Efficient face retrieval using synecdoches

Brendan F. Klare Austin Blanton Benjamin Klein

Noblis
Falls Church, VA, U.S.A.

Abstract

The number of images and videos available for search on the Internet is in the order of a trillion images, making current brute force search techniques prohibitively inefficient on such large scales. As society continues to increase our desire and ability to search these vast collections of data, improving upon traditional face recognition search techniques becomes an important problem to address. Because face recognition (and other biometric) algorithms are only commercially available as black box systems, any indexing scheme developed to perform efficient search must operate without access to the underlying feature vectors used to measure facial similarity. To address this restriction, we propose a structured search that separates the facial feature space into clusters derived from sets of prototype subjects we refer to as "synecdoches". After an offline training step, our proposed method assigns each gallery image to a cluster in the face space based on its similarity to a set of synecdoche clusters. In turn, query images are compared to the target gallery images based on the closest synecdoche cluster in sequence. Our results show a minimal drop in accuracy when only considering half of the clusters, thus reducing the search space in half. Additional experiments demonstrate the viability of our proposed approach to improve search efficiency amidst the common restriction of a black box matcher.

1. Introduction

The number of images that exist on the web is daunting. For example, a single website, Instagram, reports having twenty billion photos, with users uploading an average of sixty million photos every day. At the same time, there is an increasing need to effectively search such images using information other than manually labelled text. Consider systems that can search a known criminal's social networks for his associates to aid in an investigation, find all images of a given person to assist individuals in managing online identities, or organize photo collections based on image similarity or the people present in the photos. Such scenarios are not yet operationally feasible, but are of increasing demand.

In order to perform efficient retrieval on the vast quantities of data, search methods that do not perform a brute force comparison to all available face templates are needed. Such efficient search methods maximize hardware efficiency while minimizing search times. In order to generalize to common applications, these methods must be compatible with black box algorithms where the feature vector representations are not readily available. Ideally, efficient search methods will minimize search time without a drop in accuracy; however, this is generally unavoidable while searching a reduced feature space.

We propose a method for efficient search using clusters developed offline from a small subset of the similarity search space. We call the small subset a "synecdoche", a figure of speech where a word or phrase is representative of a larger whole. We use this term to indicate that the synecdoche set is indicative of a much larger space that shares similar patterns. The clusters are formed using the self-similarity matrix of the synecdoche templates. Target (or gallery) images are assigned to a cluster based on their similarity to the synecdoche set. During enrollment, query (or probe) images are compared to the synecdoche set, yielding a list of the closest clusters. In turn, the query can be compared to each cluster from closest to farthest until a match is found, with the option of stopping the search process before comparing to all gallery instances as the likelihood a mated image is in one of the more distant clusters is low. Our results show that the search space can indeed be pruned significantly with only a small decrease in the accuracy.

The most critical aspect of our proposed method is that it operates without access to the features, which is amenable for use with commercial black-box systems. This property is paramount as the providers of face recognition algorithms (e.g., commercial SDK vendors) and the integrators of face recognition algorithms are generally not the same, thus limiting access to the system as a black box matcher. This restriction is not likely to change as commercial vendors need to restrict access to the underlying feature vectors to main-

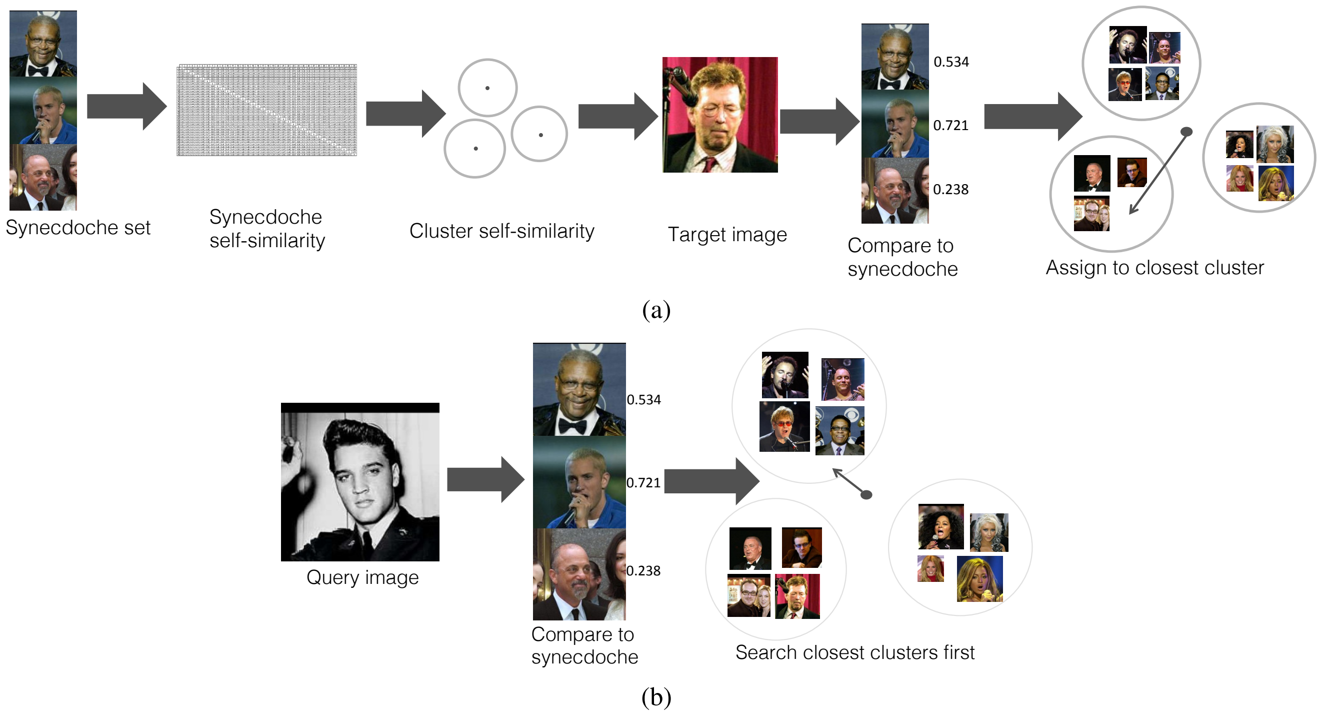


Figure 1. Shown is an overview of the proposed method to perform efficient search when restricted to a black box face recognition algorithm. (a) Using a self-similarity matrix from a set of prototype face images (called the "synecdoche" set), the implied feature space is clustered using the self similarity matrix. The cluster centers of the matrix are then used to partition the database by comparing each database image to the synecdoche set, and using the vector of similarities to assign that image to the gallery corresponding to the closest cluster center. (b) When a query is performed, the template from the query image is compared against the synecdoche set. The resultant similarity vector is then used to search the galleries in order of distance to the cluster center. Experimental results indicate that only a portion of these galleries need to be searched with minimal loss in accuracy.

tain biometric security via encryption [7], maintain privacy, and protect their intellectual property. Thus, while a partitioning of the feature space may be more accurate, the generality of working within the similarity space makes our proposed approach usable within any system.

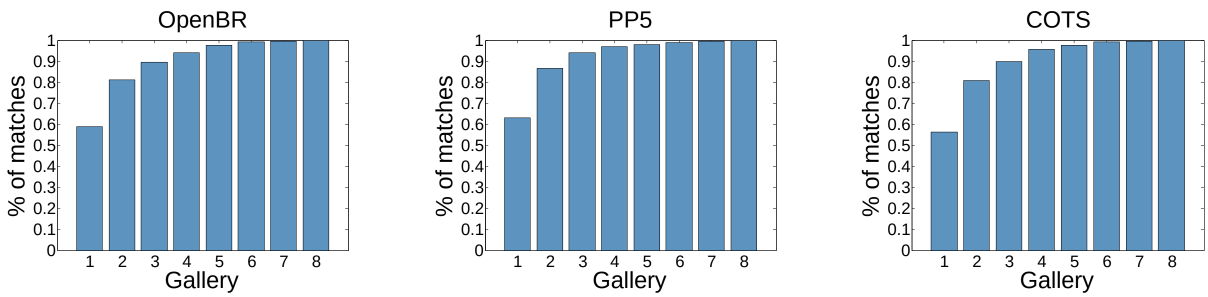
2. Related research

Indexing methods for efficient retrieval in pattern recognition systems is a well studied subject. These approaches can be dichotomized into feature-based and metric-based indexing schemes. Feature-based schemes are generally more accurate as they assume access to the underlying feature vector representations used to measure the similarity between two objects. Metric-based indexing is constrained to only operate on the similarity (or distances) between two objects, and thus must use the similarity information to infer the distribution of the underlying feature space.

Feature-based indexing is not relevant to this work, as we are motivated by real world biometrics applications where we do not have access to the underlying feature vectors. However, readers who are interested in this topic are encouraged to read Datar *et al.*'s seminal work on locality sensitive hashing (LSH) [1], as well as other influential works on this topic [9, 6]. It is notable that while feature-based methods have access to the underlying feature vector representation, these methods for efficient search still come at a cost of search accuracy. While methods such as LSH minimize any loss in recall, the methods are still fundamentally constrained by noise in the data and the curse of dimensionality when operating on high dimensional datasets. As such, these methods will always be imperfect and instead strive for a probabilistic underpinning of the likelihood that indexing was performed successfully.

Metric-based indexing is a special case of feature-based indexing where only the measured similarities or distances can be used to partition the target database. Thus, given that feature-based indexing is rife with difficulties, metric-based indexing is essentially bounded by the success of the more informed feature-based methods. However, despite this pessimistic reality, many operational scenarios dictate the use of metric-based indexing. The reasons vary, but most commonly this is due to either a proprietary feature vector representation, or template security requirements to maintain security [7].

A comprehensive survey of metric-based indexing is provided by Hjaltason and Samet [4]. Most of these ap-



(a) (b) (c)

Figure 2. The cumulative percentage of correct identifications as progressively more of the galleries are searched. Results are from the PCSO dataset using the spectral clustering method. With all three matchers, searching only half of the galleries will result an comparison to the mated image over 90%. The failure cases generally occur when the probe and gallery mate would not yield a successful match anyways.

proaches operate under the assumption of the triangle inequality as the term "metric" is used strictly. Contrarily, when indexing black box face recognition algorithms, the assumption of the triangle inequality rarely holds. In fact, even more confounding, we have found that certain face recognition algorithms do not even satisfy the symmetry property of a distance metric. As such, score-level indexing schemes are required that are robust to the noise present in biometrics algorithms, and are designed with the very high number of classes (i.e., subjects) used in biometric systems.

Searching the biometrics literature for relevant methods in search space indexing leads to methods that focus on the feature-based case. For example, Feng and Jain proposed a method for performing indexing using fingerprint features [3], which was expanded on by Paulino *et al.* [13]. Other feature-based indexing methods have been proposed for iris recognition [14], hand geometry matching [11], and multi-biometrics systems [8].

Our review of relevant literature did not yield any face (or biometric) recognition methods for performing indexing amid the constraint of similarity score information. Thus, we are providing information regarding our initial study of the problem. The solutions proposed in this paper target the common scenario where a proprietary recognition algorithm is used to perform retrieval searches on a large database. Based on our operational applications of interest, experiments conducted in this paper focus on face recognition. However, it should generalize to most any biometric modality that involves searching across a large population.

3. Gallery Partitioning using Synecdoches

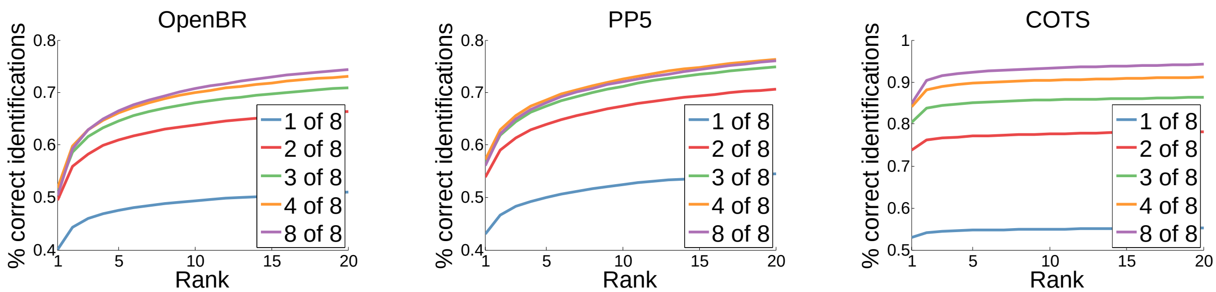
3.1. Training

Our proposed method is premised on the ability to infer clusters of similar subjects from facial similarities. This is achieved by using a set of prototype subjects that ideally span the distribution of the facial similarity space. Referred to as "synecdoches," the similarity of a newly presented image to the synedoche set is the basis for our partitioning of a large gallery database. The intuition is that with meaningful, distinct clusters from a representative synecdoche, the similarity to a large collection can be estimated very rapidly using only similarity values. This estimation allows systems to intelligently partition the search space and possibly find a suitable match in a much shorter period of time.

The first step in the proposed algorithm is to develop a set of n^s images to serve as our synecdoche set. Given the set of synecdoche images x_i^s, where $i = 1 \dots n^s$, and a black box face recognition algorithm $y(x_1, x_2) = s$ (where s is the resultant simmilarity between two face images x_1 and x_2) we generate the self similarity matrix $S^s \in \mathbb{R}^{n^s, n^s}$ between all the images. Similar to the kernel trick in support vector machines, here we are using the columns of the self similarity matrix to serve as feature vectors that infer our higher dimensional space.

Using the set of similarity vectors, the synecdoche images are clustered into k groups, which in turn yields a set of k cluster centers $c_k \in \mathbb{R}^{n_s}$. We explored the use of three different clustering algorithms in the work: k-means [2], k-medoids [12], and spectral clustering [15]. Despite what clustering algorithm is used, the outcome is the same: the set of cluster centers c_k. Each cluster center acts as the presumed center for a location in the face space of some underlying matcher. Thus, for a large database of face images, each cluster corresponds to a sub-gallery where images that fall into the cluster can be assigned.

Given the cluster centers and the set of synecdoche images, we can proceed to partition our database to facilitate a more efficient search. Given n^g gallery images $x_i^g, i = 1 \dots n^g$, we measure the similarity between $f(x_i^g, x_j^s)$ between the i-th gallery image and all $j = 1 \dots n^s$ synecdoche images, resulting in the vector of synecdoche similarities $l(x_i^g) \in \mathbb{R}^{n_s}$. In turn, we measure the distance between $l(x_i^g)$ and each cluster center c_k and assign the image



(a) (b) (c)

Figure 3. CMC curves for each of the three matchers on the PCSO dataset using the spectral clustering algorithm. All three matchers report CMC curves for searching half of the galleries that are at most only slightly worse than searching all galleries. At low ranks, OpenBR and PP5 actually report a higher percentage of correct identifications when searching four and three of the galleries than searching all galleries. See Figure 4 for details on this phenomenon.

to the gallery k corresponding to the closest cluster. Note that gallery enrollment is not exactly a training step, but it is instead placed in this section as it can occur offline to batch enroll an existing gallery.

3.2. Enrollment and search

When a new face image x is enrolled into the gallery database, the process remains the same as in the offline step: the similarity of the image is measured against the synecdoche set (i.e., we generate vector $l(x)$)), and the image is compared against each cluster center c_k to assign the image to the gallery k corresponding to the closest cluster center.

When a query face image x^q is provided, we measure the similarity between the image and the synecdoche set to yield the vector of similarities $l(x^q)$. However, when comparing the probe similarity vector to the cluster centers, we instead maintain a sorted list of the distance to each sub-gallery (i.e., cluster). Thus, when querying the database, the galleries can be searched in order of closest to farthest. Based on operational considerations, the search process could search only a portion of these galleries. Thus, in the case of $k = 10$ clusters, if we only search the galleries corresponding to the two closest clusters, then we remove roughly 80% of our face comparisons.

4. Experiments

Several experiments were conducted to measure the efficacy of the proposed approach. The overall intent of these experiments is to measure the trade off between the percentage of the database searched, and what (if any) error we incur by not searching portions of the database. Unless noted otherwise, all results are presented using the spectral clustering algorithm.

4.1. Datasets and Matchers

Two different datasets are used in this study. The first dataset is the Labeled Faces in the Wild (LFW) database [5], which is a smaller scale database but allows us to understand the effects of unconstrained imagery typically encountered in internet-sourced media. Our experiments on LFW use 4068 images in the synecdoche set, and 1680 images from non-overlapping subjects in the operational set.

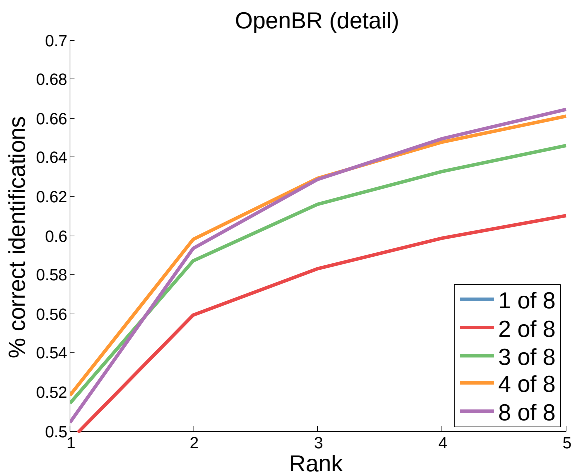
The second dataset is a mug shot database from the Pinellas County Sherrif's Office (PCSO). These are constrained frontal images from alleged criminal offenders. Experiments on the PCSO database used 4,385 images in the synecdoche set, and 50,000 subjects in the operational set.

The operational set for both datasets contained one probe image and one target image per subject. For both datasets, no subject in the operational set was in the synecdoche set (i.e., the synecdoche and operational datasets were non-overlapping). As described in Section 3, the target images are assigned to a gallery based on the closest cluster center c_k. The probe images will search these galleries in order of closest to farthest.

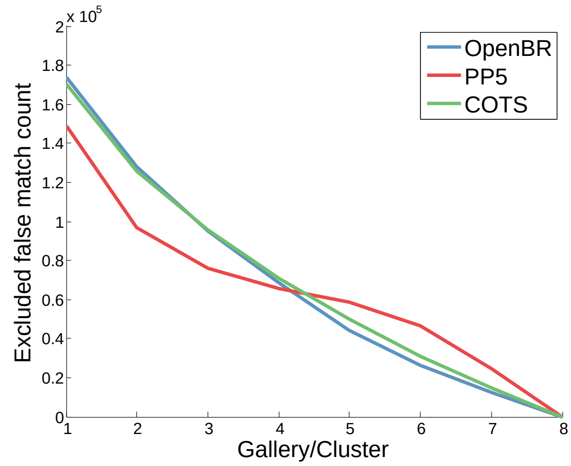
We used three different face recognition algorithms in this study. The first is the open source matcher, OpenBR [10]. The second matcher is version 5.2 of the PittPatt face recognition algorithm (labelled as PP5). Finally, an anonymous, and state of the art, commercial off the shelf (COTS) face recognition algorithm was used.

4.2. Results

Figure 2 shows a cumulative sum of the percentage of true matches that fall within the k-th gallery searched for the PCSO dataset and the spectral clustering algorithm. In this experiment, if the assignment to galleries were random, the first gallery would only contain $1/k$ of the matches, assuming targets distributed evenly among galleries. Instead, the clusters are successfully inferring the structure of the feature space using the similarity space. For example, when searching only half the database, all three matchers are able to search the gallery corresponding to their mate over 90% of the time. This is indeed impressive given that the major-



(a)



(b)

Figure 4. (a) Detail of Figure 3(a) shows that at low ranks, searching only half the galleries actually outperforms searching all galleries. (b) The count of excluded false matches decreases as more galleries are added. This graph explains the effect in Figure 4 where a search through less images gives a higher percentage of correct identifications than by searching the full set.

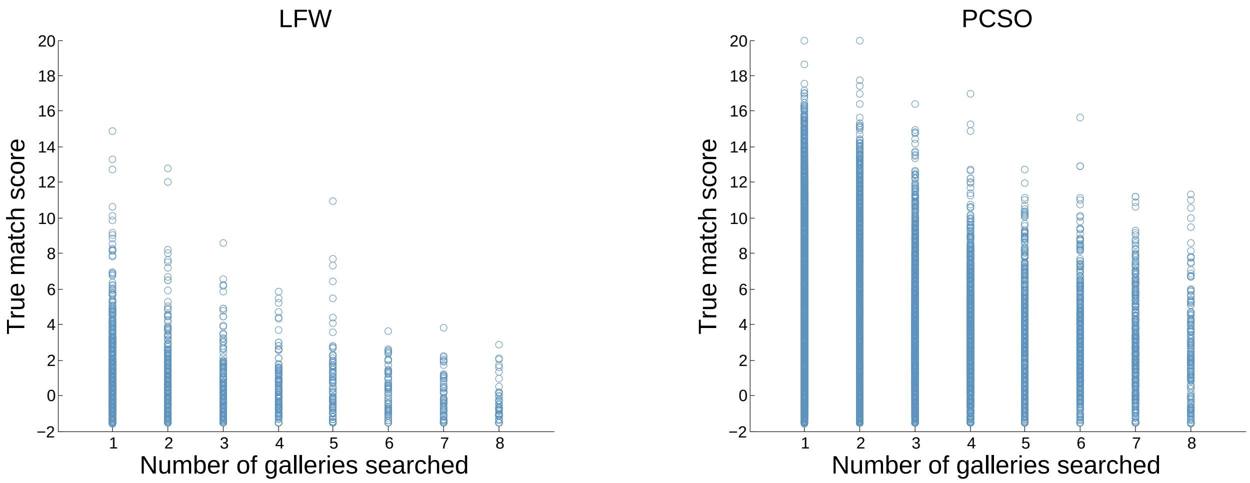
ity of the cases where the mate is not searched, a successful match would not have been achieved anyways (as we subsequently show). While a perfect algorithm would show almost all the matches within the first gallery, this elusive achievement is not currently possible using even feature-based indexing methods.

Next, we measure the impact on recognition accuracy using the proposed approach. Shown in in Figure 3 are the CMC plots for all three matchers on the PCSO database. While only searching the first gallery has a significant impact on the percentage of correctly identified subjects, there is little to no decrease in recognition accuracy when searching half the galleries. That is, given 50,000 images in the gallery, we can only search 25,000 of the images with a negligible change in accuracy using multiple different matchers.

Interestingly, in some cases recognition accuracy actually improves at low ranks for two of the matchers when only searching half the galleries (see Figure 4(a)). This is because the galleries that are being excluded contain false matches with higher scores than the true match. Figure 4(b) shows how a count of false match scores at low ranks are excluded in a sequential gallery search. Despite better performance, this is not preferable in general; it would be better if the highest match scores to a given image were all in the first cluster, regardless of the correctness. However, the COTS matcher often does not exhibit the same property. We postulate matchers that are more accurate overall will show more consistent clusters, with high match scores in the first galleries. Thus, while our proposal is not dependent on a specific algorithm and independent of feature representation, it still depends on the accuracy of the underlying algorithm. Regardless, searching half the space with only modest losses in accuracy is very desirable in a situation where the number of images to search is very large and speed is a priority. Indeed, such scenarios will be increasingly common as available image resources and demand for services continue to grow.

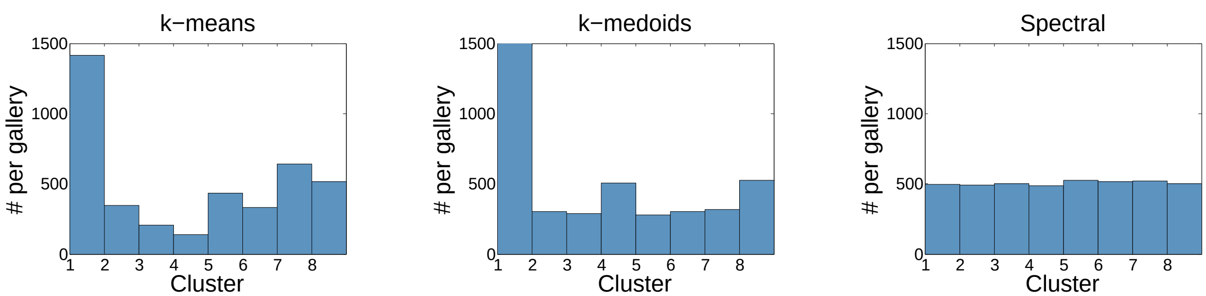
Another way to view the efficacy of the approach is to consider the match scores of true matches against the number of galleries searched before finding the true match. The ideal curve would show a negative linear relationship; high match scores would be contained in the first rank gallery and low scores in the last. This would confirm the intuition that lower scoring true matches are simply hard examples and the underlying algorithms are not capturing the similarity well enough to assign a gallery accurately. We indeed see that in the results when examining LFW using PP5 (Figure 5a) (similar results were observed for other matchers). Note that while low-scoring true matches are scattered throughout, one can still detect a linear trend; high matches are much less numerous in the higher-rank galleries. Such a trend is also apparent in PCSO (Figure 5b), where the data points are denser and middle scores seemingly uniform distributed among the rank galleries.

We conducted additional experiments to examine the effect of clustering algorithm on the partitioning of the search space. Spectral clustering was chosen over k-means and k-medoids clustering because the clusters tend to be more evenly divided (see Figure 6). Evenly divided clusters in the synecdoche set would indicate evenly assigned target images to galleries. This is good because if a few galleries contain the majority of the dataset, searching those galleries will be costly and the overall runtime savings meager.



(a) (b)

Figure 5. True match scores for each image versus in which gallery/cluster rank they were found (using PP5). There is a noticeable linear relationship; higher true match scores are often found in the first gallery searched.



(a) (b) (c)

Figure 6. The distribution of assignments to clusters in the LFW PP5 synecdoche set using three different clustering methods. Note that spectral clustering exhibits the most evenly distributed clustering. This is a good property for indexing purposes, as gallery assignments divide the search space evenly.

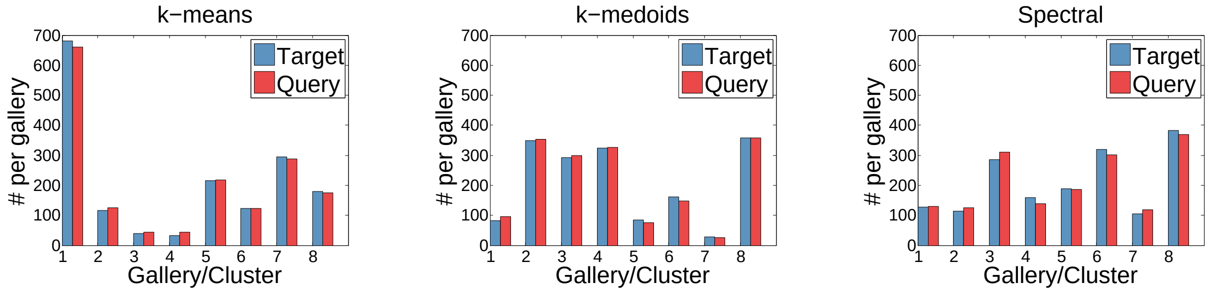
While k-medoids is fast, it still exhibits the uneven assignment of k-means, which is unacceptable. We see in Figure 7 that the cluster assignments follow the general trend of the distribution of the synecdoche set for each clustering algorithm, as expected. The assignments using the k-means clusters are heavily imbalanced to a particular cluster while spectral clustering exhibits a more even distribution. Having balanced and yet distinct clusters will enable retrieval to be both efficient and accurate.

A critique of this approach is that it is difficult to ensure that the synecdoche set indeed follows the distribution of the intended operational data. However, this is an unfortunate reality of training-based pattern recognition methods in general. For example, the vast majority of face recognition algorithms are developed using training data. The implications of non-representative training data are the same for these algorithms as our proposed method. In both cases, the use of empirical experiments are paramount to demonstrating the practicality of a given method, as is the case in this work.

5. Conclusions

In this paper we proposed a method for efficient indexing by partitioning the similarity search space into clusters that are inferred using only black box face recognition algorithms. Experiments demonstrated that without access to the underlying features, we could still reduce our search space in half with little to no decrease in recognition accuracy.

Given such promising results, and the confounding operational use case motivating our study, we will continue to improve the proposed method. For example, one way to enhance this approach might be a cluster center representation that is more indicative of the full diversity of data points within the cluster. That is, instead of using the center point of a cluster, a subset of cluster instances could be used to measure the distance of a newly seen instance to



(a) (b) (c)

Figure 7. The distribution of assignments to clusters for both target and query using three different clustering methods on the LFW PP5 similarity scores. The k-means plot shows one cluster that dominates the assignments. This is not ideal; if one gallery contains most of the subjects, there is only a small difference between using the proposed method and searching the entire space. Both k-medoids and spectral clustering exhibit more evenly distributed clusters, but k-medoids still has several clusters that are noticeably larger than average and others that are noticeably smaller.

the cluster. In cases of large within-cluster variance, this scheme would presumably alleviate inaccurate cluster assignment. Another area of inquiry is the use of hierarchical clusters that separate the search space into smaller chunks. Such a system would traverse the hierarchy, with a depth first search into the closest clusters until it reaches a cluster with no sub-clusters. The query would be compared to images in each leaf sub-cluster it traverses until a suitable match is found.

Finally, given the vast number of images cited in the introduction, this study is on a relatively small scale. Yet, even in this modest approximation of an enormous data collection, there were still considerable obstacles in terms of both speed and memory. Such limitations will be mitigated as we continue to operationalize the proposed method, allowing us to conduct larger scale studies at greater efficiency.

Acknowledgement

This research was supported by the Noblis Sponsored Research program.

References

[1] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni. Locality-sensitive hashing scheme based on p-stable distributions. In *Proceedings of the 20th Annual Symposium on Computational geometry*, pages 253–262. ACM, 2004.

[2] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. John Wiley & Sons, 2012.

[3] J. Feng and A. K. Jain. Filtering large fingerprint database for latent matching. In *International Conference on Pattern Recognition*, pages 1–4, 2008.

[4] G. R. Hjaltason and H. Samet. Index-driven similarity search in metric spaces (survey article). *ACM Transactions on Database Systems (TODS)*, 28(4):517–580, 2003.

[5] 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, Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[6] P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *Proceedings of the 30th annual ACM symposium on Theory of computing*, pages 604–613. ACM, 1998.

[7] A. K. Jain and K. Nandakumar. Biometric authentication: System security and user privacy. *IEEE Computer*, 45(11):87–92, 2012.

[8] U. Jayaraman, S. Prakash, and P. Gupta. Indexing multimodal biometric databases using kd-tree with feature level fusion. In *Information Systems Security*, pages 221–234. Springer, 2008.

[9] J. M. Kleinberg. Two algorithms for nearest-neighbor search in high dimensions. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 599–608. ACM, 1997.

[10] J. C. Klontz, B. F. Klare, S. Klum, A. K. Jain, and M. J. Burge. Open source biometric recognition. In *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, pages 1–8. IEEE, 2013.

[11] A. Mhatre, S. Chikkerur, and V. Govindaraju. Indexing biometric databases using pyramid technique. In *Audio-and Video-Based Biometric Person Authentication*, pages 841–849, 2005.

[12] H.-S. Park and C.-H. Jun. A simple and fast algorithm for k-medoids clustering. *Expert Systems with Applications*, 36(2):3336–3341, 2009.

[13] A. A. Paulino, E. Liu, K. Cao, and A. K. Jain. Latent fingerprint indexing: Fusion of level 1 and level 2 features. In *Proc. IEEE Biometrics: Theory, Applications, and Systems*, pages 1–8, 2013.

[14] C. Rathgeb and A. Uhl. Iris-biometric hash generation for biometric database indexing. In *International Conference on Pattern Recognition*, pages 2848–2851. IEEE, 2010.

[15] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.