Domain Adaptation for Visual Recognition

Raghuraman Gopalan
AT&T Labs-Research
raghuram@research.att.com

Ruonan Li
Harvard University
ruonanli@seas.harvard.edu

Vishal M. Patel
University of Maryland
pvishalm@umd.edu

Rama Chellappa
University of Maryland
rama@umiacs.umd.edu

the essence of knowledge
Boston — Delft

Foundations and Trends® in Computer Graphics and Vision

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

R. Gopalan, R. Li, V.M. Patel and R. Chellappa. *Domain Adaptation for Visual Recognition*. Foundations and Trends® in Computer Graphics and Vision, vol. 8, no. 4, pp. 285–378, 2012.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-68083-031-6

Foundations and Trends® in Computer Graphics and Vision
Volume 8, Issue 4, 2012

Editorial Scope

Topics

Foundations and Trends® in Computer Graphics and Vision publishes survey and tutorial articles in the following topics:

- Rendering
- Shape
- Mesh simplification
- Animation
- Sensors and sensing
- Image restoration and enhancement
- Segmentation and grouping
- Feature detection and selection
- Color processing
- Texture analysis and synthesis
- Illumination and reflectance modeling
- Shape representation
- Tracking
- Calibration
- Structure from motion
- Motion estimation and registration
- Stereo matching and reconstruction
- 3D reconstruction and image-based modeling
- Learning and statistical methods
- Appearance-based matching
- Object and scene recognition
- Face detection and recognition
- Activity and gesture recognition
- Image and video retrieval
- Video analysis and event recognition
- Medical image analysis
- Robot localization and navigation

Information for Librarians

Foundations and Trends® in Computer Graphics and Vision, 2012, Volume 8, 4 issues. ISSN paper version 1572-2740. ISSN online version 1572-2759. Also available as a combined paper and online subscription.

Foundations and Trends® in
Computer Graphics and Vision
Vol. 8, No. 4 (2012) 285–378

DOI: 10.1561/0600000057

Domain Adaptation for Visual Recognition

Raghuraman Gopalan
AT&T Labs-Research
raghuram@research.att.com

Ruonan Li
Harvard University
ruonanli@seas.harvard.edu

Vishal M. Patel
University of Maryland
pvishalm@umd.edu

Rama Chellappa
University of Maryland
rama@umiacs.umd.edu

Contents

1 Introduction **2**

2 Unsupervised Adaptation **8**
2.1 Overview 8
2.2 Subspace Approaches 10
2.3 Low-rank Representations 17
2.4 Discussion 21
2.5 Evaluations 24

3 Semi-supervised Adaptation **30**
3.1 Overview 30
3.2 Transform-based Approaches 32
3.3 Max-margin Techniques 34
3.4 Dictionary-based Approaches 36
3.5 Boosting and Attributes 41
3.6 Discussion 42
3.7 Evaluations 43

4 Adaptation with Multiple, Heterogeneous Domains **51**
4.1 Multiple Domain Adaptation 51
4.2 Heterogeneous Domain Adaptation 57
4.3 Evaluations 61

5 Recent Advances and Open Problems **64**
5.1 Adaptation for Detection and Continuous Estimation . . . 64
5.2 Scalable Adaptation in the Era of Big Visual Data 70
5.3 Open-set Adaptation 75
5.4 Statistical and Physically Informed Adaptation 77

6 Conclusion **80**

References **83**

Abstract

Domain adaptation is an active, emerging research area that attempts to address the changes in data distribution across training and testing datasets. With the availability of a multitude of image acquisition sensors, variations due to illumination, and viewpoint among others, computer vision applications present a very natural test bed for evaluating domain adaptation methods. In this monograph, we provide a comprehensive overview of domain adaptation solutions for visual recognition problems. By starting with the problem description and illustrations, we discuss three adaptation scenarios namely, (i) unsupervised adaptation where the "source domain" training data is partially labeled and the "target domain" test data is unlabeled, (ii) semi-supervised adaptation where the target domain also has partial labels, and (iii) multi-domain heterogeneous adaptation which studies the previous two settings with the source and/or target having more than one domain, and accounts for cases where the features used to represent the data in each domain are different. For all these topics we discuss existing adaptation techniques in the literature, which are motivated by the principles of max-margin discriminative learning, manifold learning, sparse coding, as well as low-rank representations. These techniques have shown improved performance on a variety of applications such as object recognition, face recognition, activity analysis, concept classification, and person detection. We then conclude by analyzing the challenges posed by the realm of "big visual data", in terms of the generalization ability of adaptation algorithms to unconstrained data acquisition as well as issues related to their computational tractability, and draw parallels with the efforts from vision community on image transformation models, and invariant descriptors so as to facilitate improved understanding of vision problems under uncertainty.

R. Gopalan, R. Li, V.M. Patel and R. Chellappa. *Domain Adaptation for Visual Recognition*. Foundations and Trends® in Computer Graphics and Vision, vol. 8, no. 4, pp. 285–378, 2012.
DOI: 10.1561/0600000057.

1

Introduction

Over the last few years, we have witnessed a widespread impact of computer vision techniques in practical applications pertaining to surveillance, robotics, human computer interaction and user content personalization. Typical examples include biometric authentication using face, iris, fingerprint and gait, object localization and scene understanding for autonomous agents, human gesture interpretation systems such as Kinect, and visual analytic apps from web-scale images and videos. While the foundations for these techniques date back to atleast three decades ago, the main catalyst enabling the transition of these methods to real applications is the availability of data, which from the early 2000's has seen an exponential increase in part due to the widespread availability of cameras. The performance improvement facilitated by large quantity of data has been well documented in several computer vision applications that involve unconstrained variations in the entities of interest. Examples range from face verification on the LFW dataset (Huang et al., 2007; Taigman et al., 2014), the Pascal VOC challenge for object recognition (Everingham et al., 2010; Girshick et al., 2014), activity analysis on the HMDB dataset (Kuehne et al., 2011; Shao et al., 2014), and the MIT scene categorization challenge (Xiao et al., 2010, 2014), to name a few, where we have seen substantial gains by properly harnessing the information conveyed by data using modeling tools that are relevant to the specific problem of interest.

While the benefits derived from data-centric approaches are many, they however come with their own set of problems which have begun to surface in the recent past. Primary among them are (i) the variations in data properties when obtained from different sources, even for a specific data category, (ii) imbalance in the amount of data obtained for different data categories, since certain categories are more common than others, and (iii) the absence of category labels and/or the presence of noisy category labels since vast quantities of data available from the web are unstructured. There have been several attempts in the literature to address these issues, and in particular, the problem of change in data properties acquired from different sources is tackled by domain adaptation (Jiang, 2008). An example would be the pictures of Eiffel Tower obtained from expert photographers when compared with those obtained from casual visitors. In this case, although the scene that is being captured is the same, the properties of the images could vary vastly due to differences in capabilities. Hence to perform visual recognition on data from these different sources or domains, it is important to account for the change in data distribution. Transfer learning, on the other hand, deals with the notion of transferring the information learnt on some data categories to other/newer categories which may not have sufficient amount of data to begin with (Pan and Yang, 2010). For one interested in studying visual appearance of mammals, a typical example could be to learn properties of image categories corresponding to tiger, lion, antelope and cow, which are common mammals, and utilize them to perform inference on Saola which is one of the world's rarest mammals for which one may not have enough data.

While transfer learning and domain adaptation problems originate from a distribution mismatch between the source and target data, the underlying causes for such mismatches are traditionally considered different. Thus, even though transfer learning and domain adaptation algorithms are designed to address different issues, one might argue that these problems are just different manifestations of learning to learn, i.e. the ability to leverage over prior knowledge when attempting to solve a new task, such as the one studied by a recent work from Patricia and Caputo (2014). Practical applications usually involve challenges overlapping these problems, thereby giving rise to techniques to deal with them in unison. Hence for the sake of clarity, the primary focus of this monograph will be on domain adaptation while accommodating some key efforts pertaining to transfer learning and big data techniques with an adaptation flavor.

Circumstances requiring domain adaptation arise very naturally in visual recognition, where the change in data distribution is caused by variations in lighting conditions, viewpoint, blur, resolution, and occlusion, in addition to different types of imaging sensors such as RGB, RGB-D, and infrared among others. Initial attempts for addressing this problem started around the year 2010 in the context of object recognition, and since then there have been several efforts that expanded to problems involving faces, events, concepts, activity and attributes in general. The technical approaches proposed for these problems derive motivation from several existing modeling frameworks such as max-margin learning, transform coding, manifolds, and dictionary learning, where the broad goal is to modify the cost function of these frameworks to account for the change in data distribution (or the domain shift) between the training and testing datasets (or the source and target domains). In doing so, an inherent assumption is that each domain contains data drawn from a similar distribution, for instance the source domain consisting of objects imaged under ambient lighting while the target domain contains same objects captured under low light.

In the following chapters we discuss these techniques by grouping them into three adaptation settings. For all these settings the source domain is either fully labeled, or is partially labeled with labels available for all data categories. Hence one can think of the source domain as the reliable data prepared under human supervision for the application of interest, using which the inference needs to be performed on the target domain. We first begin with *unsupervised adaptation* in Chapter 2 where the target domain is completely unlabeled. This could correspond to scenarios where the test data comes from real-time feeds for which it is not possible to obtain labels beforehand. We then discuss *semi-supervised adaptation* in Chapter 3 that considers the target domain to have partial labels. This could correspond to cases where some human effort has been spent on labeling a new data corpus, or when the data itself has weak supervision in the form of text labels accompanying images. The advantage of this setting over the unsupervised one is the presence of valuable correspondence information on how data belonging to (some) categories has transformed across the domains. While for these two settings we have assumed there is only one domain in both the source and target, there could be cases with multiple domains for either. A typical scenario would be to use the data crawled from web image collections as the source domain. Since the properties of such data could vary greatly as the images may be acquired from dslr cameras, webcams, hand drawings, and paintings, one needs to separate them into multiple similar-looking source domains using which adaptation can be performed to infer the target domain data. We discuss such *multi-domain adaptation* settings for both unsupervised and semi-supervised adaptation in Chapter 4, where we also consider *heterogeneous* adaptation where the feature types and dimensionalities for data in each domain could be different. One practical example is to perform adaptation between depth images and intensity images, given the increased availability of RGB-D sensors such as Kinect.

Finally, with the emergence of "big visual data" the role of adaptation becomes increasingly important in extracting pertinent information from a humongous data corpus that would positively contribute to the final problem objective. This is critical since there are studies that suggest training with increased data quantities alone may not guarantee a good performance, and that if data is not utilized in the right way it could actually be detrimental to the objective. Hence a concerted effort is required to address these issues from both a computational aspect as well as from a generalization standpoint, in being robust across possible variations in data and to accommodate the presence of new categories in the test set that are not present during training (referred to as the open-set problem). Moreover, with the utility of adaptation extending beyond visual recognition to problems such as detection, continuous parameter estimation, reconstruction and segmentation among others, it is only natural to encode the valuable information conveyed by image transformation models that have been studied for several decades in the vision community. A vast majority of existing approaches tackle the adaptation problem in a pure statistical sense by extracting or learning features from the image data and minimizing the domain shift with respect to classification accuracy. While this could be due to the nature of unconstrained data variations, for which the assumptions inherent to a model-based treatment may be restrictive, one nevertheless stands to gain by integrating the data-driven techniques with pertinent model-induced geometry as it has been shown to have the potential for accomplishing more with less data. We discuss recent advances related to such themes in Chapter 5, and draw conclusions in Chapter 6.

While our main focus in this monograph will be on approaches proposed for visual recognition, in the discourse we will also discuss earlier work of domain adaptation used for other signal modalities such as natural language and speech. In support of the conceptual discussion, we performed an empirical comparison of a couple of language modeling adaptation algorithms on the visual office object recognition dataset. By doing so we attempt to answer the question whether there are unique challenges posed by visual domain shift that require more specialized techniques than those used in other signal modalities. Last but least, we will focus throughout the discussion on the efficiency and scalability of different approaches, and analyze how the algorithms scale with data size.

Notations. We refer to the training dataset with plenty of labeled data as the source domain and the test dataset with a few labeled data or no labeled data as the target domain. Let $\mathcal{S} = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^{N_s}$, where $\mathbf{x}^s \in^N$ denote the labeled data from the source domain. Here, $\mathbf{x}^s$ is referred to as an observation and y^s is the corresponding class label. Labeled data from the target domain is denoted by $\mathcal{T}_l = \{(\mathbf{x}_i^{tl}, y_i^{tl})\}_{i=1}^{N_{tl}}$ where $\mathbf{x}^{tl} \in^M$. Similarly, unlabeled data in the target domain is denoted by $\mathcal{T}_u = \{\mathbf{x}_i^{tu}\}_{i=1}^{N_{tu}}$ where $\mathbf{x}^{tu} \in^M$. Unless specified otherwise, we assume $N = M$ and $N_s \gg N_{tl}$ in general. Let $\mathcal{T} = \mathcal{T}_l \cup \mathcal{T}_u$. As a result, the total number of samples in the target domain is denoted by N_t which is equal to $N_{tl} + N_{tu}$. Denote $\mathbf{S} = [\mathbf{x}_1^s, \cdots, \mathbf{x}_{N_s}^s]$ as the matrix of N_s data points from $\mathcal{S}$. Denote $\mathbf{T}_l = [\mathbf{x}_1^{tl}, \cdots, \mathbf{x}_{N_{tl}}^{tl}]$ as the matrix of N_{tl} data from $\mathcal{T}_l$, $\mathbf{T}_u = [\mathbf{x}_1^{tu}, \cdots, \mathbf{x}_{N_{tu}}^{tu}]$ as the matrix of N_{tu} data from $\mathcal{T}_u$ and $\mathbf{T} = [\mathbf{T}_l | \mathbf{T}_u] = [\mathbf{x}_1^t, \cdots, \mathbf{x}_{N_t}^t]$ as the matrix of N_t data from $\mathcal{T}$. It is assumed that both the target and source data pertain to C classes or categories.

References

O. Aghazadeh and S. Carlsson. Properties of datasets predict the performance of classifiers. In *British Machine Vision Conference*, pages 22–31, 2013a.

O. Aghazadeh and S. Carlsson. Datasets, sample selection, and test performance. In *The First International Workshop on Visual Domain Adaptation and Dataset Bias*, 2013b.

M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54:4311 – 4322, November 2006.

T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila. Recognition of blurred faces using local phase quantization. In *International Conference on Pattern Recognition*, pages 1–4, 2008.

Y. Aytar and A. Zisserman. Tabula rasa: Model transfer for object category detection. In *IEEE International Conference on Computer Vision*, pages 2252–2259, 2011.

M. Baktashmotlagh, M.T. Harandi, B.C. Lovell, and M. Salzmann. Unsupervised domain adaptation by domain invariant projection. In *IEEE International Conference on Computer Vision*, pages 769 – 776, 2013.

C.J. Becker, C.M. Christoudias, and P. Fua. Non-linear domain adaptation with boosting. In *Advances in Neural Information Processing Systems*, pages 485–493, 2013.

S. Ben-David and R. Urner. Domain adaptation as learning with auxiliary information. In *Advances in Neural Information Processing Systems*, 2013.

S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira. Analysis of representations for domain adaptation. *Advances in Neural Information Processing Systems*, 19:137–143, 2007.

S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J.W. Vaughan. A theory of learning from different domains. *Machine Learning*, 79:151–175, January 2010.

Y. Bengio. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1):1–127, 2009.

Y. Bengio. Deep learning of representations for unsupervised and transfer learning. *Journal of Machine Learning Research*, 27:17–36, 2012.

T.L. Berg, A.C. Berg, and J. Shih. Automatic attribute discovery and characterization from noisy web data. In *European Conference on Computer Vision*, pages 663–676. 2010.

A. Bergamo and L. Torresani. Exploiting weakly-labeled web images to improve object classification: a domain adaptation approach. In *Advances in Neural Information Processing Systems*, pages 181–189, 2010.

S. Biswas, G. Aggarwal, and R. Chellappa. Robust estimation of albedo for illumination-invariant matching and shape recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:884–899, May 2009.

J. Blitzer, R. McDonald, and F. Pereira. Domain adaptation with structural correspondence learning. In *Conference on Empirical Methods in Natural Language Processing*, pages 120–128, 2006.

J. Blitzer, S. Kakade, and D.P. Foster. Domain adaptation with coupled subspaces. In *International Conference on Artificial Intelligence and Statistics*, pages 173–181, 2011.

J. Bruna and S. Mallat. Invariant scattering convolution networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:1872–1886, August 2013.

O. Chapelle, B. Schölkopf, and A. Zien. *Semi-supervised learning*, volume 2. MIT press Cambridge, 2006.

L. Chen, L. Duan, D. Xu, and D. Xu. Event recognition in videos by learning from heterogeneous web sources. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2666–2673, 2013.

M. Chen, K.Q. Weinberger, and J. Blitzer. Co-training for domain adaptation. In *Advances in Neural Information Processing Systems*, pages 2456–2464, 2011.

Y. Chikuse. *Statistics on special manifolds*, volume 174 of *Lecture Notes in Statistics*. Springer Verlag, 2003.

S. Chopra, S. Balakrishnan, and R. Gopalan. Dlid: Deep learning for domain adaptation by interpolating between domains. In *ICML Workshop on Representation Learning*, 2013.

W-S. Chu, F. De la Torre, and J.F. Cohn. Selective transfer machine for personalized facial action unit detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3515–3522, 2013.

K. Crammer, M. Kearns, and J. Wortman. Learning from multiple sources. *Journal of Machine Learning Research*, 9:1757–1774, 2008.

W. Dai, Gui-Rong Xue, Q. Yang, and Y. Yu. Transferring naive Bayes classifiers for text classification. In *National Conference on Artificial Intelligence*, pages 540–545, 2007.

W. Dai, Y. Chen, Gui-Rong Xue, Q. Yang, and Y. Yu. Translated learning: Transfer learning across different feature spaces. In *Advances in Neural Information Processing Systems*, pages 353–360, 2008.

H. Daumé III. Frustratingly easy domain adaptation. In *ACL*, 2007.

H. Daumé III, A. Kumar, and A. Saha. Frustratingly easy semi-supervised domain adaptation. In *Workshop on Domain Adaptation for Natural Language Processing*, pages 53–59, 2010.

J. Donahue, J. Hoffman, E. Rodner, K. Saenko, and T. Darrell. Semi-supervised domain adaptation with instance constraints. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 668–675, 2013.

L. Duan, I.W. Tsang, D. Xu, and Tat-Seng Chua. Domain adaptation from multiple sources via auxiliary classifiers. In *International Conference on Machine Learning*, pages 289–296, 2009a.

L. Duan, I.W. Tsang, D. Xu, and S.J. Maybank. Domain transfer SVM for video concept detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1375–1381, 2009b.

L. Duan, I.W. Tsang, and D. Xu. Domain transfer multiple kernel learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34:465–479, March 2012a.

L. Duan, D. Xu, and Shih-Fu Chang. Exploiting web images for event recognition in consumer videos: A multiple source domain adaptation approach. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1338–1345, 2012b.

L. Duan, D. Xu, and I. Tsang. Learning with augmented features for heterogeneous domain adaptation. In *International Conference on Machine Learning*, 2012c.

L. Duan, D. Xu, Ivor Wai-Hung Tsang, and J. Luo. Visual event recognition in videos by learning from web data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34:1667–1680, September 2012d.

M. Elad. *Sparse and Redundant Representations: From theory to applications in Signal and Image processing*. Springer, 2010.

M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15:3736–3745, December 2006.

M. Elad, M. A. T. Figueiredo, and Y. Ma. On the role of sparse and redundant representations in image processing. *Proceedings of the IEEE*, 98:972–982, June 2010.

M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL visual object classes (voc) challenge. *International Journal of Computer Vision*, 88:303–338, June 2010.

B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars. Unsupervised visual domain adaptation using subspace alignment. In *IEEE International Conference on Computer Vision*, pages 2960 – 2967, 2013.

D. Forsyth, J.L. Mundy, A. Zisserman, C. Coelho, A. Heller, and C. Rothwell. Invariant descriptors for 3D object recognition and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:971–991, October 1991.

R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587, 2014.

B. Gong, Y. Shi, F. Sha, and K. Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2066–2073, 2012.

B. Gong, K. Grauman, and F. Sha. Reshaping visual datasets for domain adaptation. In *Advances in Neural Information Processing Systems*, pages 1286–1294, 2013a.

B. Gong, K. Grauman, and F. Sha. Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation. In *International Conference on Machine Learning*, pages 222–230, 2013b.

R. Gopalan. Learning cross-domain information transfer for location recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 731–738, 2013.

R. Gopalan, R. Li, and R. Chellappa. Domain adaptation for object recognition: An unsupervised approach. In *IEEE International Conference on Computer Vision*, pages 999–1006, 2011.

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*, 36:2288–2302, November 2014.

R. Gross, I. Matthews, and S. Baker. Appearance-based face recognition and light-fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:449–465, April 2004.

H.-T. Ho and R. Gopalan. Model-driven domain adaptation on product manifolds for unconstrained face recognition. *International Journal of Computer Vision*, 109:110–125, 2014.

J. Hoffman, B. Kulis, T. Darrell, and K. Saenko. Discovering latent domains for multisource domain adaptation. In *European Conference on Computer Vision*, pages 702–715, 2012.

J. Hoffman, E. Rodner, J. Donahue, K. Saenko, and T. Darrell. Efficient learning of domain-invariant image representations. In *International Conference on Learning Representations*, 2013a.

J. Hoffman, E. Tzeng, J. Donahue, Y. Jia, K. Saenko, and T. Darrell. One-shot adaptation of supervised deep convolutional models. *ArXiv e-prints*, December 2013b.

G.B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

V. Jain and S.S. Farfade. Adapting classification cascades to new domains. In *IEEE International Conference on Computer Vision*, pages 105–112, 2013.

V. Jain and E. Learned-Miller. Online domain adaptation of a pre-trained cascade of classifiers. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 577–584, 2011.

N. Japkowicz and S. Stephen. The class imbalance problem: A systematic study. *Intelligent data analysis*, 6:429–449, May 2002.

I. Jhuo, D. Liu, D.T. Lee, and Shih-Fu Chang. Robust visual domain adaptation with low-rank reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2168–2175, 2012.

J. Jiang. A literature survey on domain adaptation of statistical classifiers. *URL: http://sifaka. cs. uiuc. edu/jiang4/domainadaptation/survey*, 2008.

L. Jie, T. Tommasi, and B. Caputo. Multiclass transfer learning from unconstrained priors. In *IEEE International Conference on Computer Vision*, pages 1863–1870, 2011.

S.M. Kakade, S. Shalev-Shwartz, and A. Tewari. Regularization techniques for learning with matrices. *The Journal of Machine Learning Research*, 13: 1865–1890, January 2012.

H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30:509–541, May 1977.

A. Khosla, T. Zhou, T. Malisiewicz, A.A. Efros, and A. Torralba. Undoing the damage of dataset bias. In *European Conference on Computer Vision*, pages 158–171. 2012.

A. Kovashka and K. Grauman. Attribute adaptation for personalized image search. In *IEEE International Conference on Computer Vision*, pages 3432–3439, 2013.

H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre. Hmdb: a large video database for human motion recognition. In *IEEE International Conference on Computer Vision*, pages 2556–2563, 2011.

B. Kulis, K. Saenko, and T. Darrell. What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1785–1792, 2011.

I. Kuzborskij and F. Orabona. Stability and hypothesis transfer learning. In *International Conference on Machine Learning*, pages 942–950, 2013.

C.H. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 951–958, 2009.

A. Lapedriza, H. Pirsiavash, Z. Bylinskii, and A. Torralba. Are all training examples equally valuable? *arXiv:1311.6510*, 2013.

K-C. Lee, J. Ho, and D. Kriegman. Nine points of light: Acquiring subspaces for face recognition under variable lighting. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 512–519, 2001.

H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang. Probabilistic elastic part model for unsupervised face detector adaptation. In *IEEE International Conference on Computer Vision*, pages 793–800, 2013a.

H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang. Probabilistic elastic matching for pose variant face verification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3499–3506, 2013b.

R. Li and T. Zickler. Discriminative virtual views for cross-view action recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2855–2862, 2012.

W. Li, L. Duan, D. Xu, and I. Tsang. Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1–1, 2013c.

X. Li and J. Bilmes. A bayesian divergence prior for classiffier adaptation. In *International Conference on Artificial Intelligence and Statistics*, pages 275–282, 2007.

J. Liu, M. Shah, B. Kuipersy, and S. Savarese. Cross-view action recognition via view knowledge transfer. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3209–3216, 2011.

Y. Mansour, M. Mohri, and A. Rostamizadeh. Domain adaptation with multiple sources. In *Advances in Neural Information Processing Systems*, pages 1041–1048, 2009.

F. Mirrashed, V.I. Morariu, and L.S. Davis. Sampling for unsupervised domain adaptive object detection. In *International Conference on Image Processing*, pages 323–330, 2013.

E. Morvant, A. Habrard, and S. Ayache. Parsimonious unsupervised and semi-supervised domain adaptation with good similarity functions. *Knowledge and Information Systems*, 33:309–349, February 2012.

H. Nguyen, V. Patel, N. Nasrabadi, and R. Chellappa. Sparse embedding: A framework for sparsity promoting dimensionality reduction. In *European conference on Computer vision*, pages 414–427, 2012.

J. Ni, Q. Qiu, and R. Chellappa. Subspace interpolation via dictionary learning for unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 692–699, 2013.

J. Niu, Z. Li, and G. Salvendy. Mathematical methods for shape analysis and form comparison in 3D anthropometry: a literature review. In *Digital Human Modeling*, pages 161–170. Springer, 2007.

B.A Olshausen and D.J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research*, 37(23):3311–3325, 1997.

S.J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22:1345–1359, October 2010.

S.J. Pan, I.W. Tsang, J.T. Kwok, and Q. Yang. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, 22:199–210, February 2011.

V. M. Patel and R. Chellappa. *Sparse representations and compressive sensing for imaging and vision*. SpringerBriefs, 2013.

N. Patricia and B. Caputo. Learning to learn, from transfer learning to domain adaptation: A unifying perspective. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1442–1449, 2014.

G. Patterson and J. Hays. Sun attribute database: Discovering, annotating, and recognizing scene attributes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2751–2758, 2012.

Q. Qiu and R. Chellappa. Compositional dictionaries for domain adaptive face recognition. *CoRR*, abs/1308.0271, 2013.

Q. Qiu, V.M. Patel, P. Turaga, and R. Chellappa. Domain adaptive dictionary learning. In *European Conference on Computer Vision*, pages 631–645. 2012.

A. Quattoni, M. Collins, and T. Darrell. Transfer learning for image classification with sparse prototype representations. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

J. Quionero-Candela, M. Sugiyama, A. Schwaighofer, and N.D. Lawrence. *Dataset Shift in Machine Learning*. The MIT Press, 2009. ISBN 0262170051, 9780262170055.

K. Rematas, B. Fernando, T. Tommasi, and T. Tuytelaars. Does evolution cause a domain shift? In *The First International Workshop on Visual Domain Adaptation and Dataset Bias*, 2013.

E. Rodner, J. Hoffman, J. Donahue, T. Darrell, and K. Saenko. Towards adapting imagenet to reality: Scalable domain adaptation with implicit low-rank transformations. *arXiv preprint arXiv:1308.4200*, 2013.

M.T. Rosenstein, Z. Marx, L.P. Kaelbling, and T.G. Dietterich. To transfer or not to transfer. In *In NIPS Workshop, Inductive Transfer: 10 Years Later*, 2005.

A. Roy-Chowdhury, R. Chellappa, and R. Gupta. 3D face modeling from monocular video sequences. *Face Processing: Advanced Modeling and Methods*, 2005.

R. Rubinstein, A.M. Bruckstein, and M. Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98:1045–1057, June 2010.

K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting visual category models to new domains. In *European Conference on Computer Vision*, pages 213–226. 2010.

R. Salakhutdinov, A. Torralba, and J. Tenenbaum. Learning to share visual appearance for multiclass object detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1481–1488, 2011.

W.J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T.E. Boult. Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:1757–1772, July 2013.

L. Shao, X. Zhen, D. Tao, and X. Li. Spatio-temporal laplacian pyramid coding for action recognition. *IEEE Transactions on Cybernetics*, 44:817–827, June 2014.

A. Sharma and D.W. Jacobs. Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, 2011.

A. Sharma, A. Kumar, H. Daume, and D.W. Jacobs. Generalized multiview analysis: A discriminative latent space. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2160–2167, 2012.

D.A. Shaw and R. Chellappa. Combined direction estimation for dimension reduction in the presence of inhomogeneous data. *JASA, Under Review*.

S. Shekhar, V.M. Patel, H.V. Nguyen, and R. Chellappa. Generalized domain-adaptive dictionaries. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 361–368, 2013.

X. Shen, Z. Lin, J. Brandt, and Y. Wu. Detecting and aligning faces by image retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3460 – 3467, 2013.

X. Shi, Q. Liu, W. Fan, P.S. Yu, and R. Zhu. Transfer learning on heterogenous feature spaces via spectral transformation. In *International Conference on Data Mining*, pages 1049–1054, 2010.

Y. Shi and F. Sha. Information-theoretical learning of discriminative clusters for unsupervised domain adaptation. In *International Conference on Machine Learning*, pages 1079–1086, 2012.

H. Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of statistical planning and inference*, 90:227–244, February 2000.

T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:1615–1618, December 2003.

J. Sun, X. Wu, S. Yan, Loong-Fah Cheong, Tat-Seng Chua, and J. Li. Hierarchical spatio-temporal context modeling for action recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2004–2011, 2009.

Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, 2014.

K.D. Tang, V. Ramanathan, Fei-Fei Li, and D. Koller. Shifting weights: Adapting object detectors from image to video. In *Advances in Neural Information Processing Systems*, pages 647–655, 2012.

T. Tommasi and B. Caputo. Frustratingly easy NBNN domain adaptation. In *IEEE International Conference on Computer Vision*, pages 897 – 904, 2013.

T. Tommasi, F. Orabona, and B. Caputo. Safety in numbers: Learning categories from few examples with multi model knowledge transfer. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3081–3088, 2010.

A. Torralba and A.A. Efros. Unbiased look at dataset bias. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1521–1528, 2011.

G. Tur. Co-adaptation: Adaptive co-training for semi-supervised learning. In *IEEE Conference on Acoustics, Speech and Signal Processing*, pages 3721–3724, 2009.

P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa. Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33:2273–2286, November 2011.

T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: a survey. *Foundations and Trends® in Computer Graphics and Vision*, 3(3):177–280, 2008.

D. Vazquez, A. Lopez, J. Marin, D. Ponsa, and D. Geronimo. Virtual and real world adaptation for pedestrian detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1–1, 2013.

P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 503–511, 2001.

C. Wang and S. Mahadevan. Heterogeneous domain adaptation using manifold alignment. In *International Joint Conference on Artificial Intelligence*, pages 1541–1546, 2011.

X. Wang, G. Hua, and T.X. Han. Detection by detections: Non-parametric detector adaptation for a video. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 350–357, 2012.

D. Weinland, E. Boyer, and R. Ronfard. Action recognition from arbitrary views using 3D exemplars. In *IEEE International Conference on Computer Vision*, pages 1–7, 2007.

I. Weiss. Geometric invariants and object recognition. *International Journal of Computer Vision*, 10:207–231, March 1993.

J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:210–227, February 2009.

J. Wright, Y. Ma, J. Mairal, G. Sapiro, T.S. Huang, and S. Yan. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 98:1031–1044, June 2010.

P. Wu and T.G. Dietterich. Improving SVM accuracy by training on auxiliary data sources. In *International conference on Machine learning*, pages 110–116, 2004.

X. Wu, H. Wang, C. Liu, and Y. Jia. Cross-view action recognition over heterogeneous feature spaces. In *IEEE International Conference on Computer Vision*, pages 609–616, 2013.

J. Xiao, J. Hays, K.A. Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3485–3492, 2010.

J. Xiao, K.A. Ehinger, J. Hays, A. Torralba, and A. Oliva. Sun database: Exploring a large collection of scene categories. *International Journal of Computer Vision*, Preprint:1–20, 2014.

D. Xing, W. Dai, Gui-Rong Xue, and Y. Yu. Bridged refinement for transfer learning. In *Knowledge Discovery in Databases*, pages 324–335. 2007.

M. Yamada, L. Sigal, and M. Raptis. No bias left behind: Covariate shift adaptation for discriminative 3D pose estimation. In *European Conference on Computer Vision*, pages 674–687, 2012.

J. Yang, R. Yan, and A.G. Hauptmann. Cross-domain video concept detection using adaptive SVMs. In *ACM International conference on Multimedia*, pages 188–197, 2007.

M. Yang, L. Zhang, X. Feng, and D. Zhang. Fisher discrimination dictionary learning for sparse representation. In *IEEE International Conference on Computer Vision*, pages 543–550, 2011.

Y. Yao and G. Doretto. Boosting for transfer learning with multiple sources. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1855–1862, 2010.

Z. Zhang, C. Wang, B. Xiao, W. Zhou, S. Liu, and C. Shi. Cross-view action recognition via a continuous virtual path. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2690–2697, 2013.

J. Zheng, Z. Jiang, J.P. Phillips, and R. Chellappa. Cross-view action recognition via a transferable dictionary pair. In *British Machine Vision Conference*, pages 7–13, 2012a.

J. Zheng, Ming-Yu Liu, R. Chellappa, and J. Phillips. A Grassmann manifold-based domain adaptation approach. In *International Conference on Pattern Recognition*, pages 2095–2099, 2012b.

X. Zhu, C. Vondrick, D. Ramanan, and C. Fowlkes. Do we need more training data or better models for object detection? In *British Machine Vision Conference*, pages 1–11, 2012.

Y. Zhu, Y. Chen, Z. Lu, S.J. Pan, Gui-Rong Xue, Y. Yu, and Q. Yang. Heterogeneous transfer learning for image classification. In *AAAI Conference on Artificial Intelligence*, 2011.