tremendously reduce the human effort. In this work, we provide the database for open set identification and verification problems, as well as introducing the protocol for identity clustering.

A lot of existing databases provide the images which are captured during controlled setup in the laboratory to simulate the real life operational scenarios. These databases are very useful in understanding the variability of parameters in the database and unknown labels at the time of capture. However, there is a concern in how much these variability in images represent a real life scenario. As mentioned by the authors in [9], "that it is difficult to gauge exactly which distributions of various parameters one should use in order to make the most useful database. What percentage of subjects should wear sunglasses? What percentage should have beards? How many should be smiling? How many backgrounds should contain cars, boats, grass, deserts, or basketball courts?"

Previous attempts in designing a more unconstrained setup involved collecting the databases from the web, such as LFW [9], PubFig [11]. However, the work in [18] suggests that the algorithm can exploit accidental irrelevant information in such images to improve performance. They showed that a simple pixel based algorithm can perform comparable or better than complex algorithms in such databases and pointed a need of designing datasets that are close to real life scenarios. They proposed a synthetic database which does not represent a real world scenario. One of the problems with the images collected from the web is that there is a huge bias in the background. The work of [11] shows that humans were able to predict the same or different people with the accuracy of 94.27% with the face images masked, and the human performance decreased when only faces were shown without the background. Celebrities tend to be present in similar background conditions. A tennis player will most of the times have a tennis court in the background or wear a similar hat, a political figure or a singer can have a microphone in front of her in most of the images. The attribute classifiers as presented in [11] might not consider the background images, but the face recognition algorithms which use context might be highly biased due to the background in the database.

To summarize, the main contributions of this work are:

1. Provide a large scale unconstrained database and protocols for open set face recognition with the results of baseline algorithms for open set recognition and clustering algorithms.

2. Provide ground truth imaging conditions and identity for the usability of the database of wide ranges of challenges in face recognition such as pair matching, identification, sequence to sequence matching, and clustering.

2 Existing databases

There are almost 50 different databases that have been provided in the past to study the challenges in face recognition. The website [2] provides the summary of these databases and [9] summarizes some of these databases. We discuss some of the relevant databases which are more unconstrained and have been widely popular in the literature. We briefly state the statistics of the provided data and discuss the shortcomings of those databases.

Labeled Faces in the Wild (LFW): Labeled Faces in the Wild has been a very popular database. Most of the recent state-of-the-art algorithms have been evaluated on this database. The database provides images collected from the web with pair matching protocol. It consists of 1100 match and 1100 non match pairs for training and 500 match and 500 non-match pairs for testing.

However, this database has some limitations. First, though there are 5749 subjects, the database contains only 610 subjects with 3 or more images. This makes it difficult to use the database for identification problems using learning-based methods. The second limitation is the background bias present in the database. As shown by Kumar at al. the human performance decreases when the background is removed. Even without the face mask, the verification rate is around 95% [11]. Pinto at al. showed that even a simple pixel based algorithm can sometimes outperform the complex algorithms due to accidental use of irrelevant background information used by face recognition algorithms. Another problem is that the database does not have any defined variation, such as the amount of blur, scale, noise, or occlusion. Lastly, the database is not big enough for open set recognition.

PubFig Database:[11] Another database that is similar to LFW is PubFig. Images in this database are collected from the web. The database contains 60,000 images from 200 people. The verification protocol of this database consists of 20,000 pairs of images from 140 people. The database also provides the partitions in terms of variability such as pose, lighting and expression. Each category is further divided into easy and difficult subsets. A subset of PubFig is constructed to define an identification protocol ([18]). This subset of PubFig is called PubFig83.

Though PubFig was able to define the variability of the database and have more images per person, it still contains only 200 people. PubFig83 contains only 83 people with 100 or more images. This setup has similar limitations to LFW in terms of representation of operational scenarios and limitation in openset evaluation.

SCface Databse: Surveillance Camera face database (SCface) is another example of a reasonably unconstrained database. Images are collected using 6 different surveillance cameras at 3 different distances. It consists of 130 subjects and a total of 4160 images. The database consists of mostly frontal test images at different distances. Though the database is not popularly used, this is one of the hardest database for identification. The performance of baseline face recognition algorithm is very low as mentioned in [7]. However, this database is limited by the small number of subjects and it does not contain operational variability in face images such as expressions, pose, occlusion.

Multi-PIE [8] : To overcome the limitations of PIE database, a Multi-PIE database was proposed. The Multi-PIE database contains images under 15 view points and 19 illumination conditions. Though both PIE and Multi-PIE provide a wide ranges of variations across pose, illumination and expression and help advance the face recognition research, the database is limited to controlled indoor images. This database does not contain the uncontrolled images in a practical sense. The expressions and poses are guided and not natural.

Remote Face Database [14] : One of the recent databases in long range face recognition is provided by [14]. It contains unconstrained images from 17 subjects. Images are captured from 5m to 250m distance. This database is very useful for studying unconstrained operational scenario face recognition. However, the small number of subjects and the limited availability of the database makes it difficult for wide use and large scale evaluation.

3 Database design challenges and annotations

Our goal is to provide a large scale database for unconstrained and open set face recognition. We describe and release part 1 of this database with this paper. [1] Part 1 of this database consists of 6,337 images from 308 individuals. We provide identity ground truth labels and facial attributes along with the data We also provide the results on baseline algorithms on clustering and open set identification problems. In this section, we describe the data acquisition process, challenges in designing the database, and the description of the annotations we provide.

For acquiring the images, we use a Canon 7D camera fitted with Sigma 800mm F5.6 EX APO DG HSM lens. The camera is placed inside an office room and is focused on the outdoor sidewalk at 100m distance from the office room, resulting in 18 Megapixels scene images. The camera is programmed to start capturing images at specific time intervals between classes to maximize the number of faces being captured. Images are captured at an interval of 100msec, resulting in around 10 pictures of a person at different focal points, with multiple views and expressions at each particular interval. The chances of the same person appearing in front of the camera the next day at the same interval is high. For example, a student taking Monday-Wednesday classes at 12:30 PM will show up in the camera on almost every Monday and Wednesday. This results in multiple sequences of an individual on multiple days. The images contain various weather conditions such as sunny versus snowy days. They also contain various occlusions such as sunglasses, winter caps, fur jackets, etc., and occlusion due to tree branches, poles, etc. Capturing of images started in February 2012 and is still going on. So far, we have collected more than 50000 scene images and more than 100,000 face chips.

Face detection and bounding box: One of the major challenges with the images captured in the above unconstrained setting is that of face detection. Even state-of-the-art face detection algorithms fail to detect all faces. However, we would like to include easy to difficult face images in our database. In order to maximize the number of faces, we used a software to manually crop the face images from the captured scene images. This task is performed locally in the lab machines with the help of students for two reasons. First is that this task of cropping images needs seriousness and careful attention. Using platforms like Amazon mechanical turk [1] are prone to human negligence. Another reason is that young students who are looking for career direction get the insight of research problems and get motivated. We provide the bounding box coordinates for face detection problems.

Variations in database and annotations: Though the database is captured in unconstrained situations, it is useful to divide the database into multiple partitions based on their difficulty level for evaluating algorithms. Faces captured in a controlled setting usually do not have this issue as the parameters are known during the time of capture. However, we do not know the underlying distribution of challenges in images in the collected database. After cropping the face chips from the scene images, various parameters of the database are annotated locally by students. We provide the following annotations: Occlusion: High, Medium, Low, No; Blur: High, Medium, Low, No; Pose: High, medium Low, No; Illumination: High, Medium, Low

Fiducial points: The challenge of alignment of faces is surmountable as most of the face recognition algorithms depend on face alignment during the preprocessing stage. We provide the ground truth eyes, nose and mouth corners coordinates for alignment. This is equally useful to evaluate alignment algorithms and fiducial point detection algorithms. We again use local machines and resources to extract the ground truth information. We provide the following ground truth fiducial points for alignment problem: Left eye inner corner, left eye outer corner, right eye inner corner, right eye outer corner, nose tip, mouth left corner, mouth right corner.

Identity labels: The most difficult problem we face during designing this database is identity labeling of the images. To make the data useful for verification, pair matching, identification, sequence to sequence matching and identity clustering problems, we need to label the face images with the individual identities. The major problem is that this is not a controlled image capturing setting and we do not have

[1] http://vast.uccs.edu/UCCSfaces

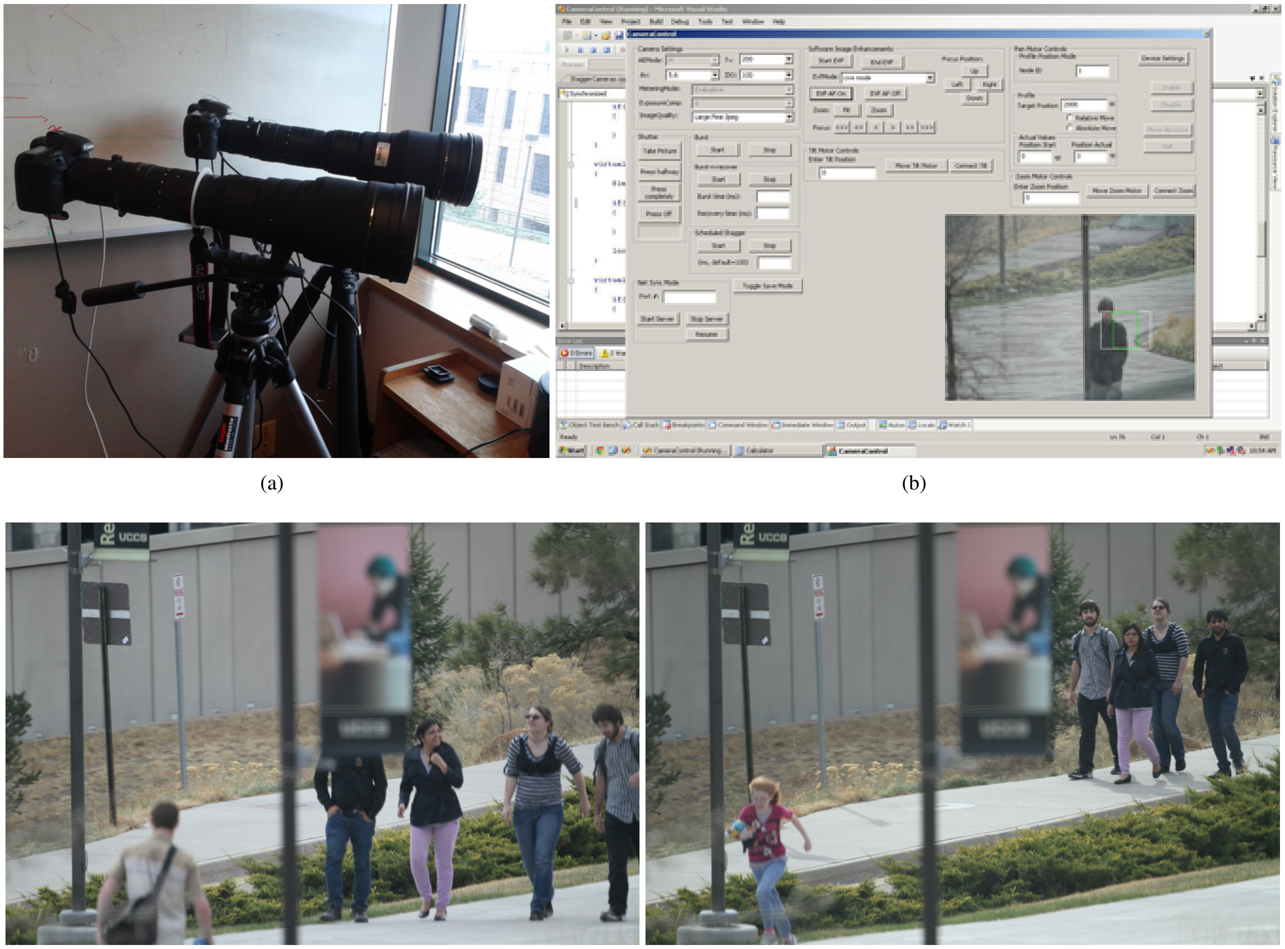
Database	Subjects	Total Images	source	Features
LFW[9]	5749	13233	web	Verification/Pair Matching setup: 1100 match, 1100 non match pairs for training, 500 match and 500 non-match pairs for testing. 10 cross validation sets. Available: eye coordinates, aligned images, attributes (newly available)
PUBFIG[11]	200	60,000	web	Verification setup: 20,000 pairs of 140 subjects. Pose, lighting and expression subsets, divided into easy and difficult
PUBFIG83	200	60,000	web	Verification setup: 20,000 pairs of 140 subjects. Pose, lighting and expression subsets, divided into easy and difficult Available: identity labels
AT&T database (formerly ORL) [19]	40	400	Controlled, lab	Occlusion, illumination variations, expressions. Identification protocol Available: identity labels
FRGC [15]	>466	>50000	controlled and uncontrolled	illumination, expression, background variations, 6 experiments setup with varied number of samples
Scface[7]	130	4160	lab, surveillance cameras	Identification setup: Multiple resolution and various quality surveillance cameras used. Images captured at 1m, 2.6m, 4.2m distance. Verification setup possible. Available: identity labels, fiducial points, imaging conditions
FERET [17]	1199	14126	semi-controlled lab setup	4 different identification subsets. Available: eye coordinates, image labels
GBU	437	1085	uncontrolled illumination indoor and outdoor	Good, Bad and Ugly protocols, verification protocol
LDHF-DB [13]	100	800	Indoor and outdoor settings	Cross distance face verification: Frontal images captured at 60m, 100, and 150m distance. Gallery images captured indoor at 1m.
Remote Face Database [14]	17	2106	Unconstrained, very close to our setup, outdoor	Long range face images, illumination, blur, occlusion variations. Identification protocol

Table 1. Summary of related existing databases.

information about who appears in front of the camera. Some people might appear in front of camera multiple times a day and multiple times during the entire capture period. This task is extremely human intensive if done using face image by image comparision. We would also like to create an open set face database. We therefore need to make sure that everyone who is in the gallery is well labeled and there is no one in the open set non-labeled person who is in the gallery. To conduct this task, we use Picasa [3] because of its easy interface to tag people. When images are loaded, Picasa first clusters images with similarities, and, as people are tagged, it suggests new images to be added into the same group. We observed that Picasa was good in clustering the images of an individual on several small clusters; however, it was unable to cluster or suggest correct tags for all the images of the same person. Even with the help of Picasa interface, the image annotation task is very error prone and it needs careful attention and a lot of seriousness. We conduct this task of image annotation in the local machines in the lab with the help of students. Some imaging conditions make determining matches so challenging that, sometimes, it is hard for humans to recognize whether the images are from the same person or not. As the number of people in the database increases, one has to manually go through all the earlier annotated images and assign a new tag depending on whether the person is already present or not. We provide the identity labels of the images in the database.

4 Database challenge problems

The database is provided to support a number of challenge problems and we describe some here.



(a) (b)

(c) (d)

Figure 1. Setup for image acquisition and example frames. (a) Canon 7D camera fitted on Sigma 800mm F5.6 EX APO DG HSM lens (b) Screenshot of the GUI of software for capturing images. (c) and (d) example of image frames captured at the distance of approximately 100m and 150m respectively.

Face detection: Face detection is an important entity in the face recognition pipeline. Prior to recognition, it is important that face locations are detected in arbitrary images. This is especially difficult in images captured in an unconstrained setting because the presence of faces and the number of faces, if any, is unknown, unlike the images captured in small experimental lab settings. Also the challenges due to scale, pose, and other variations make this task extremely difficult. For the evaluation of face detection algorithms, we provide the ground truth hand labeled face bounding boxes.

Identity Clustering: Given the images of different individuals, identity clustering refers to clustering the people based on their identities when multiple images of each individuals are available. This is applicable to large scale image annotation, organizing photos in personal photo collection, etc. Clustering can be unsupervised or semi-supervised. We provide the identity labels of face images for unsupervised as well as semi-supervised identity clustering.

Identification: Identification[16] returns an identity of the individual. During an initial enrollment, face images, either with or without additional labels, are stored in the database. During testing, an image of an individual is presented to the system. The system then tells whether the person is already enrolled and with whom she/he matches. This is a 1 to n match problem where n is the number of stored images in the gallery. There are different settings in which identification can be performed:

a. Closed set identification: Closed set identification [16] answers : I know you are enrolled, which one are you? In this setting, the identification of the test subject is unknown, but it is from the known list of subjects. This kind of setup is not very common in other applications of face recognition.

b. Open set identification: Open set identification answers: Do I know you? If so, who are you? This is the most realistic setting in face recognition. During biometric enrollment, the system should correctly reject the individuals

Figure 2. Examples of cropped faces from the frames in Figure 1. These images show the variations in images of an individual on a sequence of images captured at a particular time.

which are in the system and enroll the individuals who are not in the system. For this, each time an individual presents to the system, the face image should be matched to all the individuals in the gallery. Similarly, in surveillance systems, the individual in the camera may or may not be in the face database. In this situation, the system has to correctly reject if the person in not in the gallery and correctly identify if the person is in the gallery. Open set identification can also be used for background searches.

c. Automated re-identification: Given a sequence of images from day1 or time1, the problem of automated reidentification is finding the sequence of images that are at a different time or different day. One way of looking at this problem is clustering followed by identification of clusters.

Pair matching: Given a pair of images, the problem is to find whether the images are from the same person or not. This protocol is similar to [9] and is intended to evaluate the algorithms that have been previously designed for this setup.

Verification: Verification answers the question Is the person who she/he claims to be? A system protected through face biometrics uses the verification method to authenticate an individual. During original enrollment, an individuals face image and/or additional identity information is stored. During verification, this individual appears in front of the system and tries to authenticate himself/herself with face image. If the match score of the image of this individual with the stored face image is above an operating threshold, this individual is authenticated. Verification is a one-to-one match problem. Verification is applied in secure facility access, cyber security, authentication in home and personal devices, etc.

Face sequence to sequence matching: When a large number of images are available from each individual, it is often convenient and useful to compare the sequence of face images. Similar to pair matching, this will answer the question "Are these two sequences of images are from the same person or not?"

5 Baseline algorithms

There are several standard face recognition algorithms for well known face recognition problems such as closed set identification, verification, and pair matching. The baseline for the well known problems will be posted in the database website. In this paper, we provide the baseline for less obvious problems, such as identity clustering and open set recognition.

5.1 Clustering algorithm

We provide a baseline algorithm for identity clustering. One of the major challenges in designing this database is image identity annotation. Since the subjects walking in front of the camera are from an unknown pool of people, it is difficult to label the face images with the identity. Moreover, one person can appear in front of the camera at multiple times of the day and on multiple days. We use Picasa iteratively and manually, verifying the tagged images several times. Picasa was good at clustering the images of an individual on the same day; however, it was not able to cluster the same individual on multiple days automatically without human interaction. Once we label images using Picasa iteratively, we conduct the following baseline experiment to evaluate the performance of clustering.

KMeans with random seeding: We use the widely popular KMeans [10] clustering algorithm for identity clustering. KMeans is an unsupervised clustering algorithm that partitions the data into K different partitions, iteratively minimizing the distance measure between the data points and the cluster center points. The initial cluster centers of KMeans algorithms are determined by using values generated by a random seed. For evaluation of the clustering algorithm, we use 1800 images from 180 subjects with 10 images each and with varied number of K. We follow 5-fold cross validation scheme where in each set of experiment, we use different image samples from the subjects.

Algorithm 1 gives an overview of KMeans algorithm we use.

Figure 3 shows the F measure of KMeans clustering for different values of K. F Measure has been widely used in biometrics and computer vision application. It has also been used for evaluation of clustering algorithms [5, 12]. We use the following definition of F measure.

$$\text{F measure} = \frac{(\beta^2 + 1) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}} \qquad (1)$$

We use a pair wise comparison of the samples for computing true positive, true negative, false positive and false

Algorithm 1 K-Means algorithm with random seeding

Input: $X = \{(x_1, x_2, ...x_n)\}$, $x_i \in R^d$, Number of clusters: K, Seed for random number generator
Output: K partitions of the data X such that KMeans objective function is optimized.
1. Initialize: K clusters centers with random seed.
2. Run until convergence: RLD>thresh
2a.Initial assignment of data to K clusters.
2b.Compute new cluster centers
2c.Compute relative distortion loss (RDL):
$RLD = \frac{\text{Initial Distortion - Current Distortion}}{\text{Initial Distortion}}$

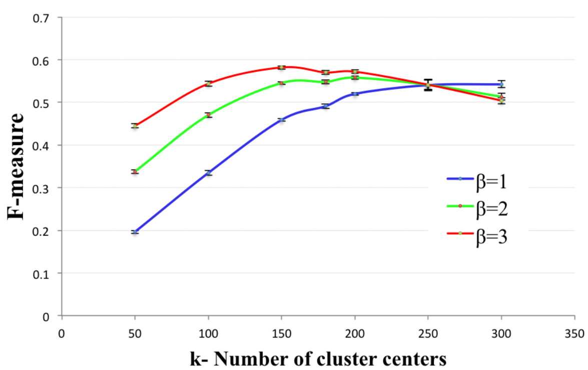


Figure 3. F Measure of KMeans clustering algorithm on our database. F Measure increases as the value of K approaches the actual number of classes in a database. However, as the value of β increases, meaning we give higher penalty for recall, the F Measure increases initially but decreases eventually.

negative. A true positive is when two images from same person fall within a cluster while a true negative is when a pair of dissimilar images fall in different clusters. Similarly, a false positive is when two images from different people are clustered together, and a false negative is when a pair of images from same person fall in different clusters.

Figure 3 shows the interesting results. When the size of K, number of clusters, is less than the actual clusters in the data, F measure, with $\beta > 1$ is better than with $\beta = 1$. But eventually when K is more than the actual number of clusters in the data, F measure with $\beta > 1$ decreases. β gives more weight to recall, which means it penalizes false negatives more than false positives. When K is less, the chances of two images from same person falling under different clusters is low. But, when K increases, the chances of two images from same person falling under different clusters is high resulting in low recall. And since $\beta > 1$ gives more weight to recall, the F measure drops. There is a tradeoff in choosing the β value and the cluster size.

5.2 Open set Identification

We propose and use the following protocol for open set identification. Table 5.2 shows the number of classes used

Training Classes	Target Classes	Test Classes	Openness	Accuracy
180	180	180	0.0%	0.78 ± 0.0053
180	180	308	14.11%	0.62 ± 0.0043
150	150	308	19.06%	0.51 ± 0.0034
100	100	308	29.98%	0.35 ± 0.0012
50	50	308	47.14%	0.15 ± 0.0009
25	25	308	61.23%	0.08 ± 0.0011

Table 2. Open Set identification using face attributes and Support Vector Machine. For the training, subjects which have 15 or more images are selected. 10 images are used for training and the remaining 5 or more images from each subjects are used to create closed set testing. Images from the subjects which have less than 15 images are used to construct openset testing set along with the closed set testing images. As shown in the table as the openness increases, the accuracy decreases.

for the experiments and the accuracy. For the closed set experiments, we select the subjects which have 15 or more images. We use 10 images from each subject for training and remaining samples from the same subjects for testing. For open set experiments, we use the same images as closed set for training but for testing set we include the images from remaining subjects which contain less than 15 images. The size of training set in terms of number of images is 1800 and size of closed set testing is 3635 whereas the size of open set testing is 4537. To design the experiments with different openness we reduce the subjects in training set and keep the same testing set. Again, we follow 5-fold cross validation scheme for each experiment.

We evaluate the performance of one of the most popular and best performing attributes-based recognition system [11]. 73 face attributes are extracted from each face images. These attributes form the feature vector for a particular images. We then use linear SVM [6] with probability output for classification. We take the following definition of openness as mentioned in [20] for our experiments:

$$\text{openness} = 1 - \sqrt{\frac{2 \times |\text{training classes}|}{|\text{testing classes}| + |\text{target classes}|}} \quad (2)$$

While we have shown results for just attributes, we expect other research to use this open set paradigm to evaluate other types of representations/features. The release dataset will include these partitions as well as the actual attribute scores. Is has also been aruged in [20] that accuracy may not be a meaningful measure for open-set recognition. Researchers are also encouraged to explore more meaningful measures for evaluation.

6 Conclusions

We propose an unconstrained face database to help evaluate face recognition algorithms in operational scenarios. The unconstrained and open set nature of this database adds additional challenges to the existing databases for face recognition. Results show that a very good face recognition algorithm also tends to perform low when tested with these images on open set scenarios. Our future goal is to release the database of around 100,000 images from almost 1,000 individuals.

7 Acknowledgements

The authors thank the financial support of ONR MURI N00014-08-1-0638, Army SBIR W15P7T-12-C-A210, ONR STTR N000014-07-M-0421 and SOCOM SBIR H92222-07-P-0020. We also acknowledge the work and help on the paper from Lohitha Ayyanar, Tanya Stere, Jon Parris, Lalit Jain, Michael Wilber, Vijay Iyer, Brian Parks and all the students in VAST lab who helped in various stages in data collection and organization.

References

[1] Amazon Mechanical Turk. https://www.mturk.com/mturk/welcome. 3

[2] Face Recognition Homepage. http://www.face-rec.org/databases. 1, 2

[3] Google Picasa. http://picasa.google.com/. 4

[4] J. R. Barr, K. W. Bowyer, and P. J. Flynn. Detecting questionable observers using face track clustering. In *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*, pages 182–189. IEEE, 2011. 1

[5] F. Beil, M. Ester, and X. Xu. Frequent term-based text clustering. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 436–442. ACM, 2002. 6

[6] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. 7

[7] M. Grgic, K. Delac, and S. Grgic. Scface — surveillance cameras face database. *Multimedia Tools Appl.*, 51(3):863–879, Feb. 2011. 3, 4

[8] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image and Vision Computing*, 28(5):807 – 813, 2010. ¡ce:title¿Best of Automatic Face and Gesture Recognition 2008¡/ce:title¿. 3

[9] 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, University of Massachusetts, Amherst, October 2007. 2, 4, 6

[10] T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Wu. An efficient k-means clustering algorithm: analysis and implementation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):881–892, 2002. 6

[11] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *In IEEE International Conference on Computer Vision (ICCV*, 2009. 2, 4, 7

[12] B. Larsen and C. Aone. Fast and effective text mining using linear-time document clustering. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 16–22. ACM, 1999. 6

[13] H. Maeng, S. Liao, D. Kang, S.-W. Lee, and A. Jain. Nighttime face recognition at long distance: Cross-distance and cross-spectral matching. In K. Lee, Y. Matsushita, J. Rehg, and Z. Hu, editors, *Computer Vision ACCV 2012*, volume 7725 of *Lecture Notes in Computer Science*, pages 708–721. Springer Berlin Heidelberg, 2013. 4

[14] J. Ni and R. Chellappa. Evaluation of state-of-the-art algorithms for remote face recognition. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1581–1584. IEEE, 2010. 3, 4

[15] P. Phillips, K. Bowyer, T. Scruggs, E. Ortiz, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 947–954 vol. 1, 2005. 4

[16] P. Phillips, P. Grother, and R. Micheals. Evaluation Methods on Face Recognition. In A. Jain and S. Li, editors, *Handbook of Face Recognition*, pages 329–348. Springer, 2005. 5

[17] P. J. Phillips, P. J. Phillips, S. Z. Der, P. J. Rauss, P. J. Rauss, O. Z. Der, and O. Z. Der. Feret (face recognition technology) recognition algorithm development and test results, 1996. 4

[18] N. Pinto, J. DiCarlo, and D. Cox. How far can you get with a modern face recognition test set using only simple features? In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2591 –2598, june 2009. 2

[19] F. S. Samaria and A. C. Harter. Parameterisation of a stochastic model for human face identification. In *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*, pages 138–142, 1994. 4

[20] W. Scheirer, A. Rocha, A. Sapkota, and T. Boult. Towards open set recognition, 2012. 1, 7

[21] Y. Tian, W. Liu, R. Xiao, F. Wen, and X. Tang. A face annotation framework with partial clustering and interactive labeling. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1 –8, june 2007. 1