

Face Recognition via Archetype Hull Ranking

Yuanjun Xiong[†] Wei Liu[‡] Deli Zhao[†] Xiaoou Tang[†]

[†]Information Engineering Department, The Chinese University of Hong Kong, Hong Kong

[‡]IBM T. J. Watson Research Center, Yorktown Heights, NY, USA

{yjxiong, xtang}@ie.cuhk.edu.hk weiliu@us.ibm.com zhaodeli@gmail.com

Abstract

The archetype hull model is playing an important role in large-scale data analytics and mining, but rarely applied to vision problems. In this paper, we migrate such a geometric model to address face recognition and verification together through proposing a unified archetype hull ranking framework. Upon a scalable graph characterized by a compact set of archetype exemplars whose convex hull encompasses most of the training images, the proposed framework explicitly captures the relevance between any query and the stored archetypes, yielding a rank vector over the archetype hull. The archetype hull ranking is then executed on every block of face images to generate a blockwise similarity measure that is achieved by comparing two different rank vectors with respect to the same archetype hull. After integrating blockwise similarity measurements with learned importance weights, we accomplish a sensible face similarity measure which can support robust and effective face recognition and verification. We evaluate the face similarity measure in terms of experiments performed on three benchmark face databases Multi-PIE, Pubfig83, and LFW, demonstrating its performance superior to the state-of-the-arts.

1. Introduction

The primary purpose of face analysis is to compute a robust and effective similarity measure between any input pair of face images. Such a measure is expected to suppress intra-personal face variations due to varying expressions, poses, and illumination conditions. Nowadays, rapidly growing face image resources stemming from online photo albums as well as social networks provide new opportunities and meanwhile pose new challenges to existing face processing approaches. How can we take advantage of the gigantic amount of face information on the Web? One feasible approach is to upgrade current face processing systems by augmenting web-crawled face images into their training datasets, which therefore requires the face systems to be easy for re-training and scalable to accommodate massive web data.

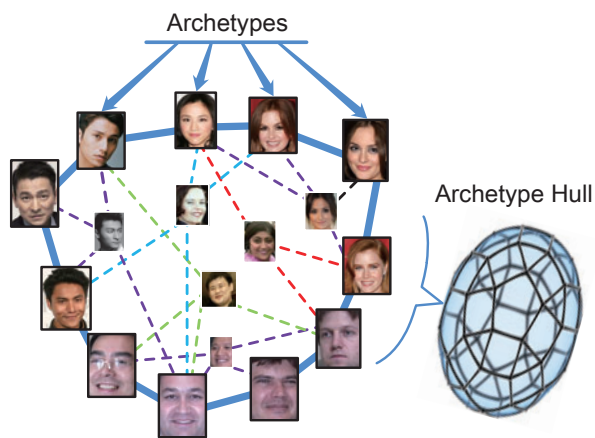


Figure 1. One visual example showcasing an archetype hull of face images. We view face images as points in the image space, where the polytope composed of a few archetype faces encloses almost all points. Any point can be represented by a convex combination of the archetypes, and these archetypes hence form a convex hull of the entire point set.

To pursue the scalability, we leverage a small set of *archetype* exemplars to represent a large training set of face images. These archetypes constitute a convex hull which encompasses most faces in the training set. To the best of our knowledge, the *archetype hull* model has not been applied to the face area. The use of archetypes along with the produced archetype hull may open up a new avenue to enable traditional face processing approaches scale up to massive face datasets. To illustrate, Fig. 1 showcases face archetypes and an archetype hull to model face images. In this paper, we seek such archetypes using an efficient simplex volume maximization algorithm. Subsequently, we build a scalable graph by virtue of the archetypes whose size is much smaller than the training data size. Moreover, we propose a novel graph-based ranking framework which explicitly captures the relevance between any query and the stored archetypes, yielding a rank vector over the archetype hull. The archetype hull ranking is then applied to every block of face images, leading to a blockwise similar-

ity measure between any face pair through comparing two different rank vectors with respect to the same archetype hull. After integrating blockwise similarity measurements with learned importance weights, we eventually achieve a sensible face similarity measure that is readily applicable to both face recognition and verification tasks. We evaluate the face similarity measure in terms of experiments carried out on three benchmark face databases Multi-PIE [7], Pubfig83 [20], and LFW [9], and demonstrate that the performance of the proposed face similarity measure is superior to the state-of-the-arts.

2. Related Work

In the face recognition literature, a large number of subspace methods [25][3][19][17][26][27][13][8][28] working on holistic facial features have been proposed. Recently, local facial descriptors [18][1][5] achieved greater accuracy gains on many benchmark datasets. The local descriptors attempt to extract distinctive features of image textures like SIFT [18] or local micro-patterns of face shapes like LBP [1]. However, intra-personal variations caused by varying expressions, poses, and illumination conditions remain a potential obstacle to these appearance-based methods.

The success of sparse representation [30] in face recognition has inspired the face community to consider the sparse properties of the subspaces hidden in face image ensembles. Nevertheless, the subspace sparsity holds if there are sufficient face examples in each class to cover intra-personal variations, which could not be satisfied under unconstrained recognition environments such as the LFW database.

As mentioned before, one of the major challenges of modern face recognition is the explosive growth of face data. Some efficient learning methods could provide promising solutions. For example, the LARK method [22] developed a special kind of features to obtain a training free classifier. Lately, a scalable neighborhood graph, *Anchor Graph* [14], was proposed to accommodate massive training samples, and has shown excellent performance in large-scale semi-supervised learning [16] and image retrieval [31]. In this paper, we employ the Anchor Graph model to deal with large quantities of face images since it scales linearly with the training set size in terms of both space and time complexities.

3. Archetypes and Archetype Hull

Archetype is initially termed in psychology [11] to describe a very typical and universally understood exemplar among a group of objects. Archetypal analysis aims to find a compact set of “pure types”, *i.e.*, so-called archetypes, such that the typical patterns of this group of objects are covered in the archetypes and the other objects are simply emulations or combinations of the archetypes.

3.1. Concept of Archetypes

The concept of archetypes actually exists in a variety of disciplines including literature, philosophy, psychology, marketing [12], and statistics [6]. The idea of archetypes is recently introduced into pattern recognition [23] and informatics [24][21]. Understanding the concept of archetypes in face recognition is intuitive: people can easily remember someone with a very distinctive facial appearance; some people are thought of being very similar to a few distinctive faces. Such phenomena serve as the evidence of identifying archetypes and correlating unknown faces to known archetypes in a recognition process.

Mathematically, the convex hull of archetypes encompasses almost all points in a data set, and the archetypes inherently reside on the border of the point cloud formed by the data set. To exploit the idea of archetypes along with their produced archetype hull, there are two core problems: 1) how to find the archetypes, and 2) how to correlate one input sample with the archetypes.

3.2. Archetype Seeking

The time efficiency of finding archetypes is an important concern. It is known that for an input data set \mathcal{X} of n points with d dimensions, the time complexity for computing a convex hull enclosing \mathcal{X} is as high as $O(n^{d/2})$ [23][24]. For a set of tens of thousands of images with high-dimensional descriptors, exactly solving this convex hull problem quickly becomes computationally intractable. To this end, a few algorithms have been designed to achieve approximate solutions, of which the simplex volume maximization algorithm [24] can provide a good approximate solution in a linear time complexity $O(n)$. Given a point set $\mathcal{X} = \{\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^n$, we intend to search for m archetypes $\mathcal{U} = \{\mathbf{u}_j \in \mathcal{X}\}_{j=1}^m$ whose convex hull forms an $(m-1)$ -simplex in geometry. Because searching for the archetypes \mathcal{U} is geometrically equivalent to maximizing the volume of the simplex produced by \mathcal{U} , the simplex volume maximization algorithm uses a sequential seeking strategy to find m archetypes such that their produced simplex has the maximum volume.

In distance geometry [4], the volume of an $(m-1)$ -simplex $\mathcal{S}(\mathbf{u}_1, \dots, \mathbf{u}_m)$ is obtained by the *Cayley-Menger determinant*. With a mild assumption that all simplex vertices are equidistant, the Cayley-Menger determinant gives a simplified volume formula as

$$\begin{aligned} Vol(\mathcal{S})_{m-1}^2 = & \frac{a^{2(m-1)}}{2^{m-1}((m-1)!)^2} \left(\frac{2}{a^4} \sum_{j=1}^{m-1} \sum_{j'=j+1}^{m-1} d_{j,m}^2 d_{j',m}^2 + \right. \\ & \left. \frac{2}{a^2} \sum_{j=1}^{m-1} d_{j,m}^2 - \frac{m-2}{a^4} \sum_{j=1}^{m-1} d_{j,m}^4 - (m-2) \right), \end{aligned} \quad (1)$$

Algorithm 1 Archetype Seeking via Simplex Volume Maximization

Input: An input data set $\mathcal{X} = \{\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^n$, the number m ($\ll n$) of required archetypes.

Initialize

randomly sample a point \mathbf{x}_t from \mathcal{X} ,

$$\mathbf{u}_1 = \arg \max_{\mathbf{x}_i \in \mathcal{X}} \|\mathbf{x}_i - \mathbf{x}_t\|,$$

$$a = 0, D_i^{(1)} = D_i^{(2)} = D_i^{(3)} = 0 \forall i \in [1 : n];$$

for $k = 1, \dots, m - 1$ **do**

for $i = 1, \dots, n$ **do**

$$D_i^{(3)} \leftarrow D_i^{(3)} + \|\mathbf{u}_k - \mathbf{x}_i\| D_i^{(1)}$$

$$D_i^{(1)} \leftarrow D_i^{(1)} + \|\mathbf{u}_k - \mathbf{x}_i\|$$

$$D_i^{(2)} \leftarrow D_i^{(2)} + \|\mathbf{u}_k - \mathbf{x}_i\|^2$$

end for,

$$a \leftarrow \max \{a, \{\|\mathbf{u}_k - \mathbf{x}_i\|\}_{i=1}^n\},$$

$$i^* \leftarrow \arg \max_{i \in [1:n]} (a D_i^{(1)} - \frac{k-1}{2} D_i^{(2)} + D_i^{(3)}),$$

$$\mathbf{u}_{k+1} \leftarrow \mathbf{x}_{i^*},$$

end for.

Output: The m archetypes $\mathcal{U} = \{\mathbf{u}_k\}_{k=1}^m$.

where $a > 0$ is the constant distance among m vertices, and $d_{j,m} = \|\mathbf{u}_j - \mathbf{u}_m\|$. Based on the volume formula in eq. (1), the sequential archetype seeking principle is expressed as follows

$$\mathbf{u}_{k+1} = \arg \max_{\mathbf{x}_i \in \mathcal{X}} \sum_{j=1}^k \tilde{d}_{j,i} \left(a_k - \frac{k-1}{2} \sum_{j=1}^k \tilde{d}_{j,i} + \sum_{j'=j+1}^k \tilde{d}_{j',i} \right), \quad (2)$$

where $k \in [1 : m - 1]$, $a_k = \max_{\mathbf{x}_i \in \mathcal{X}} \max_{k' \in [1:k]} \|\mathbf{x}_i - \mathbf{u}_{k'}\|$, and

$\tilde{d}_{j,i} = \|\mathbf{u}_j - \mathbf{x}_i\|$. This principle immediately leads to the efficient *Simplex Volume Maximization* algorithm depicted in Algorithm 1. The time complexity of this algorithm is $O(mn)$. Since $m \ll n$ is kept as a constant, such a complexity is linear in the data set size n . A visual example of the algorithm applied to a 2D toy dataset is displayed in Fig. 2.

4. Archetype Hull Ranking

In this section, we address the problem that we raised in Section 3: how to correlate an input sample with archetypes. Note that this sample may be an unseen sample outside the set of available training samples, and that these archetypes can be stored in memory for real-time computations.

Since the archetype hull

$$\mathcal{S}(\mathcal{U}) = \left\{ \sum_{j=1}^m \theta_j \mathbf{u}_j \mid \theta_j \geq 0, 1 \leq j \leq m, \sum_{j=1}^m \theta_j = 1 \right\} \quad (3)$$

encloses most training samples in \mathcal{X} , we can project any input sample $\mathbf{x} \in \mathbb{R}^d$ onto \mathcal{S} , thereby obtaining a new (lossy) representation of \mathbf{x} in terms of the stored archetypes in \mathcal{U} . Formally, we reconstruct the input \mathbf{x} using a convex combination of the archetypes in \mathcal{U} as

$$\begin{aligned} \min_{\mathbf{z}(\mathbf{x}) \in \mathbb{R}^m} \quad & \|\mathbf{x} - \mathbf{U}\mathbf{z}(\mathbf{x})\|^2 \\ \text{s.t.} \quad & \mathbf{z}(\mathbf{x}) \geq 0, \mathbf{1}^\top \mathbf{z}(\mathbf{x}) = 1, \end{aligned} \quad (4)$$

where $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_m] \in \mathbb{R}^{d \times m}$, and $\mathbf{z}(\mathbf{x})$ contains the nonnegative reconstruction coefficients. It turns out that if $d < m$ the coefficient vector $\mathbf{z}(\mathbf{x})$ is sparse and the number of its nonzero elements is usually far smaller than m . In an algebraic perspective, $\mathbf{z}(\mathbf{x})$ acts as a vectorial mapping that maps any input sample \mathbf{x} to a sparse code in \mathbb{R}^m . The convex optimization problem in eq. (4) can be efficiently solved using a fast constrained least squares solver.

Through a simple archetype hull projection, we acquire the new representation $\mathbf{U}\mathbf{z}(\mathbf{x})$ for any input sample \mathbf{x} , where the coefficient vector $\mathbf{z}(\mathbf{x})$ directly reveals that which archetypes are relevant to the input \mathbf{x} ($z_j(\mathbf{x}) > 0$ implies that the archetype \mathbf{u}_j is relevant to \mathbf{x}). Although projecting \mathbf{x} onto the archetype hull $\mathcal{S}(\mathcal{U})$ gives rise to a correlation between the input \mathbf{x} and the archetypes \mathcal{U} , such a correlation is somewhat weak in the sense that the information of the training set \mathcal{X} is not fully exploited. In what follows, we discover a stronger correlation between \mathbf{x} and \mathcal{U} by proposing a novel graph-based ranking framework.

We first build a data-to-archetype affinity matrix between the training data \mathcal{X} and the archetypes \mathcal{U} , that is, $\mathbf{Z} =$

$$\begin{bmatrix} \mathbf{z}^\top(\mathbf{x}_1) \\ \dots \\ \mathbf{z}^\top(\mathbf{x}_n) \end{bmatrix} \in \mathbb{R}^{n \times m} \text{ which satisfies } \mathbf{Z} \geq \mathbf{0} \text{ and } \mathbf{Z}\mathbf{1} = \mathbf{1}.$$

With the data-to-archetype affinity matrix \mathbf{Z} , the theory [14] of *Anchor Graph* unveils an approximate data-to-data affinity matrix in an elegant manner. The Anchor Graph affinity matrix is in a low-rank form of $\mathbf{A} = \mathbf{Z}\mathbf{\Lambda}^{-1}\mathbf{Z}^\top \in \mathbb{R}^{n \times n}$, where $\mathbf{\Lambda} = \text{diag}(\mathbf{Z}^\top \mathbf{1}) \in \mathbb{R}^{m \times m}$. The low-rank nature of \mathbf{A} lends prominent advantages to storage and computations, thus making the Anchor Graph scale up to massive datasets. Let us define a diagonal matrix $\mathbf{D} \in \mathbb{R}^{n \times n}$ whose diagonal elements stem from the diagonal elements of \mathbf{A} , i.e., $\mathbf{D}_{ii} = \mathbf{z}^\top(\mathbf{x}_i)\mathbf{\Lambda}^{-1}\mathbf{z}(\mathbf{x}_i)$ for $1 \leq i \leq n$. As such, we derive the normalized Anchor Graph Laplacian matrix as follows (notice $\mathbf{A}\mathbf{1} = \mathbf{1}$)

$$\begin{aligned} \bar{\mathbf{L}} &= \mathbf{I} - (\mathbf{I} - \mathbf{D})^{-\frac{1}{2}} (\mathbf{A} - \mathbf{D}) (\mathbf{I} - \mathbf{D})^{-\frac{1}{2}} \\ &= \mathbf{I} - (\mathbf{I} - \mathbf{D})^{-\frac{1}{2}} \mathbf{A} (\mathbf{I} - \mathbf{D})^{-\frac{1}{2}} + \mathbf{D} (\mathbf{I} - \mathbf{D})^{-1} \\ &= (\mathbf{I} - \mathbf{D})^{-1} - (\mathbf{I} - \mathbf{D})^{-\frac{1}{2}} \mathbf{Z} \mathbf{\Lambda}^{-1} \mathbf{Z}^\top (\mathbf{I} - \mathbf{D})^{-\frac{1}{2}}. \end{aligned} \quad (5)$$

By means of this normalized Anchor Graph Laplacian $\bar{\mathbf{L}}$, we propose a novel graph-based ranking framework which is capable of scaling up the well-known manifold ranking

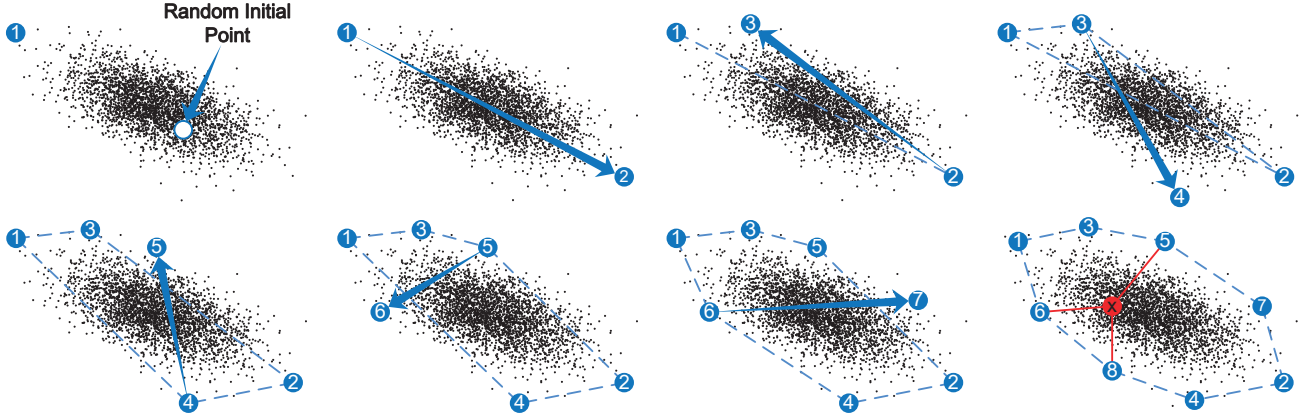


Figure 2. An example of archetype seeking on a 2D toy dataset consisting of 3000 2D points. Every subfigure shows one step of the archetype seeking process. The area surrounded by a dashed line denotes the current archetype hull produced by the currently chosen archetypes. The last subfigure shows the final archetype hull which encloses almost all data points, and shows a data point \mathbf{x} that is represented by a convex combination of three archetypes.

approach [33][15]. The original objective of manifold ranking is

$$\min_{\mathbf{f} \in \mathbb{R}^n} (1 - \alpha) \|\mathbf{f} - \mathbf{y}\|^2 + \alpha \mathbf{f}^\top \bar{\mathbf{L}} \mathbf{f}, \quad (6)$$

where $\mathbf{y} \in \mathbb{R}^n$ keeps the initial rank scores, $\mathbf{f} \in \mathbb{R}^n$ saves the final rank scores, and the parameter $0 < \alpha < 1$ controls the trade-off between proximity to the initial ranks in \mathbf{y} and graph-consistent smoothness of \mathbf{f} . Now we adapt the manifold ranking task in eq. (6) into the context of the archetype hull described in eq. (3). Complying with the convex reconstruction scheme in eq. (4) led by archetype hull projection, we assume that the rank score of any training sample \mathbf{x}_i is also a convex combination of rank scores $\{r_j\}_{j=1}^m$ of the archetypes (archetype \mathbf{u}_j has the rank score r_j) with the same combination coefficients $\mathbf{z}(\mathbf{x}_i)$ as those obtained in the archetype-based reconstruction. Thus, this assumption leads to $f_i = \mathbf{z}^\top(\mathbf{x}_i) \mathbf{r}$ in which $\mathbf{r} = [r_1, \dots, r_m]^\top$ is the rank vector over the archetype hull. Aggregating f_i 's for all training samples, we have $\mathbf{f} = \mathbf{Z} \mathbf{r}$. After substituting $\mathbf{f} = \mathbf{Z} \mathbf{r}$ into eq. (6) and using the Anchor Graph Laplacian $\bar{\mathbf{L}}$ written in eq. (5), we arrive at the following archetype-driven ranking framework:

$$\min_{\mathbf{r} \in \mathbb{R}^m} (1 - \alpha) \|\mathbf{Z} \mathbf{r} - \mathbf{y}\|^2 + \alpha \mathbf{r}^\top \mathbf{Z}^\top \bar{\mathbf{L}} \mathbf{Z} \mathbf{r}. \quad (7)$$

The optimal solution to eq. (7) is easily achieved by

$$\mathbf{r}^* = \left(\mathbf{Z}^\top \mathbf{Z} + \frac{\alpha}{1 - \alpha} (\mathbf{M}_1 - \mathbf{M}_2 \mathbf{\Lambda}^{-1} \mathbf{M}_2) \right)^{-1} \mathbf{Z}^\top \mathbf{y}, \quad (8)$$

where $\mathbf{M}_1 = \mathbf{Z}^\top (\mathbf{I} - \mathbf{D})^{-1} \mathbf{Z}$ and $\mathbf{M}_2 = \mathbf{Z}^\top (\mathbf{I} - \mathbf{D})^{-1/2} \mathbf{Z}$.

Eq. (8) indicates that when a training sample \mathbf{x}_i becomes a query (corresponding to $\mathbf{y} = [0, \dots, 0, 1, 0, \dots, 0]^\top$), the

resulting rank vector is

$$\mathbf{r}^\alpha(\mathbf{x}_i) = \left(\mathbf{Z}^\top \mathbf{Z} + \frac{\alpha}{1 - \alpha} (\mathbf{M}_1 - \mathbf{M}_2 \mathbf{\Lambda}^{-1} \mathbf{M}_2) \right)^{-1} \mathbf{z}(\mathbf{x}_i), \quad (9)$$

which relies on the parameter α . In a natural means, we can generalize eq. (9) to

$$\mathbf{r}^\alpha(\mathbf{x}) = \left(\mathbf{Z}^\top \mathbf{Z} + \frac{\alpha}{1 - \alpha} (\mathbf{M}_1 - \mathbf{M}_2 \mathbf{\Lambda}^{-1} \mathbf{M}_2) \right)^{-1} \mathbf{z}(\mathbf{x}), \quad (10)$$

which allows any sample \mathbf{x} as a query.

So far, the proposed archetype-driven ranking framework in eq. (7), namely *Archetype Hull Ranking* (AHR), employs the scalable Anchor Graph capturing the manifold structure of the whole training set \mathcal{X} to accomplish a strong correlation (or relevancy) in eq. (10) between any input \mathbf{x} and m archetypes in \mathcal{U} . Moreover, computing the relevance between \mathbf{x} and \mathcal{U} by eq. (10) is quite convenient, since $(\mathbf{Z}^\top \mathbf{Z} + \frac{\alpha}{1 - \alpha} (\mathbf{M}_1 - \mathbf{M}_2 \mathbf{\Lambda}^{-1} \mathbf{M}_2))^{-1}$ can be pre-computed in $O(mn)$ space and $O(m^2n)$ time complexities that still scale linearly with the training set size n .

It is worthwhile to clarify that our archetype hull ranking approach differs from the ranking approach proposed in [31] in the sense that our aim is to solve rank scores (\mathbf{r}) on much fewer archetypes while the latter directly solves full rank scores (\mathbf{f}) on entire training samples.

5. Face Modeling with Archetypes

To ensure the robustness resilient to intra-personal variations caused by varying poses, expressions, and illumination conditions, we adopt a block (*i.e.*, image patch) based

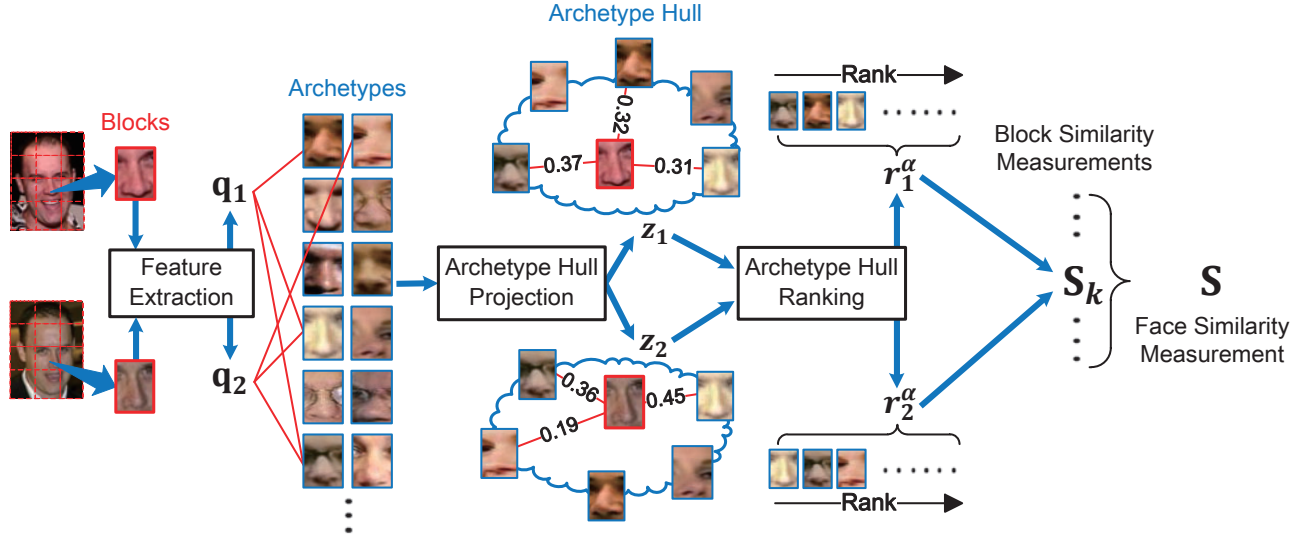


Figure 3. The flowchart of obtaining the proposed face similarity measure.

decomposition to model face images. In detail, we divide each face image into a fixed number of blocks, then extract low-level features from each block, and seek archetypes in block-level feature spaces by applying Algorithm 1.

In this section, we propose to tackle face recognition and verification together using the archetype hull ranking approach proposed in Section 4. First, we acquire a blockwise face similarity measure through harnessing the results of the archetype hull ranking in an unsupervised or supervised manner. Then, we achieve a holistic face similarity measure by integrating blockwise similarity measurements with equal or learned importance weights. The flowchart of our face modeling approach is plotted in Fig. 3.

5.1. Blockwise Similarity Measures

5.1.1 Unsupervised Approaches

Here we present two approaches for attaining blockwise similarity measures. Note that both of them do not use any class label information of the training set.

Baseline. Given any pair of feature vectors (q_1, q_2) at the k -th block position in face images, the baseline similarity measure is formulated as

$$S_k^{basic}(q_1, q_2) = z^\top(q_1) \Lambda^{-1} z(q_2), \quad (11)$$

which directly uses the affinity function of an Anchor Graph established over a training set $\mathcal{X}_k = \{\mathbf{x}_{ki}\}_{i=1}^n$ corresponding to the k -th block position.

Archetype Hull Ranking. The baseline measure merely takes advantage of archetype hull projection, and essentially takes the intersection between the resulting sparse coefficient vectors $z(q_1), z(q_2)$ as the similarity. By virtue of archetype hull ranking, we herewith obtain two rank vectors

$r^\alpha(q_1), r^\alpha(q_2)$ which can capture the more reliable relevancy between q_1, q_2 and the archetypes in \mathcal{U}_k . As such, we take the cosine similarity between the two rank vectors

$$S_k^{AHR}(q_1, q_2) = \frac{(r^\alpha(q_1))^\top (r^\alpha(q_2))}{\|r^\alpha(q_1)\| \|r^\alpha(q_2)\|}, \quad (12)$$

which measures the extent of the overlapping between q_1 and q_2 's relevant archetypes disclosed by the archetype hull ranking with respect to the queries q_1 and q_2 .

5.1.2 Supervised Approaches

Now we intend to leverage the label information of the training set \mathcal{X}_k as well as the archetypes in \mathcal{U}_k . Suppose that there are totally c classes (*i.e.*, persons) in the training image set. We specially assign a label indicator vector $\mathbf{Y}_i = [0, \dots, 0, 1, 0, \dots, 0] \in \{1, 0\}^{1 \times c}$ to each training

image $I_i = \{\mathbf{x}_{ki}\}_k$ belonging to the j -th class. Then each sample in \mathcal{X}_k has a ground-truth label and $\mathcal{U}_k \subset \mathcal{X}_k$ does

so. We write a label matrix $\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 \\ \dots \\ \mathbf{H}_m \end{bmatrix} \in \{1, 0\}^{m \times c}$ each

row of which corresponds to an archetype in \mathcal{U}_k . Like the previous subsection, we propose two supervised approaches which access the label matrix \mathbf{H} in yielding blockwise similarity measures.

Baseline. For any sample q , the archetype hull projection results in the convex reconstruction scheme $q \approx \mathbf{U}z(q)$, which motivates us to make the assumption that such a scheme is preserved in the label space. Hence, we can estimate the label vector of q as $\hat{\mathbf{Y}}(q) = z^\top(q)\mathbf{H}$. Subse-

quently, we take the inner-product between the estimated label vectors of an input pair $(\mathbf{q}_1, \mathbf{q}_2)$ as the similarity:

$$S_k^{basic}(\mathbf{q}_1, \mathbf{q}_2) = \hat{\mathbf{Y}}(\mathbf{q}_1)(\hat{\mathbf{Y}}(\mathbf{q}_2))^\top = \mathbf{z}^\top(\mathbf{q}_1)\mathbf{H}\mathbf{H}^\top\mathbf{z}(\mathbf{q}_2). \quad (13)$$

Archetype Hull Ranking. While eq. (13) offers a semantic similarity measure, we try to discover a more discriminative semantic similarity measure by controlling the archetype hull ranking. Let us keep the indexes of \mathbf{q} 's relevant archetypes revealed by $\mathbf{z}(\mathbf{q})$ into an initial reference set $R(\mathbf{q}) = \{j | z_j(\mathbf{q}) > 0, j \in [1 : m]\}$. Afterwards, we apply the archetype hull ranking to obtain the rank vector $\mathbf{r}^\alpha(\mathbf{q})$ from which we acquire another reference set

$$\tilde{R}^\alpha(\mathbf{q}) = \{j | r_j^\alpha(\mathbf{q}) \in \Delta, j \in [1 : m]\},$$

in which the set $\Delta = \{\text{largest } |R(\mathbf{q})| \text{ elements in } \mathbf{r}^\alpha(\mathbf{q})\}$ and $|R(\mathbf{q})|$ denotes the cardinality of the set $R(\mathbf{q})$. Through intersecting the two reference sets $R(\mathbf{q})$ and $\tilde{R}^\alpha(\mathbf{q})$, we can refine \mathbf{q} 's relevant archetypes and thereby re-estimate the label vector of \mathbf{q} as

$$\hat{\mathbf{Y}}^\alpha(\mathbf{q}) = \frac{\sum_{j \in R(\mathbf{q}) \cap \tilde{R}^\alpha(\mathbf{q})} \mathbf{r}_j^\alpha(\mathbf{q}) \mathbf{H}_j}{\sum_{j \in R(\mathbf{q}) \cap \tilde{R}^\alpha(\mathbf{q})} \mathbf{r}_j^\alpha(\mathbf{q})}. \quad (14)$$

A new semantic similarity measure is then given by

$$S_k^{AHR}(\mathbf{q}_1, \mathbf{q}_2) = \hat{\mathbf{Y}}^\alpha(\mathbf{q}_1)(\hat{\mathbf{Y}}^\alpha(\mathbf{q}_2))^\top. \quad (15)$$

Critically, we are able to make the semantic similarity measure S_k^{AHR} in eq. (15) as discriminative as possible by tuning the model parameter $\alpha \in (0, 1)$. The optimal α^* is achieved through performing the semantic similarity validation on the training set \mathcal{X}_k as follows

$$\alpha^* = \arg \min_{\alpha \in (0, 1)} \left\| \mathbf{Y}\mathbf{Y}^\top - \hat{\mathbf{Y}}^\alpha(\hat{\mathbf{Y}}^\alpha)^\top \right\|_F^2, \quad (16)$$

where $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \dots \\ \mathbf{Y}_n \end{bmatrix}$ and $\hat{\mathbf{Y}}^\alpha = \begin{bmatrix} \hat{\mathbf{Y}}^\alpha(\mathbf{x}_{k1}) \\ \dots \\ \hat{\mathbf{Y}}^\alpha(\mathbf{x}_{kn}) \end{bmatrix}$. Eq. (16) implies that with the optimal parameter α^* , the archetype hull ranking yields the discriminative semantic similarity measure in eq. (15) that can best fit the ground-truth semantic similarity $\mathbf{Y}\mathbf{Y}^\top$ in the least-squares sense.

5.2. Face Similarity Measures

In order to acquire a holistic similarity between an input face pair $Q_1 = \{\mathbf{q}_{k1}\}_k$ and $Q_2 = \{\mathbf{q}_{k2}\}_k$, we can simply integrate all blockwise similarity measurements $\{S_k\}_{k=1}^b$ (b is the total number of blocks) with equal weights, leading to

$$F(Q_1, Q_2) = \frac{\sum_{k=1}^b S_k(\mathbf{q}_{k1}, \mathbf{q}_{k2})}{b}, \quad (17)$$

where the blockwise similarity measure $S_k(\cdot, \cdot)$ can be the two unsupervised measures in eqs. (11)(12) or the supervised baseline measure in eq. (13).

During learning the supervised AHR measure in eq. (15) on the training set \mathcal{X}_k across all b blocks, we are already aware of the importance of each block, which is reflected into the squared fitting error $\epsilon_k^2 = \|\mathbf{Y}\mathbf{Y}^\top - \hat{\mathbf{Y}}^{\alpha^*}(\hat{\mathbf{Y}}^{\alpha^*})^\top\|_F^2$. The smaller ϵ_k^2 , the more reliable the supervised measure $S_k^{AHR}(\cdot, \cdot)$. Consequently, we introduce the importance weights $\exp(-\lambda\epsilon_k^2/n^2)$ ($\lambda > 0$) to achieve the holistic face similarity measure as

$$\tilde{F}(Q_1, Q_2) = \frac{\sum_{k=1}^b \exp(-\frac{\lambda\epsilon_k^2}{n^2}) S_k^{AHR}(\mathbf{q}_{k1}, \mathbf{q}_{k2})}{\sum_{k=1}^b \exp(-\frac{\lambda\epsilon_k^2}{n^2})}. \quad (18)$$

6. Experiments

We conduct experiments on three widely used benchmark face databases Multi-PIE [7], Pubfig83 [20], and LFW [9]. The face recognition experiments on Multi-PIE and Pubfig83 are conducted under a gallery-query recognition mode. For the face verification experiments on LFW, we follow the unsupervised protocol of the LFW benchmark for evaluations. For all datasets, each face image is divided into 8×10 blocks, and a 59-dimensional uniform LBP [1] is extracted from each block as facial features. We apply the two proposed supervised approaches (called as ‘‘supervised basic AHR’’ and ‘‘supervised AHR’’) for face recognition, and the two proposed unsupervised approaches (called as ‘‘unsupervised basic AHR’’ and ‘‘unsupervised AHR’’) for face verification, respectively. Note that ‘‘supervised AHR’’ learns the optimal α on the training set, while ‘‘unsupervised AHR’’ uses the fixed $\alpha = 0.16$.

Archetype Type. We first study the influences of different types of archetypes on the face recognition accuracy achieved by our proposed ‘‘supervised basic AHR’’ and ‘‘supervised AHR’’. The comparison results are listed in Table 1, where ‘‘Random Exemplars’’ refers to randomly selecting data exemplars as archetypes, ‘‘K-Medoids’’ refers to using the output of K-Medoids clustering as archetypes, and ‘‘Archetype Seeking’’ refers to using the output of archetype seeking Algorithm 1 as archetypes. This group of experimental results demonstrate that the archetype seeking algorithm always produces the archetypes of the best quality, which lead to the highest recognition accuracy for both ‘‘supervised basic AHR’’ and ‘‘supervised AHR’’.

Archetype Number. The number of archetypes m is also an important factor to affecting the recognition/verification performance. To shed light on the effect of varying m , we conduct recognition and verification experiments with varying m and show the corresponding results in Figs. 4 and 5. The experimental results clearly indicate that the recognition/verification accuracy increases as m grows.

Table 1. Comparison of different types of archetypes on the Multi-PIE dataset. The number of archetypes is fixed to $\lfloor n/10 \rfloor$.

Archetype Type	Approach	
	supervised basic AHR	supervised AHR
Random Exemplars	79.4%	84.6%
K-Medoids	82.3%	86.9%
Archetype Seeking	86.4%	92.1%

Table 2. The recognition accuracy on the Multi-PIE and Pubfig83 datasets. The number of archetypes is fixed to $\lfloor n/10 \rfloor$.

Approach	Accuracy (%)	
	Multi-PIE	Pubfig83
LBP	57.7	56.3
Eigenfaces [25]	61.9	56.3
Fisherfaces [3]	67.4	60.2
SRC [30]	81.5	75.2
supervised basic AHR	86.4	80.2
supervised AHR	92.1	85.5
supervised AHR (cross-dataset)	88.6	83.3

Table 3. The verification accuracy on the LFW dataset under the unsupervised setting. The number of archetypes is fixed to 3,000.

Approach	Accuracy (%)
LBP	68.50
unsupervised basic AHR	73.68
unsupervised AHR	83.18
MRF-MLBP [2]	80.08
I-LPQ [10]	86.20
PAF [32]	87.77

Face Recognition. We conduct face recognition experiments on the Multi-PIE and Pubfig83 datasets. All face images are uniformly aligned according to an affine transform. On Multi-PIE, the training set includes 20,000 images from 37 persons, the gallery set contains 6,000 images from 300 persons, and the query set contains 2,000 images. On Pubfig83, the training set consists of 4,000 images from 83 persons, the gallery set has 30 images from each of 83 persons, and the query set has 1,000 images. We compare the proposed supervised approaches against several representative recognition methods. The experimental results are listed in Table 2 which shows that “supervised AHR” is the best and “supervised basic AHR” is the second best in terms of recognition accuracy. In Table 2, “LBP” refers to directly using concatenated LBP features; “Eigenfaces” and “Fisherfaces” learn subspaces over concatenated LBP features; “LBP”, “Eigenfaces”, and “Fisherfaces” all use the cosine similarity for the final recognition. To verify the generalization power of our approaches, we also conduct cross-dataset experiments shown as “cross-dataset” in Table 2. For the Multi-PIE test data, we use the Pubfig83 training set for cross-training; for the Pubfig83 test data,



Figure 4. The recognition accuracy achieved by our approaches with varying archetype proportions m/n on the Multi-PIE and Pubfig83 datasets.

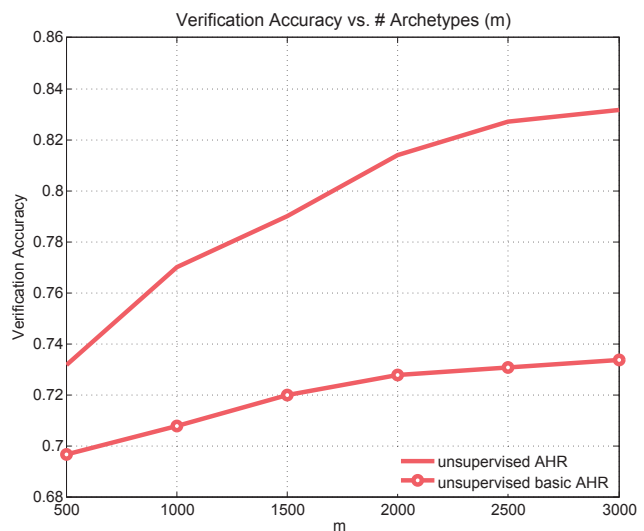


Figure 5. The verification accuracy achieved by our approaches with a varying number m of archetypes on the LFW dataset.

we use the Multi-PIE training set for cross-training. The cross-dataset recognition results illustrate that our approach “supervised AHR” is robust to the training source, and can still achieve higher accuracy than the other methods even if the training set is changed to a different data source.

Face Verification. We follow the standard unsupervised protocol of the LFW benchmark. The roughly aligned images of LFW are from the LFW-a dataset [29]. Without using labels, faces images outside LFW are permitted for training. Specifically, our unsupervised approaches combine the two training sets of the Multi-PIE and Pubfig83 datasets for training. The verification accuracy averaged

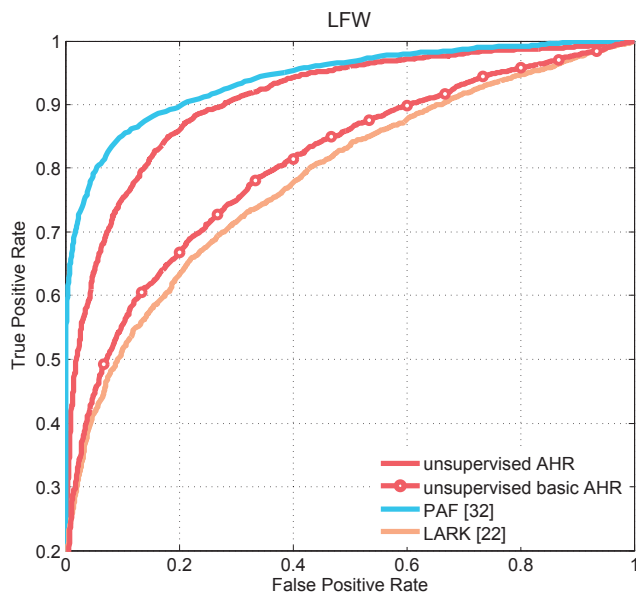


Figure 6. The ROC curves of face verification approaches with the standard unsupervised LFW protocol.

over ten test folds is reported for all compared methods in Table 3 which shows that our “unsupervised AHR” is ranked third among the recently reported highest unsupervised results. We also plot the ROC curves in Fig. 6, where the curve of “unsupervised AHR” is the second highest.

7. Conclusions

In this paper, the geometric archetype hull model is leveraged into face modeling. Based on a scalable graph characterized by a compact set of archetypes, we propose an archetype-driven ranking framework that can scale up to massive datasets. This framework generates a rank vector over the archetype hull, which well captures the relevance between any query and the archetypes. The archetype hull ranking is then carried out on each block of face images, leading to a blockwise face similarity measure. By integrating blockwise similarity measurements with learned importance weights, a sensible face similarity measure is eventually yielded. The performance of the face similarity measure is corroborated through recognition and verification experiments performed on three benchmark face databases.

Acknowledgements This work is supported by the General Research Fund sponsored by the Research Grants Council of the Hong Kong SAR (Project No. CUHK 416312 and CUHK 416510) and Guangdong Innovative Research Team Program (No. 201001D0104648280). Wei Liu is supported by Josef Raviv Memorial Postdoctoral Fellowship.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *TPAMI*, 28(12):2037–2041, 2006.

[2] S. R. Arashloo and J. Kittler. Efficient processing of mrfs for unconstrained-pose face recognition. In *Proc. Biometrics: Theory, Applications and Systems*, 2013.

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *TPAMI*, 19(7):711–720, 1997.

[4] L. M. Blumenthal. *Theory and Applications of Distance Geometry*. Oxford University Press, UK, 1953.

[5] Z. Cao, Q. Yin, X. Tang, and J. Sun. Face recognition with learning-based descriptor. In *Proc. CVPR*, 2010.

[6] M. J. Eugster and F. Leisch. Weighted and robust archetypal analysis. *Computational Statistics & Data Analysis*, 55(3):1215–1225, 2011.

[7] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. In *Proc. FGR*, 2008.

[8] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang. Face recognition using laplacianfaces. *TPAMI*, 27(3):328–340, 2005.

[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, 2007.

[10] S. U. Hussain, T. Napoléon, and F. Jurie. Face recognition using local quantized patterns. In *Proc. BMVC*, 2012.

[11] C. C. G. Jung. *The archetypes and the collective unconscious*, volume 9. Bollingen, 1981.

[12] S. Li, P. Wang, J. Louviere, and R. Carson. Archetypal analysis: A new way to segment markets based on extreme individuals. In *Proc. Australian and New Zealand Marketing Academy Conference*, 2003.

[13] Z. Li, W. Liu, D. Lin, and X. Tang. Nonparametric subspace analysis for face recognition. In *Proc. CVPR*, 2005.

[14] W. Liu, J. He, and S.-F. Chang. Large graph construction for scalable semi-supervised learning. In *Proc. ICML*, 2010.

[15] W. Liu, Y.-G. Jiang, J. Luo, and S.-F. Chang. Noise resistant graph ranking for improved web image search. In *Proc. CVPR*, 2011.

[16] W. Liu, J. Wang, and S.-F. Chang. Robust and scalable graph-based semisupervised learning. *Proceedings of the IEEE*, 100(9):2624–2638, 2012.

[17] W. Liu, Y. Wang, S. Z. Li, and T. Tan. Null space approach of fisher discriminant analysis for face recognition. *Biometric Authentication, Lecture Notes in Computer Science*, 3087:32–44, 2004.

[18] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[19] B. Moghaddam, T. Jebara, and A. Pentland. Bayesian face recognition. *Pattern Recognition*, 33(11):1771–1782, 2000.

[20] N. Pinto, Z. Stone, T. Zickler, and D. Cox. Scaling up biologically-inspired computer vision: A case study in unconstrained face recognition on facebook. In *Proc. CVPR Workshop on Biologically-Consistent Vision*, 2011.

[21] C. Seiler and K. Wohlrahe. Archetypal scientists. *Journal of Informetrics*, 7(2):345–356, 2013.

[22] H. J. Seo and P. Milanfar. Face verification using the lark representation. *IEEE Transactions on Information Forensics and Security*, 6(4):1275–1286, 2011.

[23] C. Thureau. Nearest archetype hull methods for large-scale data classification. In *Proc. ICPR*, 2010.

[24] C. Thureau, K. Kersting, and C. Bauckhage. Yes we can: simplex volume maximization for descriptive web-scale matrix factorization. In *Proc. CIKM*, 2010.

[25] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):7186, 1991.

[26] X. Wang and X. Tang. Dual-space linear discriminant analysis for face recognition. In *Proc. CVPR*, 2004.

[27] X. Wang and X. Tang. A unified framework for subspace face recognition. *TPAMI*, 26(9):1222–1228, 2004.

[28] X. Wang and X. Tang. Random sampling for subspace face recognition. *IJCV*, 70(1):91–104, 2006.

[29] L. Wolf, T. Hassner, and Y. Taigman. Effective unconstrained face recognition by combining multiple descriptors and learned background statistics. *TPAMI*, 33(10):1978–1990, 2011.

[30] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *TPAMI*, 31(2):210–227, 2009.

[31] B. Xu, J. Bu, C. Chen, D. Cai, X. He, W. Liu, and J. Luo. Efficient manifold ranking for image retrieval. In *Proc. SIGIR*, 2011.

[32] D. Yi, Z. Lei, and S. Z. Li. Towards pose robust face recognition. In *Proc. CVPR*, 2013.

[33] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf. Ranking on data manifolds. In *NIPS 16*, 2003.