Image Clustering Under Domain Shift

Raghuraman Gopalan
Dept. of AI and Automation, AT&T Labs-Research

***Abstract*—We address domain adaptation in the context of clustering where we are given a set of unlabeled data, coming from several domains, and the goal is to group data into different categories regardless of the domain they come from. This is a challenging problem since we do not have any supervision unlike most adaptation scenarios studied earlier, and is very relevant in practical industry applications where labeled data often comes at a premium especially while deploying services that do not have a comparable predecessor. Our philosophy in addressing this problem draws motivation from the concept of dirty paper coding, a communications technique where the signal being transmitted through a noisy channel is encoded with priors on the possible noise patterns to assist reliable decoding of the signal at the receiver. We focus on image applications in this paper, where we encode priors on possible image domain shift factors such as viewpoint, lighting, blur and appearance variations and utilize geometric adaptation mechanisms to perform clustering. We illustrate the utility of our approach on standard datasets involving objects and faces, by obtaining around 18% improvement on average over existing approaches.**

I. Introduction

Domain adaptation has become mainstream in classification problems owing to practical constraints that arise when the training (source domain) and testing (target domain) datasets have different distributions [1]. There have been several approaches presented in the computer vision community to address this problem, which utilize tools related to manifolds, sparse coding, transform learning and max-margin learners, and have successfully demonstrated their utility for a wide range of classification problems involving objects, scenes, faces, actions and much more [2]. Besides classification, there have also been adaptation efforts focusing on other learning problems such as detection [3] and regression [4].

However one adaptation problem that has not received much attention is clustering, where we do not have any supervision (data labels) and this forms the focus of this paper. While there exist several unsupervised adaptation methods [5], they do not assume labels in target domain but assume partial labels in the source domain, in part because their focus has been on classification and thus do not pertain to the typical unsupervised clustering scenario where we do not have 'any' label for data from 'any' domain. In this paper we study this problem by assuming we are given one large pool of unlabeled data D, corresponding to N different categories and coming from unknown number of domains, and the goal is to group data from same categories regardless of the domain they come from. It is a hard problem since there is no demarcation between different domains, and such cases arise often in practice where one collects data from many different unknown number of sources (say from web search pulling results from several feeds) and would want a way to discover categories present in that unlabeled data pool irrespective of the domain shift present in them. While there exist a few adaptive clustering approaches in the literature, they either assume prior grouping of the unlabeled data into different domains [6] which restricts their practical utility, or their construction is inherently tied to identifying the domains before doing clustering [7] which is an ill-posed problem in itself.

Our approach to this challenging problem takes motivation from dirty paper coding [8], which addresses the scenario where the signal is being transmitted through a noisy communications channel and discusses strategies to encode the signal with priors on the potential noise patterns so that the receiver can have more probability of success in decoding the signal. In our case of clustering under domain shift, the signal (an image) is affected by different forms of noise (domain shift) such as variations in illumination, occlusion, viewpoint, intrinsic dataset characteristics among others and we are looking at strategies to encode the signal in such a way that some of the noise patterns can be reduced so as to assist proper decoding (i.e. performing good clustering). In doing so, our objective is to not require any prior grouping of data into different domains, or to necessitate identifying the different domains to enable the clustering process.

In pursuit of this goal, we utilize extensive studies in the computer vision community on developing features that are invariant to transformations caused by illumination, viewpoint etc. Examples include SIFT features that are invariant to scale [9], self-quotient image that is insensitive to illumination [10], and deconvolution features that are robust to blurring [11]. For some reason, most existing adaptation solutions do not make good use of these features and they work with just one feature representation that comes with the dataset. One reason perhaps is that they focus mostly on classification scenarios where the availability of (partial) labeled data provide reliable cues to mitigate the effects of domain shift. However for the clustering scenario we are considering in this paper, we do not have any labeled data, and hence we expand the choice of features to represent our unlabeled pool of data D. We use M features to represent D, and this results in M domains $\{D_i\}_{i=1}^{M}$ where the domain D_i contains data from D represented by the i^{th} feature.

Each of these domains provide some degree of robustness in accounting for the domain shift, and while each by themselves may not provide the invariance we desire, our hope is that by combining the information contained in them we can arrive at a reasonable data representation that mitigates the effect of domain shift. Our combination strategy uses results from Grassmannian manifold sampling to bridge domain shift [12], where we harness the power of big data by generating several data representations conveying potential information on the domain shift, and then perform clustering on the manifold to group different categories of data contained in D. We illustrate the advantages of our approach by performing experiments on standard adaptation datasets containing objects

and faces.

A. Contributions

- We study the largely under-studied problem of adaptation in the clustering scenario, where there is no labeled data from any of the domain(s), without assuming prior grouping of data into their corresponding domains or requiring identification of different domains to obtain the cluster outputs
- Motivated by dirty paper coding principles, we rely on the power of big data by generating several potentially relevant data representations by using studies on invariant image features, and geometric modeling strategies to combine information contained in them to perform clustering
- We demonstrate the utility of our method over other approaches that require prior grouping of data into different domains through empirical evaluation on two adaptation benchmarks, where we obtain an improvement in clustering accuracy of around 18% on average

We first review some related work on domain adaptation in Section II and provide more details of our approach in Section III. Experimental results are given in Section IV and Section V concludes the paper.

II. Related Work

Domain adaptation has become mainstream in machine learning, natural language processing and computer vision communities over the last decade [1]. Its importance can be attributed to the prevalence of different types of sensors that have some intrinsically different domain-specific characteristics that makes it necessary to bridge the domain shift variations before performing statistical inference. Examples include using speech recognition systems trained on indoor conversations to test on outdoor conversations, doing sentiment analysis on book reviews using models trained on electronic gadget reviews, and performing face recognition using surveillance photos for training and consumer domain photos for testing. There have been established studies documenting the importance of accounting for domain shift, with the inference accuracy dropping by as much as 60% when domain shift is not bridged [2].

In the computer vision community, the main focus has been on classifying objects, faces, events etc. across datasets that have domain shift variations. There are unsupervised adaptation approaches that assume the source domain to have partial labels and the target domain to be completely unlabeled, and semi-supervised adaptation methods that assume both the source and target domains to have partial labels. Strategies to account for domain shift have ranged from learning intermediate representations to bridge domain shift using manifolds [5], learning transformation spaces where the domain variations and the classification errors are minimized [13], hierarchical schemes that jointly learn adaptive features and classifiers from raw images [14], and classifier adaptation that re-adjusts the weights of pre-trained classifiers to minimize domain shift using max-margin principles [15].

However there have been relatively much fewer approaches that focus on the clustering scenario where there are no labels in both source and target domains, and addressing the realistic scenario where may not even know how many domains are present and we are given a single large pool of unlabeled data D containing N categories from unknown number of domains. The absence of labeled data makes it even harder to account for domain shift, since we do not have any evidence on how the domain shift has affected the data categories. One of the first efforts in this area was by [7] that approached this problem by identifying latent domains that would ideally represent each data category from each domain and then combining their information using transform coding to perform clustering. The efficacy of this method was directly dependent on identifying the latent domains accurately, which is a hard problem by itself. Another approach [6] assumes the demarcation of the domains present in the data, which makes it an easier problem, and then uses principles from sparse subspace clustering to bridge domain shift. We present an alternate approach in this paper that does not assume demarcation of the domains, nor the requirement to identify the domains in the process, by taking motivation from dirty paper coding and leveraging the power of big data by representing the data D using features invariant to different aspects of domain shift as researched by the computer vision community over the last few decades.

III. Proposed Approach

A. Motivation

We are given a completely unlabeled dataset D, where the data (images) corresponds to N different categories and has been acquired from unknown number of domains. The goal of this work is to group the data into N clusters $\{C_i\}_{i=1}^{N}$ by accounting for the domain shift such that each cluster C_i contains data belonging to the i^{th} category. While we do not know what domain shift variations are present in the dataset, we do know some possible causes for domain shift in images, and there exists a vast body of literature addressing visual domain shift variations such as viewpoint, lighting, affine transformations, blur, resolution, occlusion etc. using concepts from invariants [16], [17], image formation models [18], [19], and robust local features [20]. While one can argue what sets domain adaptation apart from these works is the lack of clear knowledge of 'what' constitutes the domain change, the above-mentioned factors does cover most, if not all, causes for visual domain change. However, most existing work on visual domain adaptation have not explicitly utilized results from this rich literature. This leads us to following question: if one has 'some' knowledge on the possible domain shift in the data, how to utilize that information in performing adaptation?

This exact question has been investigated atleast thirty years back in the communications literature as dirty paper coding [8]. The scenario considered there was, a signal sent by a transmitter is affected by dirt (or noise) in the channel before reaching the receiver. If the sender and receiver have partial information about the properties of the dirt, then is it possible to correctly interpret the transmitted signal from

the received signal using (i) appropriate coding of the signal in lieu of the dirt, and (ii) modeling strategies that reduce dirt-driven dissimilarity between the transmitted and received signals. We fit this idea in the context of clustering under domain shift by considering (i) *Signal Coding* - what features we can use to represent the images contained in D so that we get partial robustness to domain shift from each feature, and (ii) *Signal Modeling* - how to combine several such partially robust feature representations so as to account for the holistic domain shift present in D.

B. Signal Coding

Given the unlabeled dataset D containing n images, we first code them using the following feature representations to achieve partial robustness to some known visual domain shift patterns. More specifically, (i) we perform histogram equalization [23] on D and extract SIFT codebook features [21] to obtain robustness to lighting, scale and rotation, (ii) compute self-quotient image [10] from D and extract SIFT codebook features to get additional robustness to non-linear lighting, (iii) perform image deconvolution [11] on D and extract SURF codebook features [22] to get robustness to blur and rotation, and finally (iv) extract local phase quantized codebooks [24] from D to obtain robustness to blur, lighting and partial occlusions. Essentially, we now have M=4 domains D_1 to D_4, with each domain D_i containing different feature descriptions of the original n images. We set the dimension d of the codebooks across all domains as the same.

C. Signal Modeling

We then pursue modeling strategies on these domains so as to minimize the impact of domain shift to assist in clustering. We take a two-stage approach. (i) First we look at domain-level modeling so that the information contained across the domains can be combined so that the real domain shift can be tackled more effectively on a global level. (ii) We then translate the information we have learnt onto the individual images to perform more local modeling using which the data can be grouped into the appropriate categories. For both stages, we utilize subspace geometry tools pertaining to Grassmannian manifold since it has been shown effective for domain adaption by many different approaches [5], though they were tailored for the classification scenario.

1) Global Domain-level Modeling: From each of the M domains D_i, we first generate a p dimensional generative subspace S_i by performing principal component analysis (PCA) [25]. Each subspace S_i is a point on the Grassmann manifold $G_{d,p}$ which is the space of all p-dimensional subspaces in R^d [12]. We now learn how the information flows between these points, using the notion of the geodesic which is the shortest path between a pair of points on the manifold. What this conveys is how to combine or bridge the partial information conveyed about the domain shift in these points. We consider geodesics between all pairs of points (S_i, S_j) and then sample additional points along those geodesics using the notion of inverse exponential maps (Algorithm 1). Let $\bar{S} = \{\bar{S}_i\}_{i=1}^{m}$ denote the collection of all m sampled points. Each of these points is a weak representation that partially bridges the domain shift. We now seek to model this information into a usable form by learning a parametric Gaussian distribution from $\bar{S}$.

Algorithm 1: Numerical computation of the velocity matrix: The inverse exponential map [26]

Given two points S_1 and S_2 on the Grassmannian $G_{d,p}$.
(1) Compute the $d \times d$ orthogonal completion Q of S_1.
(2) Compute the thin CS decomposition of $Q^T S_2$ given by $Q^T S_2 = \begin{pmatrix} X_C \\ Y_C \end{pmatrix} = \begin{pmatrix} V_1 & 0 \\ 0 & \tilde{V}_2 \end{pmatrix} \begin{pmatrix} \Gamma(1) \\ -\Sigma(1) \end{pmatrix} V^T$
(3) Compute $\{\theta_i\}$ which are given by the arccos and arcsine of the diagonal elements of Γ and Σ respectively, i.e. $\gamma_i = \cos(\theta_i)$, $\sigma_i = \sin(\theta_i)$. Form the diagonal matrix Θ with θ's as diagonal elements.
(4) Compute $A = \tilde{V}_2 \Theta V_1^T$.

One procedure to learn intrinsic distributions on the manifold is to utilize the tangent space approximation, which is a locally Euclidean representation of the non-linear manifold space [12]. Since the Grassmannian is an analytical manifold, there are expressions to compute this approximation using inverse exponential mapping. In Algorithm 2 we summarize the generic steps involved in learning a distribution corresponding to a set of points (subspaces). While we pursue a Gaussian distribution in this work, any other parametric/non-parametric distributions can be fitted. We then randomly sample 100 representative points from the Gaussian, obtain their representation on the manifold using exponential mapping (Algorithm 3) to result in the final set of subspaces $S^\star$ that convey the consolidated bridged information on the domain shift, which we have obtained from our coded domains D_i. This is a global level modeling, as it has focused on how partial information on domain shift conveyed by each D_i can be transported across other domains.

Algorithm 2: Learning Intrinsic Gaussians on the manifold

Given a collection of m points $\bar{S}$ on $G_{d,p}$
(1) Compute the mean $\bar{\mu}$ of $\bar{S}$ using the Karcher mean algorithm given in Algorithm 4.
(2) For each point in $\bar{S}_i$ in $\bar{S}$, compute the inverse exponential map defined at $\bar{\mu}$ to obtain its tangent space approximation $\bar{S}_i'$.
(3) Fit a Gaussian on the locally Euclidean tangent space to the points $\{\bar{S}_i'\}_{i=1}^{m}$.

2) Local Data-level Modeling: We now translate the information contained in $S^\star$ onto the individual data such that it can assist in clustering them into their N categories. We project data contained in D_i's onto the subspaces $S^\star$ to get a p-dimensional representation. We then group the projections corresponding to *each* of the n original images and learn a q dimensional subspace by doing PCA. Let $T = \{T_i\}_{i=1}^{n}$ denote the collection of all such subspaces. Each T_i is a point on the Grassmannian $G_{p,q}$ which is the space of all q-dimensional subspaces[1] in R^p. We now perform clustering

[1]The subspace dimension q of T_i (and p of S_i) is determined by a heuristic tied to the number of eigen vectors required to preserve 90% of the PCA energy.

Algorithm 3: Algorithm for computing the exponential map, and sampling along the geodesic [26]

Given a point S_1 on the Grassmann manifold $G_{d,p}$ and a tangent vector $B = \begin{pmatrix} 0 & A^T \\ -A & 0 \end{pmatrix}$.
(1) Compute the $d \times d$ orthogonal completion Q of S_1.
(2) Compute the compact SVD of the direction matrix $A = \tilde{V}_2 \Theta V_1$.
(3) Compute the diagonal matrices $\Gamma(t')$ and $\Sigma(t')$ such that $\gamma_i(t') = \cos(t'\theta_i)$ and $\sigma_i(t') = \sin(t'\theta_i)$, where θ's are the diagonal elements of Θ.
(4) Compute $\Psi(t') = Q \begin{pmatrix} V_1\Gamma(t') \\ -\tilde{V}_2\Sigma(t') \end{pmatrix}$, for various values of $t' \in [0, 1]$.

Algorithm 4: Algorithm to compute the Karcher mean [27].

Given a set of u points $U = \{U_i\}_{i=1}^{u}$ on the manifold.
(1) Let $\bar{\mu}_0$ be an initial estimate of Karcher mean, by randomly picking an element of U. Set j = 0.
(2) For each $i = 1, .., u$, compute the inverse exponential map ν_i of U_i about the current estimate of the mean, i.e. $\nu_i = \exp_{\bar{\mu}_j}^{-1}(U_i)$.
(3) Compute the average tangent vector $\bar{\nu} = \frac{1}{u}\sum_{i=1}^{u} \nu_i$.
(4) If $\|\bar{\nu}\|$ is small, then stop. Else, move $\bar{\mu}_j$ in the average tangent direction using $\bar{\mu}_{j+1} = \exp_{\bar{\mu}_j}(\epsilon\bar{\nu})$, where $\epsilon > 0$ is small step size, typically 0.5, and $\exp_{\bar{\mu}_j}$ is the exponential map at $\bar{\mu}_j$.
(5) Set $j = j + 1$ and return to Step 2. Continue till $\bar{\mu}_j$ does not change anymore or till maximum iterations are exceeded.

on T to group them into N categories, by doing k-means on the Grassmannian $G_{q,p}$. The main difference between this and the standard Euclidean k-means is that we use the geodesic distance on the Grassmann manifold instead of the regular l_2 distance.

More specifically, given the set of points $T = (T_1, T_2, ..., T_n)$ on $G_{p,q}$, we seek to estimate N clusters $C = (C_1, C_2, ..., C_N)$ with cluster centers $(\mu_1, \mu_2, ..., \mu_N)$ so that the sum of geodesic-distance squares, $\sum_{i=1}^{N} \sum_{T_j \in C_i} \bar{d}^2(T_j, \mu_i)$ is minimized. Here $\bar{d}^2(T_j, \mu_i) = |\exp_{\mu_i}^{-1}(T_j)|^2$, where $\exp_{\mu_i}^{-1}$ is the inverse exponential map computed from tangent plane centered at μ_i. As is the case with standard Euclidean k-means, we can solve this problem using an EM-based approach. We initialize the algorithm with a random selection of N points as the cluster centers. In the E-step, we assign each of the points in T to the nearest cluster center. Then in the M-step, we recompute the cluster centers using the Karcher mean algorithm given in Algorithm 4.

IV. Experiments

We evaluate our clustering algorithm on two adaptation benchmarks involving objects from the Berkeley Office dataset [13] and faces from the Maryland AA01 dataset [28]. We

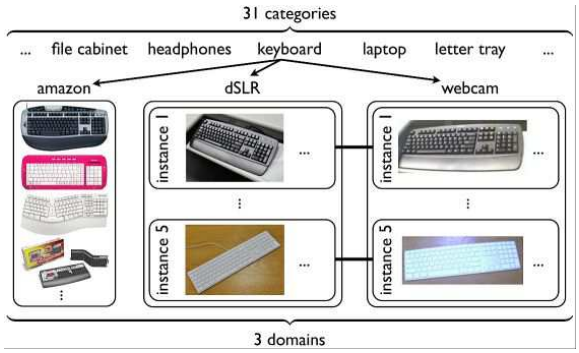

Fig. 1. Domain shift illustration from the Berkeley Office dataset [13]. The dataset contains variations in appearance, resolution, illumination among others.

Method	Object clustering accuracy (mean±std. deviation)
SSC	54.61±4.20
LRR	35.75±5.08
CO-SSC	51.44±4.99
CO-LRR	50.37±3.40
DA-SSC	49.41±5.05
DA-LRR	35.05±4.62
ED-SSC	53.32±5.69
ED-LRR	41.93±5.73
GM-SSC	53.92±5.89
GM-LRR	48.06±3.06
Ours	**71.96±3.21**

TABLE I. Comparison of object clustering accuracy on the Berkeley Office adaptation dataset [13]. Results of other approaches are given from [6].

compare our performance with other types of approaches from the literature such as (1) sparse subspace clustering (SSC) [29] and low-rank representation based subspace clustering (LRR) [30] which do not explicitly perform adaptation, (2) adaptation approaches proposed for classification such as the frustratingly easy adaptation (ED) [31], correlation alignment (CO) [32] and Grassmann manifold adaptation (GM) [5], whose learnt representations are used with the clustering algorithms SSC and LRR, and (3) adaptation versions of SSC and LRR namely DA-SSC and DA-LRR proposed by [6]. This gives us an idea of how standard clustering algorithms perform under domain shift (type 1), how adaptation approaches originally meant for classification can be used for clustering (type 2) and finally comparing with an adaptation approach actually designed for clustering (type 3). Note that all these three types of approaches do require demarcation of the domains present in the dataset, whereas we do not use such demarcations and consider the dataset holistically. Finally we also analyze the impact of design choices used in our approach and discuss their relative merits. For all these experiments, we extracted the four features (and hence four domains D_1 to D_4) mentioned in Section III-B from the dataset images.

A. Object Clustering

For clustering objects, we used the Berkeley Office dataset [13] that contains 31 categories of daily use objects such as chairs, keyboard etc. collected from three domains namely Amazon, Dslr, Webcam. The domain shift variations include,

Method	Face clustering accuracy (mean±std. deviation)
SSC	56.36±1.60
LRR	49.66±5.23
CO-SSC	55.91±2.22
CO-LRR	50.37±3.40
DA-SSC	54.15±2.33
DA-LRR	36.13±0.99
ED-SSC	57.40±2.90
ED-LRR	43.67±2.96
GM-SSC	57.31±2.09
GM-LRR	45.92±4.16
Ours	**75.36±2.89**

TABLE II. COMPARISON OF FACE CLUSTERING ACCURACY ON THE MARYLAND AA01 ADAPTATION DATASET [28]. RESULTS OF OTHER APPROACHES ARE GIVEN FROM [6].

but not limited to, illumination, resolution and viewpoint changes. Figure 1 gives an illustration. We considered the unlabeled data holistically (which forms our D), for all experiment trials in the protocol, and report clustering results in Table I where we obtain the best clustering accuracy. Note that the other approaches do not use multiple features like ours, but the experiment drives home the point that using multiple features does benefit clustering under domain shift, which supports our hypothesis. Moreover our results are without using any domain demarcation, unlike other approaches, and hence represents a harder problem setting.

B. Face Clustering

We then perform face clustering using the UMD-AA01 dataset [28], which was collected on mobile devices for the purpose of active authentication. The dataset contains facial images of 50 users (categories) over 3 sessions corresponding to different illumination conditions. In each session more than 750 sample images are taken from each face. Some sample images from the dataset are provided in Figure 2. We perform clustering on holistic unlabeled data (which forms our D), for all different experiment trials prescribed in the dataset protocol, and we report results in Table II where we again see much better accuracy over other methods. This reinforces our hypothesis on using multiple invariant image features to address the challenging problem of domain adaptive clustering, even while operating on a harder problem setting which does not use domain demarcations.

C. Design Choice Analysis

We now analyze the relative merits of the design choices we made in our approach. We first study the contribution of our modeling strategy (Section III-C) by directly performing clustering on the outputs of our signal coding stage (Section III-B). In other words, instead of pursuing Section III-C we apply clustering algorithms such as SSC and LRR directly on the features from the M domains. For each of the n images in D, we concatenate their respective feature descriptors for the four domains D_1 to D_4 into a long vector of dimensions 4^*d, and give that as the input to SSC and LRR. Those clustering results are inferior to our results by around 35% and 41% respectively. We then study if the global domain-level modeling alone is sufficient. For each of the n images in D,

Fig. 2. Domain shift illustration from the UMD Face dataset [28]. The dataset has extreme changes in illumination and some viewpoint variations.

we project their respective feature descriptors in domains D_1 to D_4 onto the collection 100 subspaces $S^\star$ and concatenate them into a long vector of dimensions 400^*p, and give that as the input to SSC and LRR to perform clustering. Those results were inferior to our results by around 22% and 27% respectively. These studies shed some light on the importance of our modeling strategies.

We then reduced the choice of features used in Section III-B by using only two of the four and three of the four features, by trying all different feature combinations, and with the rest of the approach unchanged. We saw the results reduce by around 29% and 32% respectively on average. These results point at the possible improvement in accuracy when much more features are used. Note that we need atleast two features (domains) to facilitate our signal modeling strategy (Section III-C) and hence we did not experiment using a single feature.

In all these experiments, we set the codebook dimensions d for all the domains D_i as 800, number of points sampled between a pair of subspace points (S_i, S_j) as 10, PCA energy to determine subspace dimensions as 90%, and the number of points randomly sampled from the intrinsically learned Gaussian distribution as 100. The results in Tables I and II are averaged over 200 different trials of random sampling on the Gaussian, and the low standard deviation suggests the robustness of our approach. Furthermore, we experimented with values of 600, 700, 900 and 1000 for d, 5, 8, 12 and 15 for the number of points sampled between each (S_i, S_j) pair, 80%, 85%, and 95% for the PCA energy, and 70, 90, 110, 130 for the number of random samples from the Gaussian, and the results reduced at the most by 4%, which is still a much better result than other existing approaches. These studies provide some confidence on the generalizability of our approach.

D. Discussion

While domain adaptation in classification scenarios has seen tremendous improvement over the years, with results on

standard datasets approaching 90% accuracy [2], clustering under domain shift is still far from being solved. While this sheds light on the importance of having atleast a few data labels that provide ground truth on how the domain shift has affected the data, it calls for a more concerted effort in addressing scenarios without any data labels as the data from many practical industry applications are fully unlabeled.

We now outline some strategies that may be useful to further our understanding of this problem. (i) One option is to incorporate coarse labels within the adaptive clustering approach to get a flavor of (weak) supervision. Each feature representation does provide some cues on classifying the data into different categories, and while such information could be noisy, we can use them as initial label estimates and refine the clustering process. (ii) While we have represented known domain shift factors using features, there are more involved models explaining different image variations [19] and pursuing such approaches could result in a more formal treatment of the problem. Integrating such models could also minimize the propagation of errors, especially since the signal coding step in the current approach is disjoint from the subsequent modeling process. (iii) Finally, heterogeneous signal modalities can be used, when available, to provide additional confidence since if domain shift impacts one modality much more than the others then we stand to gain by considering such information. Although there exists some studies on heterogeneous adaptation [33], they do not provide a broad treatment of practical use case scenarios.

V. Conclusion

We approached the challenging (and under-studied) problem of clustering under domain shift by borrowing results from invariant feature descriptors to provide coarse robustness to the domain change, and then using geometric modeling strategies to integrate such information to group the unlabeled data into their respective categories. While our results show a marked improvement over existing approaches on two public image datasets, we are still far away from solving this problem especially when compared to the classification version of domain adaptation where we have partial labeled data providing correspondence on the effect of domain shift. We hope the studies reported in this paper will provide some impetus to address this practically important problem in more depth.

References

[1] H. Daume III and D. Marcu, "Domain adaptation for statistical classifiers," *Journal of Artificial Intelligence Research*, vol. 26, pp. 101–126, 2006.

[2] V. M. Patel, R. Gopalan, R. Li, and R. Chellappa, "Visual domain adaptation: A survey of recent advances," *IEEE signal processing magazine*, vol. 32, pp. 53–69, 2015.

[3] Y. Aytar and A. Zisserman, "Tabula rasa: Model transfer for object category detection," in *International Conference on Computer Vision*, 2011, pp. 2252–2259.

[4] C. Cortes and M. Mohri, "Domain adaptation in regression," in *International Conference on Algorithmic Learning Theory*, 2011, pp. 308–323.

[5] R. Gopalan, R. Li, and R. Chellappa, "Unsupervised adaptation across domain shifts by generating intermediate data representations," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, pp. 2288–2302, 2014.

[6] M. Abavisani and V. M. Patel, "Domain adaptive subspace clustering," in *BTAS, Biometrics, Theory and Systems*, 2016.

[7] J. Hoffman, B. Kulis, T. Darrell, and K. Saenko, "Discovering latent domains for multisource domain adaptation," in *ECCV*, 2012, pp. 702–715.

[8] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Transactions on Information Theory*, vol. 29, pp. 439–441, 1983.

[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, pp. 91–110, 2004.

[10] H. Wang, S. Z. Li, and Y. Wang, "Face recognition under varying lighting conditions using self quotient image," in *Automatic Face and Gesture Recognition*, 2004, pp. 819–824.

[11] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE signal processing magazine*, vol. 13, pp. 43–64, 1996.

[12] P.-A. Absil, R. Mahony, and R. Sepulchre, "Riemannian geometry of grassmann manifolds with a view on algorithmic computation," *Acta Applicandae Mathematica*, vol. 80, pp. 199–220, 2004.

[13] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *ECCV*, 2010, pp. 213–226.

[14] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *ICML*, 2011, pp. 513–520.

[15] L. Bruzzone and M. Marconcini, "Domain adaptation problems: A dasvm classification technique and a circular validation strategy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 770–787, 2010.

[16] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, pp. 179–187, 1962.

[17] D. Forsyth, J. L. Mundy, A. Zisserman, C. Coelho, A. Heller, and C. Rothwell, "Invariant descriptors for 3 d object recognition and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 971–991, 1991.

[18] M. Riesenhuber and T. Poggio, "Models of object recognition," *Nature neuroscience*, vol. 3, pp. 1199–1204, 2000.

[19] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Generative models for recognition under variable pose and illumination," in *Automatic Face and Gesture Recognition*. IEEE, 2000, pp. 277–284.

[20] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," *Foundations and trends® in computer graphics and vision*, vol. 3, no. 3, pp. 177–280, 2008.

[21] F. Jurie and B. Triggs, "Creating efficient codebooks for visual recognition," in *ICCV*, 2005, pp. 604–610.

[22] B. Besbes, A. Rogozan, and A. Bensrhair, "Pedestrian recognition based on hierarchical codebook of surf features in visible and infrared images," in *Intelligent Vehicles Symposium*, 2010, pp. 156–161.

[23] H. Cheng and X. Shi, "A simple and effective histogram equalization approach to image enhancement," *Digital signal processing*, vol. 14, pp. 158–170, 2004.

[24] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *International Conference on Image and Signal Processing*, 2008, pp. 236–243.

[25] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, pp. 71–86, 1991.

[26] K. A. Gallivan, A. Srivastava, X. Liu, and P. Van Dooren, "Efficient algorithms for inferences on grassmann manifolds," in *Statistical Signal Processing Workshop*, 2003, pp. 315–318.

[27] Y. Chikuse, *Statistics on special manifolds*. Springer Science & Business Media, 2012, vol. 174.

[28] H. Zhang, V. M. Patel, S. Shekhar, and R. Chellappa, "Domain adaptive sparse representation-based classification," in *Automatic Face and Gesture Recognition*, 2015, pp. 1–8.

[29] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *CVPR*, 2009, pp. 2790–2797.

[30] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 171–184, 2013.

[31] H. Daumé III, "Frustratingly easy domain adaptation," *ACL*, 2007.

[32] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," *AAAI*, 2016.

[33] W. Li, L. Duan, D. Xu, and I. W. Tsang, "Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, pp. 1134–1148, 2014.