BALANCED WEIGHTED UNIFIED DISCRIMINANT AND DISTRIBUTION ALIGNMENT FOR OPEN-VIEW HUMAN ACTION RECOGNITION

MOHD SHAH RIZAL BIN SAMSUDIN

UNIVERSITI TEKNOLOGI MALAYSIA

BALANCED WEIGHTED UNIFIED DISCRIMINANT AND DISTRIBUTION ALIGNMENT FOR OPEN-VIEW HUMAN ACTION RECOGNITION

MOHD SHAH RIZAL BIN SAMSUDIN

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy

School of Electrical Engineering
Faculty of Engineering
Universiti Teknologi Malaysia

SEPTEMBER 2022

ACKNOWLEDGEMENT

Praise be to Allah (SWT), the Most Gracious and Most Merciful.

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof Dr. Syed Abdul Rahman bin Syed Abu Bakar for his continuous and immense support, supervision, encouragement, and understanding throughout my study. I would also like to express my gratitude to my Co. Supervisor Assoc. Prof. Dr. Musa bin Mohd Mokji for his guidance, enlightenment, support, and encouragement. Also, to other CVVIP's lecturers, Dr. Usman Ullah Sheikh and Assoc. Prof. Dr. Zaid Omar, for their constructive comments and suggestions during this research work.

Special indebtedness goes to my friends in Universiti Teknologi Malaysia (UTM), especially my Ph.D. mates, Najeeb ur Rehman Malik, Dr. Aliyu Muhammad Abdu and Dr. Ahmed Sabeeh for their assistance, understanding, and moral support.

Special thanks go to the Royal Malaysian Navy for the full-time sponsorship. Dedicated to Rear Admiral Datuk Ir. Ts. Mohd Shaiful Adli Chung for giving me the opportunity and trust to pursue this full-time study. Special thanks to First Admiral Ir. Ts. Franklin J. Joseph for suggesting my name to be a PhD candidate and encouraging me to achieve my dream of becoming a PhD holder.

Finally, heartfelt thanks to my dear wife and children, who had endured the hardships with me while pursuing this prestigious degree. I am also grateful to the rest of my family members for their support, patience, and continuous prayers.

ABSTRACT

Human action recognition (HAR) plays an increasingly important role in surveillance, robot learning, and human-computer interaction. However, there are many challenges and issues involved in achieving reliable and high-performance results. Among these challenges, view-invariant in an uncontrolled dataset where several cameras are placed at different locations received the most attention from researchers. One of the primary concerns for the uncontrolled dataset is the large difference between data distributions at the source (training) and target (testing) views. Such difference causes the data shift problem to occur and hence, decreases the performance of the HAR system. This issue has been explicitly discussed as an open-view HAR problem which aims to reduce the correlation between the source and the target views particularly when labelled data in unavailable in the target view. In addressing the issue, this thesis presents an unsupervised domain-adaptation model for the open-view HAR. Specifically, the proposed Balanced Weighted Unified Discriminant and Distribution Alignment (BW-UDDA) model has managed to handle datasets with significant variances across views. BW-UDDA balances and aligns marginal and conditional distribution features by projecting them into a low-dimensional subspace. This is to create more coordinated feature representations before feeding these features into an optimal classifier. Technically, BW-UDDA exploits two different unsupervised domain adaptation enhancement models, namely Balanced Weighted Joint Geometrical and Statistical Alignment (BW-JGSA) and Unified Discriminant and Distribution Alignment (UDDA). The BW-JGSA balances the marginal and conditional distributions in the nonparametric Maximum Mean Discrepancy (MMD) measurements on two disjointed embedded matrices. For the UDDA, two-dimensionality reduction techniques, namely linear discriminant analysis (LDA) and locality sensitivity discriminant analysis (LSDA), are incorporated to create features with global and local discriminant properties for the domain adaptation process. The enhancement models were evaluated on public image and digit datasets (Office, Caltech-256, USPS, MNIST and COIL20), while the BW-UDDA was assessed using the multi-camera action dataset (MCAD). Both enhancement models outperformed other state-of-the-art methods with average accuracies: 50.61% (object dataset) and 69.95% (digit dataset) for BW-JGSA, and 59.95% (object dataset) and 80.72% (digit dataset) for UDDA, respectively. BW-UDDA for open-view HAR was tested using two types of cross-view evaluations. The average accuracy of the first and second evaluations using the MCAD dataset outperformed the state-of-the-art with 13.38% and 61.45% higher accuracy, respectively. The BW-UDDA was also tested on a controlled multi-camera HAR dataset, the Inria Xmas Motion Acquisition Sequences (IXMAS), with an accuracy of 90.91% using the second type of cross-view evaluation. These results on MCAD and IXMAS confirmed the superiority of the proposed model for the open-view HAR.

ABSTRAK

Pengecaman tindakan manusia (HAR) memainkan peranan yang semakin penting dalam pengawasan, pembelajaran robot dan interaksi manusia-komputer. Walau bagaimanapun, terdapat banyak kekangan dan isu yang dihadapi untuk mencapai keputusan yang boleh disandarkan dan berprestasi tinggi. Antara cabaran yang mendapat perhatian penyelidik adalah isu paparan berbilang dalam set data tidak terkawal menggunakan beberapa kamera di lokasi berbeza. Salah satu masalah utama bagi set data tidak terkawal ialah perbezaan besar di antara taburan data pandangan sumber (latihan) dan sasaran (ujian). Perbezaan ini menyebabkan timbulnya masalah peralihan data, dan sekaligus menjejaskan prestasi sistem HAR. Isu ini telah dibincangkan secara khusus di bawah masalah paparan terbuka HAR, iaitu kes di mana korelasi antara paparan sumber dan sasaran dikurangkan, serta ketidaksediaan data berlabel dalam paparan sasaran. Dalam menangani isu ini, tesis ini membentangkan model penyesuaian domain yang tidak diselia untuk paparan terbuka HAR. Secara khusus, model Penjajaran Diskriminasi dan Pengagihan Bersepadu Wajaran Seimbang (BW-UDDA) mampu mengendalikan set data dengan perbezaan ketara merentas paparan. Pada asasnya, BW-UDDA akan mengimbangi dan menjajarkan ciri taburan marginal dan bersyarat dengan memindahkannya ke dalam subruang dimensi rendah. Ini untuk mencipta perwakilan ciri yang lebih diselaraskan sebelum memasukkan ciri ke dalam pengelasan optimum. Secara teknikal, BW-UDDA mengeksploitasi dua model penyesuaian domain tanpa pengawasan yang dipertingkatkan, dikenali sebagai Penjajaran Geometri dan Statistik Berwajaran Seimbang (BW-JGSA) dan Penjajaran Diskriminasi dan Pengagihan Bersatu (UDDA). BW-JGSA mengimbangi taburan marginal dan bersyarat dalam pengiraan Percanggahan Min Maksimum (MMD) pada dua matriks adaptasi yang tidak bergabung. Untuk UDDA, teknik pengurangan dua dimensi, iaitu analisis diskriminasi linear (LDA) dan analisis diskriminasi kepekaan lokaliti (LSDA), digabungkan untuk mencipta ciri dengan sifat diskriminasi global dan tempatan semasa proses penjajaran domain. Penilaian model BW-JGSA dan UDDA telah dijalankan pada set data imej awam dan digit (Office, Caltech-256, USPS, MNIST dan COIL20), manakala penilaian BW-UDDA dilakukan menggunakan set data tindakan berbilang kamera (MCAD). Kedua-dua model peningkatan mengatasi prestasi teknik semasa lain dengan ketepatan purata: 50.61% (set objek) dan 69.95% (set digit) untuk BW-JGSA, dan 59.95% (set objek) dan 80.72% (set digit) untuk UDDA. Bagi BW-UDDA untuk paparan terbuka HAR dinilai berdasarkan dua jenis penilaian pandangan silang. Ketepatan purata bagi penilaian pertama dan kedua dalam set data MCAD mengatasi prestasi teknik semasa dengan ketepatan lebih tinggi iaitu 13.38% dan 61.45%. BW- UDDA juga telah diuji pada set data HAR berbilang kamera terkawal, *Inria Xmas Motion Acquisition Sequences* (IXMAS), dengan ketepatan 90.91% menggunakan penilaian pandangan silang jenis kedua. Keputusan mengenai MCAD dan IXMAS ini mengesahkan keunggulan model yang dicadangkan untuk paparan terbuka HAR.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

$\mathrm{P}_s(x, y)$	-	Source Domain Distribution
$\mathrm{P}_t(x, y)$	-	Target Domain Distribution
$\beta(x)$	-	Reweighting Factor
D_{MMD}	-	MMD Operation in Domain
$\mathcal{H}$	-	Reproducing Kernel Hilbert Space (RKHS)
R	-	Risk Minimization
$\beta(x)$	-	Reweighting Factor for Instance Reweighting
$l(x, y, \theta)$	-	Loss Function for Instance Reweighting
$\tilde{L}_{\mathrm{c}}$	-	Cross-Entropy Loss
$\tilde{L}_{\mathrm{D}}$	-	Domain Discriminator Loss
$\hat{C}_s$, $\hat{C}_T$	-	Source and Target Covariances
$\hat{C}_{\bar{s}}$	-	Covariance of the transformed source features
D	-	Domain
D_s, D_t	-	Source and Target Domain
C_s, C_t	-	Learning Class for Source and Target Domains
X	-	Feature Space
$P(x)$	-	Marginal Distribution
Y	-	Label Space
$P(y\|x)$	-	Conditional Distribution
$n_s n_t$	-	Number Of Samples in Source and Target Domains
μ	-	Balanced Weighted Factor
$\mathbb{G}$	-	Nearest-Neighbor Graph
S_w,S_b	-	Within-Class-Scatter Matrix, Between-Class-Scatter-Matrix
L_w,L_b	-	Laplacian Matrix of $\mathbb{G}$
$P_s(x_{\mathrm{s}})$, $P_t(x_{\mathrm{t}})$	-	Marginal Distributions Source and Target Domain/View

$P_s(y_s\|x_s)$, $P_t(y_t\|x_t)$	-	Conditional Distributions Source and Target Domain/View
$\alpha, \beta, \gamma, \lambda, \eta, \zeta$	-	Parameters for Domain Adaptation Objective Function
L_{ss}, L_{tt}, L_{st}, L_{ts}	-	Marginal Distributions in MMD Computation
${L_{ss}}^{(c)}, {L_{tt}}^{(c)}$ ${L_{st}}^{(c)}, {L_{ts}}^{(c)}$	-	Conditional Distributions in MMD Computation
σ_s^2, σ_t^2	-	Variance of the source and target domains
$\sigma_{sL}^2, \sigma_{sG}^2$	-	Variance of the local and global source domain
$f(A), f(A,B), f(W)$	-	Objective function of unsupervised domain adaptation
G_w, G_b	-	Within-Class Subgraph and Between-Class Subgraph
$\hat{L}$	-	Graph Laplacian Matrix for Manifold Regularization
$\mathcal{A}$	-	Coefficient Vector for Element Z
x_s, x_t	-	Source and Target Domains Input Features Data
y_s, y_t	-	Source and Target Domains Input Labels
$\widehat{\boldsymbol{M}}$, $\boldsymbol{M}$	-	MMD Matrix with and without Balanced Weighted Factor
W_{ij}	-	Weight Matrix of $\mathbb{G}$
R_f	-	Manifold Regularization
Z_s, Z_t	-	New Representation of Source and Target Domains/Views
$\hat{z}$	-	Low-Dimensional Data for LDA/LSDA
A, B	-	Adaptation Matrices
S_t, H_t	-	Covariance Matrix, Centering Matrix
$\widehat{D}$	-	Diagonal Matrix for Manifold Regularization
$\hat{f}(Z)$	-	Theorem Representation for Manifold Regularization
θ	-	Diagonal Domain Indicator Matrix for Manifold Regularization
$K(Z_i, Z)$	-	Kernel Function for Manifold Regularization
T	-	Number of Iteration

K	-	Kernel Matrix
$\mathcal{L}$	-	Lagrange Function
Tr	-	Trace of Matrix
I	-	Identity Matrix
ω	-	Dense Optical Flow
$\widetilde{M}$	-	Median Filtering
d	-	Subspace Dimension
$\ell_{2,1}$	-	$\ell_{2,1}$ norm

LIST OF ABBREVIATIONS

HAR	-	Human action recognition
CCTV	-	Closed-circuit television
FOV	-	Field of view
KTH	-	*Kungliga Tekniska Högskolan*
IXMAS	-	INRIA Xmas Motion Acquisition Sequences
USPS	-	United States Postal Service
MNIST	-	Modified National Institute of Standards and Technology
COIL	-	Columbia Object Image Library
MCAD	-	Multi-camera Action dataset
RGB	-	Red-Green-Blue
STV	-	Space-time volumes
STIP	-	Spatio-temporal interest point
iDT	-	Improved dense trajectory
MEI	-	Motion energy image
MHI	-	Motion history image
SURF	-	Speeded-Up-Robust-Features
RANSAC	-	Random-Sample-Consensus
HOF	-	Histogram of Optical Flow
HOG	-	Histogram of Oriented Gradient
BoVW	-	Bag-of-Visual-Word
FV	-	Fisher Vector
DBN	-	Deep Belief Networks
DBMs	-	Deep Boltzmann Machines
RBMs	-	Restricted Boltzmann Machines
DNN	-	Deep Neural Networks
RNN	-	Recurrent Neural Networks
CNN	-	Convolutional Neural Networks
DSL	-	Deep Sequential Learning
LSTM	-	Long Short-Term Memory

LRCN	-	Long-Term Recurrent Convolutional Networks
RNN	-	Recurrent Neural Networks
PMK	-	Pyramid Match Kernel
HMM	-	Hidden Markov Model
SSM	-	Self-Similarity Matrix
MTL	-	Multi-task Learning
HOF3D	-	Histogram of 3-Dimensiona Optical Flow
3D-DCFF	-	3-Dimensional Distance Classifier Correlation Filter
K-NN	-	K-Nearest Neighbor
SVM	-	Support Vector Machine
HOMID	-	Histograms of Motion Intensity and Direction
BoBW	-	Bag of Bilingual Words
EM	-	Expectation-Maximization
LLR	-	Likelihood Ratio
MMC		Maximum Margin Clustering
HTDCC	-	Heterogeneous Transfer Discriminant Analysis of Canonical Correlation
SAM	-	Sample Affinity Matrix
JSRDA	-	Joint Sparse Representation and Distribution Adaptation
R-NKTM	-	Robust Non-Linear Knowledge Transfer Models
NIXMAS	-	Newer INRIA Xmas Motion Acquisition Sequences
WVU	-	West virginia university
N-UCLA	-	Northwestern University of California Los Angeles
MuHAVI	-	Multi-Camera Human Action Video
TJU	-	Tianjin University
UWA3D	-	University of Western Australia 3-Dimensional
i3DPost	-	Image 3-Dimensional Post
MMD	-	Maximum mean discrepancy
W-MMD	-	Weighted maximum mean discrepancy
RAAN	-	Reweighted adversarial adaptation network
RKHS	-	Reproducing kernel hilbert space
UDA	-	Unsupervised Domain Adaptation

TCA	-	Transfer component analysis
TJM	-	Transfer joint matching
JDA	-	Joint distribution adaptation
BDA	-	Balanced distribution adaptation
JPDA	-	Joint probability distribution adaptation
CORAL	-	Correlation alignment
SA	-	Subspace alignment
SDA	-	Subspace distribution alignment
PCA	-	Principal component analysis
SGF	-	Sampling geodesic flow
GFK	-	Geodesic flow kernel
GSL	-	Guide subspace learning
GTH	-	Guide transfer hashing
JGSA	-	Joint geometrical and statistical alignment
MEDA	-	Manifold embedded distribution alignment
CMD	-	Central moment discrepancy
GAN	-	Generative adversarial nets
ADDA	-	Adversarial discriminative domain adaptation
LDA	-	Linear discriminant analysis
BW-UDDA	-	Balanced Weighted Unified Discriminant and Distribution Alignment
BW-JGSA	-	Balanced Weighted Joint Geometrical and Statistical Alignment
UDDA	-	Unified Discriminant and Distribution Alignment
LSDA	-	Locality-Sensitive Discriminant Analysis
RKHS	-	Reproducing Kernel Hilbert Space

CHAPTER 1

INTRODUCTION

1.1 Background

Human action recognition (HAR), or human activity recognition, has been a popular research area particularly in computer vision, and man-machine interaction. According to Vishkarma *et al.* in [1], HAR can be interpreted as an activity performed by an actor, combined with multiple gestures, and is part of the high-level vision of human motion to understand human behavior. The action can also be considered as a sequence of primitive movements to fulfill a function or simple purpose [2]. Several examples of action include ‘walking,’ ‘running,’ ‘punching,’ ‘waving,’ and ‘kicking.’ Figure 1.1 shows some samples of human actions from the *Kungliga Tekniska Högskolan* (KTH) action dataset [3]. The most common application for human action recognition is intelligent video security surveillance in public places like airports, subway stations, hospitals, or areas where closed-circuit television (CCTV) is required. Additionally, HAR is useful in applications such as human-computer interaction, robot learning, entertainment, sports analysis, intelligent driver assistance systems, animation industries, and content-based video search [4]–[6].

From the perspective of human vision, it is easy for humans to understand the action and intention of an actor. A human can easily detect and recognize an action of an actor, such as waving or kicking, with high confidence. However, using human resources to monitor human actions is extremely expensive in a wide range of HAR applications. As a result, many researchers have attempted to create an automated system that mimics the visual capability of humans in understanding and describing human actions. Needless to say, this is not the most straightforward task due to the many challenges and issues involved, such as background complexity, inter and intra-class variations, noise, occlusions, poor resolution, real-time processing, and view-invariant [7]. This research focuses on view-invariant cases to recognize human action from different view angles. The challenge here is that the movement of the actor's body or posture of the human's body has changed across the views.

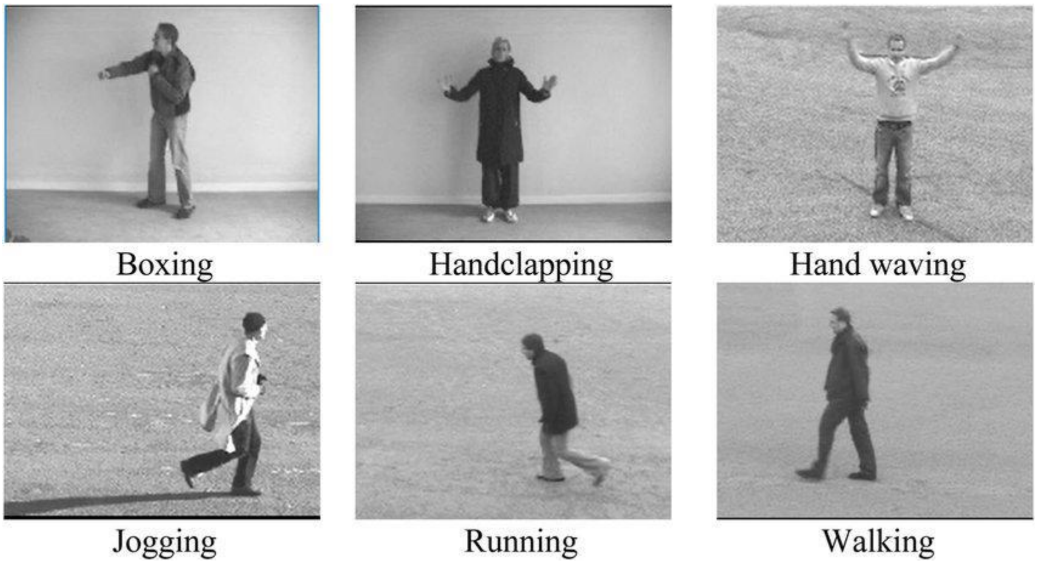

Figure 1.1 Example images from the KTH dataset [3]

In recent years, researchers for human action recognition have begun to move from studying using a single camera to multi-cameras. However, recognizing human actions automatically using multi-cameras is more complex than a single camera. For example, a different viewpoint of a camera may result in a diverse background, camera motion, a field of view, lighting condition, and occlusions. The current state-of-the-art approach is still far behind the human vision capability since most works are evaluated using a controlled environment dataset, making human action in multi-cameras still an ill-posed and unsolved complex problem.

From a multi-camera perspective, there are two types of conditions; scene condition and camera condition. Scene condition consists of elements that influence the recognition process, such as similar backgrounds, simultaneous action recorded for each view, and similar actors. The camera condition refers to the properties of a particular camera and its orientation that influence the recognition process, such as pixel resolution, camera position, and field of view (FOV). In most controlled environment datasets, the camera position is the only variable that changes between the source (training) view and the target (testing) view. However, in an uncontrolled environment that closely resembles real-world applications, other factors that can affect accuracy need to be considered, such as different backgrounds, different recording times, different camera resolutions, and different camera positions. Multi-camera scenario that specifically studies uncontrolled environment and condition is known as the open-view human action recognition [8] or the open-view HAR.

From literature [8], open-view HAR can be defined with the following characteristics; (1) Applicable only for multi-camera datasets or within cameras, (2) The correlation between cameras is minimized so that the dataset closely resembles the real-world environment. Thus, differences in parameters such as the illumination, camera type, background scene, and split action recorded are allowed, and (3) No labeled data is available in the target view.

The open-view HAR is more challenging than the conventional multi-cameras cases because the source view and target view are different. There are two main differences: (1) The previous work in multi-cameras considered an equal distribution of features between the training and the testing samples due to the assumption that both the source view and target view are highly similar. While, in the open-view HAR, this similarity can vary and causes the distribution difference across views not to be equally distributed, particularly when the view difference is large [9]. This will cause a standard classifier that has been trained in the source view not being robust enough in the target view due to the data bias, which is known as the distribution shift problem [8]. Hence, it is important to minimize the feature distribution shift between the source and the target view to mitigate large classification errors. (2) In minimizing the distribution shift between the source and the target view, the discrimination between classes also needs to be preserved. The preservation of the data discrimination will ensure that the feature from the same classes move closer to one another while those of different classes move farther away. Consequently, neglecting the preservation of data discrimination in minimizing the distribution shift will contribute to misclassification problems. Therefore, to handle the open-view cases, it is vital to minimize the distribution shift and preserve the data discrimination between classes.

The distribution shift issue is discussed in unsupervised domain adaptation as a sub-topic of transfer learning, a sub-discipline of machine learning. The theory of unsupervised domain adaptation describes the scenario in which the model trained in the source domain is used in a different (but related) target domain [10]. The domain adaptation process can minimize the feature distribution shift by projecting the source and target domains into a low-dimensional subspace. Figure 1.2 illustrates the unsupervised domain adaptation function in a low-dimensional subspace. The source and the target domains have a distribution shift resulting in poor accuracy performance if directly classified. The source and the target domains will be transformed into new representations in the common subspace using the adaptation matrix. The goal is to optimize the adaptation matrix to optimize the classification accuracy.

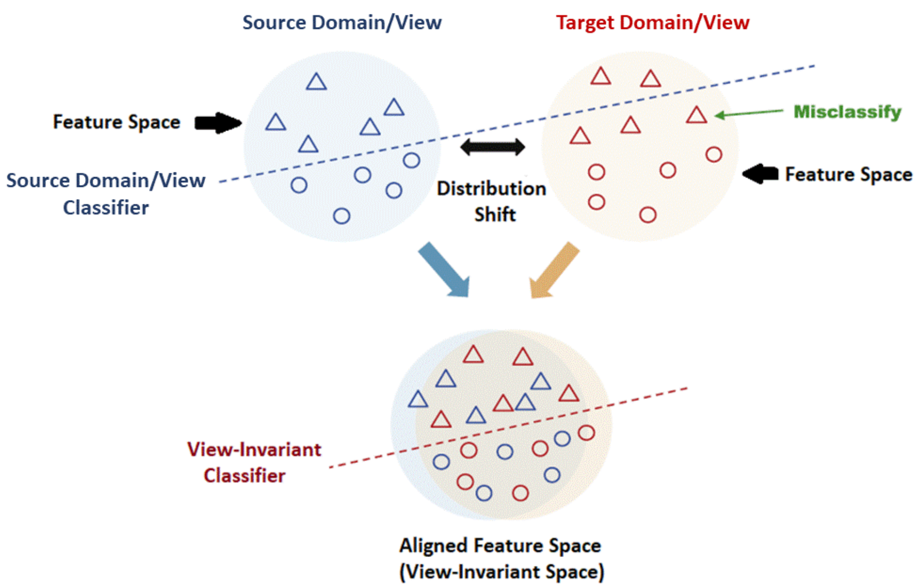

Figure 1.2 Illustration for unsupervised domain adaptation [11]. The triangles and circles denote two different classes. Before being classified by a linear classifier, both source and target view spaces will be projected and aligned as closely as possible to form a new representation in a common subspace

1.2 Problem Statement

There are two kinds of distributions in probability distribution: marginal and conditional. Most of the existing domain adaptation methods adapt either the marginal distribution, conditional distribution, or both. Recent work in [12], [14] shows that considering both distributions could perform better. However, both distributions are treated by concatenating them with a similar weight. In the open-view case, different views require different marginal and conditional weights. For instance, the marginal distributions should be more dominant if the view/domain from the source and the target are dissimilar. Whereas if the view/domain from the source and the target are more related, the conditional distributions are more dominant [14]. Furthermore, because of significant differences in open-view human action recognition, there is a possibility that a common subspace may not exist. Thus, to optimize the adaptation matrix in an unsupervised domain adaptation model, the process shall consider balancing the weights of both marginal and conditional distributions while minimizing the distance between the sample mean of the source and the target domains. In addition, the adaptation process shall consider that there is a possibility that no common subspace exists because of significant differences between views/domains.

The other issue is the preservation of data discrimination while projecting the source and target view into a new representation. It involves minimizing the distance between feature data of similar classes and maximizing the distance between feature data of different categories. In the unsupervised domain adaptation model, the labeled data is only available in the source domain, synchronizing with open-view human action recognition properties. Thus, the availability of existing labels in the source domain can be exploited to improve the class variance. Consequently, it is assumed that enhancing the class variance in the source domain can optimize the adaptation matrix.

Synergy in resolving the above two issues is believed have the potential to improve the performance of the current unsupervised domain adaptation methodology. Therefore, the idea is not only to improve existing unsupervised domain adaptation methods but also to exploit both solutions and implement them into the open-view human action recognition. Such a proposed model will lead to a better that action classification accuracy for open-view human action recognition.

1.3 Objectives of the Study

This thesis aims to implement an unsupervised domain adaptation approach in solving the open-view human action recognition challenges. Following the problem statement discussed in Section 1.2, the objectives are broken down as follows:

1. To develop the unsupervised domain adaptation enhancement model that optimizes the adaptation matrix by balancing marginal and conditional distribution weights. This model should work even if there is a possibility of unified subspaces not existing because of significant differences between the source and the target domains.

2. To develop the unsupervised domain adaptation enhancement model that optimizes the adaptation matrix by improving the source domain’s class variance while maintaining the source and target domains' discriminatory feature properties.

3. To design a dedicated unsupervised domain adaptation model for open-view human action recognition by exploiting the enhancement models proposed in objectives one and two above.

1.4 Scope of the Study

The scope of the study is limited to the following conditions:

1. The unsupervised domain adaptation enhancement models will be first evaluated using selected five public image datasets. These datasets are Office [15], Caltech-256 [16], USPS [17], MNIST [18], and COIL- 20 [19].

2. The HAR datasets used for this research for the dedicated unsupervised domain adaptation model are confined to (1) MCAD dataset [20] as the primary evaluation for open-view cases. According to Section 1.1, MCAD meets the criteria for open-view cases. MCAD datasets are recorded with both day and night actions to set the dataset to be uncontrolled and suited for open-view cases. (2) IXMAS dataset [21] as a well-known controlled multi-camera HAR dataset. IXMAS will be used as a validation dataset to prove that the model proposed is equivalent to other methods.

3. The human action recognition datasets above contain a single actor with no other moving object involved, such as a vehicle or animal. The input modality used is from an RGB camera only. In addition, low-level features are extracted based on handcrafted learning, not deep learning. The illumination is fixed to day time only.

4. The performance will be measured primarily in terms of accuracy of human actions recognition only, not in real-time. Due to the different validation approaches between multi-camera and open-view human action recognition, the comparison method with the state-of-the-art for enhancement and dedicated models is limited only to the unsupervised domain adaptation approaches.

5. Simulations and experiments were all conducted using MATLAB software. Nevertheless, some packages are applied to Ubuntu using C and C++ to extract low-level features. All experiments were run on a PC with an Intel Core i7 CPU (6 cores) and 20GB RAM.

1.5 Contributions

This study has three significant contributions, which can be summarized as follows:

1. The enhancement of an unsupervised domain adaptation model balances the weight of marginal and conditional distributions in the distribution distance computation, consequently optimizing the adaptation matrix between source and target domains during the adaptation process.

2. The enhancement of an unsupervised domain adaptation model improves the source domain's class variance while maintaining the source and target domains' discriminatory feature properties.

3. The development of a dedicated unsupervised domain adaptation model to improve human action recognition in open-view cases. This dedicated model is based on exploiting the first and second contributions above.

1.6 Organizations of Thesis

This thesis consists of five chapters. There is a brief introduction to HAR in Chapter one, including definition, applications, importance in the computer vision field, and challenges faced, along with problem statements that will introduce the issues to be discussed, objectives that will be drawn from the problem statement, scopes that will make the thesis relevant to the study constraints, research contributions from the studies, and organization of the thesis. Chapter two discusses the literature review that supports the thesis's objectives and direction. This chapter has three sub-topics: (1) the HAR approach based on single-camera approaches, (2) the HAR based on multi-cameras approaches, and (3) domain adaptation development, the basic concept, and related works. Towards the end of the chapter, a summary of the proposed human action recognition and domain adaptation work is presented. Chapter three highlights the proposed methodology for the open-view HAR case, starting with low-level feature extraction and descriptor, details of unsupervised domain adaptation block as a primary focus, and linear classifier involved. Chapter four discusses experiments conducted in this research to prove its contribution, the results obtained, and a comprehensive analysis. This chapter also discusses the datasets used and comparisons with other methods. Chapter five concludes and summarizes the entire chapter along with suggestions for future work.

REFERENCES

[1] S. Vishwakarma and A. Agrawal, “A survey on activity recognition and behavior understanding in video surveillance,” *Vis. Comput.*, vol. 29, no. 10, pp. 983–1009, 2013.

[2] J. M. Chaquet, E. J. Carmona, and A. Fernández-Caballero, “A survey of video datasets for human action and activity recognition,” *Comput. Vis. Image Underst.*, vol. 117, no. 6, pp. 633–659, 2013.

[3] C. Schüldt, I. Laptev, and B. Caputo, “Recognizing human actions: a local SVM approach,” *Proc. - Int. Conf. Pattern Recognit.*, vol. 3, pp. 32–36, 2004.

[4] M. B. Holte, T. B. Moeslund, C. Tran, and M. M. Trivedi, “Human action recognition using multiple views: a comparative perspective on recent developments,” *Proc. 2011 Jt. ACM Work. Hum. gesture Behav. Underst.*, pp. 47–52, 2011.

[5] Y. Kong and Y. Fu, “Human action recognition and prediction: a survey,” *arXiv Prepr. arXiv1806.11230*, vol. 13, no. 9, 2018.

[6] X. Ji and H. Liu, “Advances in view-invariant human motion analysis: a review,” *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.*, vol. 40, no. 1, pp. 13–24, 2010.

[7] S. A. R. Abu-Bakar, “Advances in human action recognition: an updated survey,” *IET Image Process.*, vol. 13, no. 13, pp. 2381–2394, 2019.

[8] Y. Su, Y. Li, and A. Liu, “Open-view human action recognition based on linear discriminant analysis,” *Multimed. Tools Appl.*, vol. 78, no. 1, pp. 767–782, 2019.

[9] Y. Liu, Z. Lu, J. Li, and T. Yang, “Hierarchically learned view-invariant representations for cross-view action recognition,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 8, pp. 2416–2430, Aug. 2019.

[10] W. M. Kouw and M. Loog, “An introduction to domain adaptation and transfer learning,” *arXiv Prepr. arXiv1812.11806.*, 2018.

[11] X. Li, W. Zhang, Q. Ding, and J. Q. Sun, “Multi-layer domain adaptation method for rolling bearing fault diagnosis,” *Signal Processing*, vol. 157, pp. 180–197, 2019.

[12] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, “Transfer feature learning with joint distribution adaptation,” *2013 IEEE Int. Conf. Comput. Vis.*, pp. 2200–2207, Dec. 2013.

[13] J. Zhang, W. Li, and P. Ogunbona, “Joint geometrical and statistical alignment for visual domain adaptation,” *IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2017, pp. 5150–5158, Jul. 2017.

[14] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen, “Balanced distribution adaptation for transfer learning,” *IEEE Int. Conf. Data Min.*, pp. 1129–1134, Nov. 2017.

[15] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, “Adapting visual category models to new domains,” *Eur. Conf. Comput. Vis.*, pp. 213–226, 2010.

[16] Boqing Gong, Yuan Shi, Fei Sha, and K. Grauman, “Geodesic flow kernel for unsupervised domain adaptation,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2066–2073, Jun. 2012.

[17] J. J. Hull, “A database for handwritten text recognition research,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 550–554, 1994.
[18] Y. Lecun, L. Bottou, Y. Bengio, and P. Ha, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, no. November, pp. 2278–2324, 1998.
[19] S. Nene, S. Nayar, and H. Murase, “Columbia object image library (COIL-20),” *Tech. Rep.*, vol. 95, pp. 223–303, 1996.
[20] W. Li, Y. Wong, A. A. Liu, Y. Li, Y. T. Su, and M. Kankanhalli, “Multi-camera action dataset for cross-camera action recognition benchmarking,” *IEEE Winter Conf. Appl. Comput. Vision, WACV 2017*, pp. 187–196, 2017.
[21] D. Weinland, E. Boyer, and R. Ronfard, “Action recognition from arbitrary views using 3D exemplars,” *IEEE 11th Int. Conf. Comput. Vis.*, pp. 1–7, 2007.
[22] J. K. Aggarwal and M. S. Ryoo, “Human activity analysis,” *ACM Comput. Surv.*, vol. 43, no. 3, pp. 1–43, 2011.
[23] G. Cheng, Y. Wan, A. N. Saudagar, K. Namuduri, and B. P. Buckles, “Advances in human action recognition: a survey,” *arXiv Prepr. arXiv1501.05964*, pp. 1–30, 2015.
[24] A. Sargano, P. Angelov, and Z. Habib, “A comprehensive review on handcrafted and learning-based action representation approaches for human activity recognition,” *Appl. Sci.*, vol. 7, no. 1, p. 110, 2017.
[25] F. Zhu, L. Shao, J. Xie, and Y. Fang, “From handcrafted to learned representations for human action recognition: a survey,” *Image Vis. Comput.*, vol. 55, pp. 42–52, 2016.
[26] R. Poppe, “A survey on vision-based human action recognition,” *Image Vis. Comput.*, vol. 28, no. 6, pp. 976–990, 2010.
[27] S. Herath, M. Harandi, and F. Porikli, “Going deeper into action recognition: a survey,” *Image Vis. Comput.*, vol. 60, pp. 4–21, 2017.
[28] A. F. Bobick and J. W. Davis, “The recognition of human movement using temporal templates,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 3, pp. 257–267, 2001.
[29] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, “Actions as space-time shapes,” *IEEE Trans. Pattern Anal. Mach. Intell.*, no. Nov, 2007.
[30] D. Weinland, R. Ronfard, and E. Boyer, “Free viewpoint action recognition using motion history volumes,” *Comput. Vis. Image Underst.*, vol. 104(2–3), pp. 249–257, 2006.
[31] D. L. Bruce and K. Takeo, “An iterative image registration technique with an application to stereo vision,” in *Proceedings of Image Understanding Workshop*, 1981.
[32] J. M. Alexei A. Efros, Alexander C. Berg, Greg Mori, “Action at a distance,” *Proc. Ninth IEEE Int. Conf. Comput. Vis.*, vol. 3, pp. 726–726, 2003.
[33] N. Robertson and I. Reid, “Behaviour understanding in video: a combined method,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. I, pp. 808–815, 2005.
[34] Y. Wang and G. Mori, “Hidden part models for human action recognition: probabilistic versus max margin,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1310–1323, 2011.
[35] H. Zhang, Y. Zhang, B. Zhong, Q. Lei, L. Yang, and J. Du, “A comprehensive survey of vision-based human action recognition methods,” *Sensors*, vol. 19, no. 5, p. 1005, 2019.
[36] Laptev and Lindeberg, “Space-time interest points,” *IEEE Int. Conf. Comput. Vis.*, vol. 1, pp. 432–439, 2003.

[37] C. Harris and M. Stephens, “A combined corner and edge detector,” in *Alvey vision conference*, 1988.
[38] Y. Li, R. Xia, Q. Huang, W. Xie, and X. Li, “Survey of spatio-temporal interest point detection algorithms in video,” *IEEE Access*, vol. 5, pp. 10323–10331, 2017.
[39] I. Laptev, B. Caputo, C. Schu, and T. Lindeberg, “Local velocity-adapted motion events for spatio-temporal recognition,” *Comput. Vis. image Underst.*, vol. 108, no. 3, pp. 207–229, 2007.
[40] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie, “Behavior recognition via sparse spatio-temporal features,” *Proc. - 2nd Jt. IEEE Int. Work. Vis. Surveill. Perform. Eval. Track. Surveillance, VS-PETS*, vol. 2005, pp. 65–72, 2005.
[41] H. Wang, A. Kläser, C. Schmid, and C. L. Liu, “Action recognition by dense trajectories,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3169–3176, 2011.
[42] H. Wang, A. Kläser, C. Schmid, and C. L. Liu, “Dense trajectories and motion boundary descriptors for action recognition,” *Int. J. Comput. Vis.*, vol. 103, no. 1, pp. 60–79, 2013.
[43] H. Wang and C. Schmid, “Action recognition with improved trajectories,” *IEEE Int. Conf. Comput. Vis.*, pp. 3551–3558, Dec. 2013.
[44] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (SURF),” *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.
[45] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model applications to image fitting with analysis and automated cartography,” *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[46] I. Laptev, M. Marszałek, C. Schmid, and B. Rozenfeld, “Learning realistic human actions from movies,” *26th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR*, pp. 0–7, 2008.
[47] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” *Proc. - 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, CVPR 2005*, vol. I, pp. 886–893, 2005.
[48] N. Dalal, B. Triggs, C. Schmid, N. Dalal, B. Triggs, and C. Schmid, “Human detection using oriented histograms of flow and appearance,” *Eur. Conf. Comput. Vis. (ECCV ’06)*, pp. 428–441, 2006.
[49] X. Peng, C. Zou, Y. Qiao, and Q. Peng, “Action recognition with stacked fisher vectors,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8693, no. Part 5, pp. 581–595, 2014.
[50] G. E. Hinton, S. Osindero, and T. Yee-Whye, “A fast learning algorithm for deep belief nets,” *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
[51] R. Salakhutdinov and G. Hinton, “Deep boltzmann machines,” *J. Mach. Learn. Res.*, vol. 5, no. 3, pp. 448–455, 2009.
[52] Smolensky Paul, “Information processing in dynamical systems: foundations of harmony theory,” *J. Japan Soc. Fuzzy Theory Syst.*, vol. 4, no. 2, pp. 220–228, 1986.
[53] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, “Exploring strategies for training deep neural networks,” *J. Mach. Learn. Res.*, vol. 10, pp. 1–40, 2009.
[54] A. Sherstinsky, “Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network,” *Phys. D Nonlinear Phenom.*, vol. 404, no. March, pp. 1–43, 2020.

[55] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, and W. Hubbard, “Backpropagation applied to handwritten zip code recognition,” *Neural computation*, vol. 1, no. 4. pp. 541–551, 1989.
[56] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Adv. Neural Inf. Process. Syst.*, vol. 25, pp. 1097–1105, 2012.
[57] S. Ji, W. Xu, M. Yang, and K. Yu, “3D convolutional neural networks for human action recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, 2013.
[58] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, “Sequential deep learning for human action recognition,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7065 LNCS, pp. 29–39, 2011.
[59] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, “Learning precise timing with LSTM recurrent networks,” *J. Mach. Learn. Res.*, vol. 3, no. 1, pp. 115–143, 2003.
[60] A. Karpathy and T. Leung, “Large-scale video classification with convolutional neural networks,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 10–20, 2014.
[61] K. Simonyan and A. Zisserman, “Two-stream convolutional networks for action recognition in videos,” *Adv. Neural Inf. Process. Syst.*, vol. 1, no. January, pp. 568–576, 2014.
[62] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, and K. Saenko, “Long-term recurrent convolutional networks for visual recognition and description,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 2625–2634, 2015.
[63] F. Lv and R. Nevatia, “Single view human action recognition using key pose matching and viterbi path searching,” *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1–8, 2007.
[64] Anwaar-ul-Haq, I. Gondal, and M. Murshed, “On dynamic scene geometry for view-invariant action matching,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3369–3376, 2011.
[65] I. N. Junejo, E. Dexter, I. Laptev, and P. Pérez, “View-independent action recognition from temporal self-similarities,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 1, pp. 172–185, 2011.
[66] Binlong Li, O. I. Camps, and M. Sznaier, “Cross-view activity recognition using hankelets,” *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1362–1369, Jun. 2012.
[67] M. B. Holte, B. Chakraborty, J. Gonzàlez, and T. B. Moeslund, “A local 3-D motion descriptor for multi-view human action recognition from 4-D spatio-temporal interest points,” *IEEE J. Sel. Top. Signal Process.*, vol. 6, no. 5, pp. 553–565, 2012.
[68] T.-D. Le, T.-O. Nguyen, and T.-H. Tran, “Improving multi-view human action recognition with spatial-temporal pooling and view shifting techniques,” *Proceedings-8th Int. Symp. Inf. Commun. Technol.*, pp. 348–355, 2017.
[69] A. U. Haq, I. Gondal, and M. Murshed, “On temporal order invariance for view-invariant action recognition,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 2, pp. 203–211, 2013.

[70] A. Ulhaq, X. Yin, J. He, and Y. Zhang, "On space-time filtering framework for matching human actions across different viewpoints," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1230–1242, 2018.

[71] C. Thurau and V. Hlaváč, "Pose primitive based human action recognition in videos or still images," *26th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR*, 2008.

[72] A. Fathi and G. Mori, "Action recognition by learning mid-level motion features," *26th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR*, 2008.

[73] K. Huang, Y. Zhang, and T. Tan, "A discriminative model of motion and cross ratio for view-invariant action recognition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2187–2197, 2012.

[74] F. Zhu, L. Shao, and M. Lin, "Multi-view action recognition using local similarity random forests and sensor fusion," *Pattern Recognit. Lett.*, vol. 34, pp. 20–24, 2013.

[75] F. Murtaza, M. H. Yousaf, and S. A. Velastin, "Multi-view human action recognition using 2D motion templates based on MHIs and their HOG description," *IET Comput. Vis.*, vol. 10, no. 7, pp. 758–767, 2016.

[76] A. S. Ogale, A. Karapurkar, G. Guerra-filho, and Y. Aloimonos, "View-invariant identification of pose sequences for action recognition," in *VACE*, 2004.

[77] A. B. Sargano, P. Angelov, and Z. Habib, "Human action recognition from multiple views based on view-invariant feature descriptor using support vector machines," *Appl. Sci.*, vol. 6, no. 10, p. 309, 2016.

[78] N. Kase, M. Babaee, and G. Rigoll, "Multi-view human activity recognition using motion frequency," *IEEE Int. Conf. Image Process.*, pp. 3963–3967, Sep. 2017.

[79] X. Ji, Z. Ju, C. Wang, and C. Wang, "Multi-view transition HMMs based view-invariant human action recognition method," *Multimed. Tools Appl.*, vol. 75, no. 19, pp. 11847–11864, 2016.

[80] S. Chun and C.-S. Lee, "Human action recognition using histogram of motion intensity and direction from multiple views," *IET Comput. Vis.*, vol. 10, no. 4, pp. 250–257, 2016.

[81] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[82] J. Liu, M. Shah, B. Kuipers, and S. Savarese, "Cross-view action recognition via view knowledge transfer," *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3209–3216, Jun. 2011.

[83] J. Wang, H. Zheng, J. Gao, and J. Cen, "Cross-view action recognition based on a statistical translation framework," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 8, pp. 1461–1475, 2016.

[84] A. Farhadi and M. K. Tabrizi, "Learning to recognize activities from the wrong view point," *Eur. Conf. Comput. Vis.*, pp. 154–166, 2008.

[85] Z. Zhang, C. Wang, B. Xiao, W. Zhou, and S. Liu, "Cross-view action recognition using contextual maximum margin clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 10, pp. 1663–1668, 2014.

[86] R. Li and T. Zickler, "Discriminative virtual views for cross-view action recognition," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2855–2862, 2012.

[87] Z. Zhang, C. Wang, B. Xiao, W. Zhou, S. Liu, and C. Shi, “Cross-view action recognition via a continuous virtual path,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2690–2697, 2013.
[88] C. H. Huang, Y. R. Yeh, and Y. C. F. Wang, “Recognizing actions across cameras by exploring the correlated subspace,” *Eur. Conf. Comput. Vis.*, pp. 342–351, 2012.
[89] X. Wu, H. Wang, C. Liu, and Y. Jia, “Cross-view action recognition over heterogeneous feature spaces,” *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4096–4108, 2015.
[90] W. Nie, A. Liu, W. Li, and Y. Su, “Cross-view action recognition by cross-domain learning,” *Image Vis. Comput.*, vol. 55, pp. 109–118, 2016.
[91] Y. Kong, Z. Ding, J. Li, and Y. Fu, “Deeply learned view-invariant features for cross-view action recognition,” *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 3028–3037, 2017.
[92] J. Zheng, Z. Jiang, J. Phillips, and R. Chellappa, “Cross-view action recognition via a transferable dictionary pair,” *Br. Mach. Vis. Conf.*, pp. 125.1-125.11, 2012.
[93] J. Zheng and Z. Jiang, “Learning view-invariant sparse representations for cross-view action recognition,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 3176–3183, 2013.
[94] J. Zheng, Z. Jiang, and R. Chellappa, “Cross-view action recognition via transferable dictionary learning,” *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2542–2556, 2016.
[95] C. Zhang, H. Zheng, and J. Lai, “Cross-view action recognition based on hierarchical view-shared dictionary learning,” *IEEE Access*, vol. 6, pp. 16855–16868, 2018.
[96] H. Rahmani and A. Mian, “Learning a non-linear knowledge transfer model for cross-view action recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 2458–2466, 2015.
[97] H. Rahmani, A. Mian, and M. Shah, “Learning a deep model for human action recognition from novel viewpoints,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 667–681, 2018.
[98] J. Zhang, H. P. H. Shum, J. Han, and L. Shao, “Action recognition from arbitrary views using transferable dictionary learning,” *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4709–4723, 2018.
[99] D. Weinland, M. Özuysal, and P. Fua, “Making action recognition robust to occlusions and viewpoint changes,” *Eur. Conf. Comput. Vis.*, pp. 635–648, 2010.
[100] Z. Cheng, L. Qin, Y. Ye, Q. Huang, and Q. Tian, “Human daily action analysis with multi-view and color-depth data,” *Eur. Conf. Comput. Vis.*, pp. 52–61, 2012.
[101] N. Gkalelis, H. Kim, A. Hilton, N. Nikolaidis, and I. Pitas, “The i3dpost multi-view and 3d human action/interaction database,” *CVMP 2009 - 6th Eur. Conf. Vis. Media Prod.*, pp. 159–168, 2009.
[102] V. Kulathumani, R. Kavi, and S. Ramagiri, “WVU multi-view action recognition dataset.” [Online]. Available: http://csee.wvu.edu/~vkkulathumani/wvu-action.html#download2, 2011.
[103] J. Wang, X. Nie, Y. Xia, Y. Wu, and S.-C. Zhu, “Cross-view action modeling, learning and recognition,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2649–2656, 2014.

[104] Y. Yan, E. Ricci, R. Subramanian, G. Liu, and N. Sebe, “Multitask linear discriminant analysis for view invariant action recognition,” *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5599–5611, 2014.

[105] A. Liu, Y. Su, P. Jia, Z. Gao, T. Hao, and Z. Yang, “Multiple/single-view human action recognition via part-induced multitask structural learning,” *IEEE Trans. Cybern.*, vol. 45, no. 6, pp. 1194–1208, 2015.

[106] S. Singh, S. A. Velastin, and H. Ragheb, “Muhavi: a multicamera human action video dataset for the evaluation of action recognition methods,” *IEEE Int. Conf. Adv. Video Signal Based Surveill.*, pp. 48–55, Aug. 2010.

[107] H. Rahmani, A. Mahmood, D. Huynh, and A. Mian, “Histogram of oriented principal components for cross-view action recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 12, pp. 2430–2443, 2016.

[108] T. Joachims, “Transductive inference for text classification using support vector machines,” *Int. Conf. Mach. Learn.*, pp. 200–209, 1999.

[109] J. Zhang, W. Li, P. Ogunbona, and D. Xu, “Recent advances in transfer learning for cross-dataset visual recognition: A problem-oriented perspective,” *ACM Comput. Surv.*, pp. 1–38, 2019.

[110] V. M. Patel, R. Gopalan, R. Li, and R. Chellappa, “Visual domain adaptation: a survey of recent advances,” *IEEE Signal Processing Magazine*, no. April, IEEE, pp. 53–69, 2015.

[111] J. Huang, A. J. Smola, A. Gretton, K. M. Borgwardt, and B. Schölkopf, “Correcting sample selection bias by unlabeled data,” *Adv. Neural Inf. Process. Syst.*, pp. 601–608, 2007.

[112] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, “Domain adaptation via transfer component analysis,” *IEEE Trans. Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[113] L. Zhang and X. Gao, “Transfer adaptation learning: a decade survey,” *arXiv Prepr. arXiv1903.04687*, no. March, pp. 1–26, 2019.

[114] B. Gretton, Arthur Smola, Alex Huang, Jiayuan Schmittfull, Marcel Borgwardt, Karsten Schölkopf, “Covariate shift by kernel mean matching,” *Dataset shift Mach. Learn.*, p. 5, 2009.

[115] H. Shimodaira, “Improving predictive inference under covariate shift by weighting the log-likelihood function,” *J. Stat. Plan. Inference*, vol. 90, no. 2, pp. 227–244, 2000.

[116] J. Jiang and C. X. Zhai, “Instance weighting for domain adaptation in NLP,” *ACL 2007 - Proc. 45th Annu. Meet. Assoc. Comput. Linguist.*, no. June, pp. 264–271, 2007.

[117] T. M. Harry Hsu, W. Y. Chen, C. A. Hou, Y. H. Hubert Tsai, Y. R. Yeh, and Y. C. Frank Wang, “Unsupervised domain adaptation with imbalanced cross-domain data,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 4121–4129, 2015.

[118] S. Li, S. Song, and G. Huang, “Prediction reweighting for domain adaptation,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, no. 7, pp. 1682–1695, 2017.

[119] S. Chen, F. Zhou, and Q. Liao, “Visual domain adaptation using weighted subspace alignment,” *VCIP 2016 - 30th Anniv. Vis. Commun. Image Process.*, pp. 30–33, 2017.

[120] H. Yan, Y. Ding, P. Li, Q. Wang, Y. Xu, and W. Zuo, “Mind the class weight bias: weighted maximum mean discrepancy for unsupervised domain adaptation,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2272–2281, 2017.
[121] Q. Chen, Y. Liu, Z. Wang, I. Wassell, and K. Chetty, “Re-weighted adversarial adaptation network for unsupervised domain adaptation,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 75–94, 2018.
[122] K. M. Borgwardt, A. Gretton, M. J. Rasch, H. P. Kriegel, B. Schölkopf, and A. J. Smola, “Integrating structured biological data by kernel maximum mean discrepancy,” *Bioinformatics*, vol. 22, no. 14, pp. 49–57, 2006.
[123] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, “Transfer joint matching for unsupervised domain adaptation,” *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1410–1417, Jun. 2014.
[124] W. Zhang and D. Wu, “Discriminative joint probability maximum mean discrepancy (DJP-MMD) for domain adaptation,” *Proc. Int. Jt. Conf. Neural Networks*, pp. 1–8, 2020.
[125] B. Sun, J. Feng, and K. Saenko, “Return of frustratingly easy domain adaptation,” *Proc. Conf. Artif. Intell.*, no. 30 (1), pp. 2058–2065, 2016.
[126] B. Sun and K. Saenko, “Deep coral: correlation alignment for deep domain adaptation,” *Proc. Eur. Conf. Comput. Vis.*, vol. 9915 LNCS, pp. 443–450, 2016.
[127] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, “Unsupervised visual domain adaptation using subspace alignment,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 2960–2967, 2013.
[128] B. Sun and K. Saenko, “Subspace distribution alignment for unsupervised domain adaptation,” *Proc. Br. Mach. Vis. Conf. 2015*, vol. 4, pp. 24.1-24.10, 2015.
[129] R. Gopalan, R. Li, and R. Chellappa, “Domain adaptation for object recognition: an unsupervised approach,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 999–1006, 2011.
[130] M.-W. Mak and J.-T. Chien, “Guide subspace learning for unsupervised domain adaptation,” *Mach. Learn. Speak. Recognit.*, vol. 31, no. 9, pp. 217–248, 2020.
[131] J. Wang, W. Feng, Y. Chen, H. Yu, M. Huang, and P. S. Yu, “Visual domain adaptation with manifold embedded distribution alignment,” *Proc. 26th ACM Int. Conf. Multimed.*, pp. 402–410, Oct. 2018.
[132] J. Donahue, J. Yangqing, V. Oriol, J. Hoffman, N. Zhang, and E. Tzeng, “Decaf: a deep convolutional activation feature for generic visual recognition,” *31st Int. Conf. Mach. Learn. ICML 2014*, vol. 2, pp. 988–996, 2014.
[133] W. Zellinger, E. Lughofer, and S. Saminger-platz, “Central moment discrepancy (CMD) for domain-invariant representation learning,” *Proc. Int. Conf. Learn. Represent.*, pp. 1–13, 2017.
[134] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, and S. Ozair, “Generative adversarial networks,” *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
[135] N. Silberman, D. Dohan, and S. Francisco, “Unsupervised pixel–level domain adaptation with generative adversarial networks,” *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3722–3731, 2017.

[136] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, “Adversarial discriminative domain adaptation,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2962–2971, 2017.
[137] A. Chadha and Y. Andreopoulos, “Improved techniques for adversarial discriminative domain adaptation,” *IEEE Trans. Image Process.*, vol. 29, pp. 2622–2637, 2020.
[138] J. Liu and L. Zhang, “Optimal projection guided transfer hashing for image retrieval,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 10, pp. 3788–3802, 2019.
[139] D. Cai, X. He, and J. Han, “SRDA: an efficient algorithm for large scale discriminant analysis,” *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 1, pp. 1–12, 2008.
[140] J. Wen, X. Fang, J. Cui, L. Fei, K. Yan, and Y. Chen, “Robust sparse linear discriminant analysis,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 2, pp. 390–403, 2019.
[141] J. Cai and X. Huang, “Modified sparse linear-discriminant analysis via nonconvex penalties,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 10, pp. 4957–4966, 2018.
[142] X. Li, Q. Wang, F. Nie, and M. Chen, “Locality adaptive discriminant analysis framework,” *IEEE Trans. Cybern.*, pp. 1–12, 2021.
[143] D. Cai, X. He, K. Zhou, and J. Han, “Locality sensitive discriminant analysis,” *IJCAI Int. Jt. Conf. Artif. Intell.*, no. 60633070, pp. 708–713, 2007.
[144] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, “Locality-constrained linear coding for image classification,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3360–3367, 2010.
[145] F. Zhu and L. Shao, “Weakly-supervised cross-domain dictionary learning for visual recognition,” *Int. J. Comput. Vis.*, vol. 109, no. 1–2, pp. 42–59, 2014.
[146] C. L. Chang, “Finding prototypes for nearest neighbor classifiers,” *IEEE Trans. Comput.*, vol. C–23, no. 11, pp. 1179–1184, 1974.
[147] Y. Liu, Z. Lu, J. Li, C. Yao, and Y. Deng, “Transferable feature representation for visible-to-infrared cross-dataset human action recognition,” *Complexity*, vol. 2018, pp. 1–20, 2018.
[148] W. Nie, A. Liu, J. Yu, Y. Su, L. Chaisorn, and Y. Wang, “Multi-view action recognition by cross-domain learning,” *Int. Work. Multimed. Signal Process.*, pp. 1–6, Sep. 2014.

LIST OF PUBLICATIONS

1. **MSR Samsudin**, SAR Abu-Bakar, MM Mokji, "An Improved Open-View Human Action Recognition with Unsupervised Domain Adaptation," *Multimedia Tools and Applications*, pp: 1-29, 2022. **(Q2, IF: 2.757)**

2. **MSR Samsudin**, SAR Abu-Bakar, MM Mokji, "Balanced Weight Joint Geometrical and Statistical Alignment for Unsupervised Domain Adaptation." *Journal of Advances in Information Technology Vol* 13.1, 2022. **(Indexed by SCOPUS)**

3. **MSR Samsudin**, SAR Abu-Bakar, MM Mokji, "Unified Discriminant and Distribution Alignment for Visual Domain Adaptation." *2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE, 2021. **(Indexed by SCOPUS)**