Improving Cross-resolution Face Matching using Ensemble based Co-Transfer Learning

Himanshu S. Bhatt, *Student Member, IEEE,* Richa Singh, *Senior Member, IEEE,* Mayank Vatsa, *Senior Member, IEEE,* and Nalini K. Ratha, *Fellow, IEEE*

***Abstract*—Face recognition algorithms are generally trained for matching high resolution images and they perform well for similar resolution test data. However, the performance of such systems degrade when a low resolution face image captured in unconstrained settings such as videos from cameras in a surveillance scenario are matched with high resolution gallery images. The primary challenge here is to extract discriminating features from limited biometric content in low resolution images and match it to information rich high resolution face images. The problem of cross-resolution face matching is further alleviated when there is limited labeled positive data for training face recognition algorithms. In this paper, the problem of cross-resolution face matching is addressed where low resolution images are matched with high resolution gallery. A *co-transfer learning framework* is proposed which is a cross-pollination of transfer learning and co-training paradigms and is applied for cross-resolution face matching. The transfer learning component transfers the knowledge that is learnt while matching high resolution face images during training for matching low resolution probe images with high resolution gallery during testing. On the other hand, co-training component facilitates this transfer of knowledge by assigning pseudo labels to unlabeled probe instances in the target domain. Amalgamation of these two paradigms in the proposed ensemble framework enhances the performance of cross-resolution face recognition. Experiments on multiple face databases show the efficacy of the proposed ensemble based co-transfer learning algorithm as compared to other existing algorithms and a commercial system. In addition, several high profile real world cases have been used to demonstrate the usefulness of the proposed approach in addressing the tough challenges.**

***Index Terms*—Face recognition, cross resolution, transfer learning, co-training, co-transfer learning.**

I. INTRODUCTION

It is generally believed that face recognition by computers is a solved problem in many scenarios such as user centric applications including face tagging in Google Picasa, Facebook and screen/device unlocking in Android and Windows-based systems. While significant advances have been made in last two decades, unconstrained face recognition is yet to benefit from these advances to be useful in real world applications. One such example is face recognition in low resolution surveillance images. With advancements in technology, surveillance cameras now have a profound presence and are widely used in security and law enforcement applications. There are several instances where surveillance videos have helped agencies in apprehending individuals who have committed crime or identify individuals with the intent to commit crime. For example, in 2005 subway bomb blasts in London [1], CCTV footage helped law enforcement officers in identifying the bombers. In 2008 Mumbai terrorist attacks [2], surveillance cameras installed at different locations (CST railway station, Taj Palace, and Trident hotels) helped the agencies to track the activities of terrorists and later identify them. In the 2010 car bomb case at Times Square [3], the surveillance footage captured an unidentified individual leaving the car with explosives. Later, widespread distribution and manual investigation of the video helped the investigating agencies to apprehend the individual.

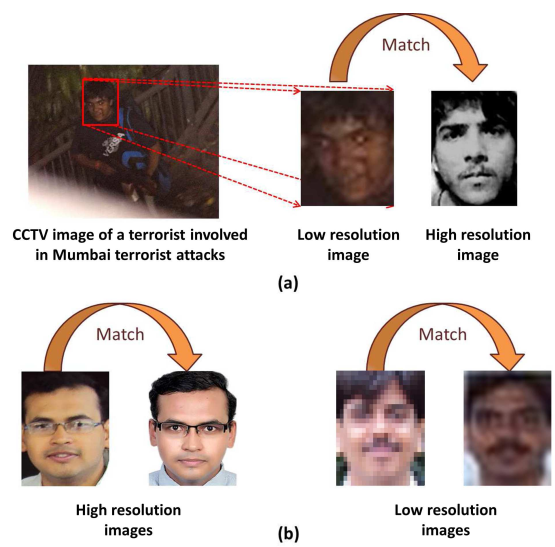


Fig. 1. Illustrating the difference in matching (a) low resolution and high resolution images, (b) two high resolution and low resolution images.

In all these cases, surveillance cameras could not foil the terrorist attacks, however, they served as the primary evidence in leading the investigation and also recognizing the individuals at the end. It is therefore desirable to build a system where surveillance cameras coupled with a face recognition algorithm can be used to automatically identify individuals from a watch-list. Along with the challenges of pose, expression, illumination [4], aging [5], disguise [6], [7], and plastic surgery [8], [9] in face recognition, matching a watch-list photograph to an image obtained from surveillance camera also requires the capability of matching across resolution. For example, the watch-list photograph could be a high resolution image

H.S. Bhatt, R. Singh,and M. Vatsa are with the Indraprastha Institute of Information Technology (IIIT) Delhi, India, e-mail: {himanshub, rsingh, mayank}@iiitd.ac.in.
N.K. Ratha is with IBM, T.J. Watson Research Center, USA, email: ratha@us.ibm.com.

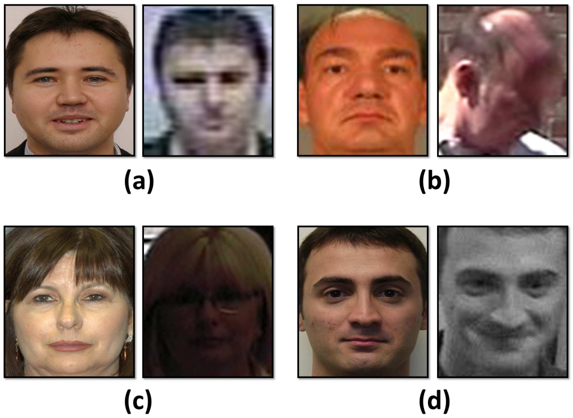


Fig. 2. Illustrates the challenge in matching low resolution images when coupled with other covariates. Low resolution challenge (a) alone, (b) with pose, (c) with illumination, and (d) with expression.

whereas the surveillance camera images are generally low resolution images. As shown in Fig. 1, even if both the images are frontal, the information content in both the images could be significantly different. The presence of pose, illumination, and expression along with different resolution could further exacerbate the problem, as shown in Fig. 2.

The conditions in which a face recognition algorithm is trained are referred to as the *source domain* where the availability of large training data helps the algorithm to efficiently learn the task. In the source domain, face recognition algorithms are trained to match high resolution images. However, for surveillance applications, the probe data i.e., the *target domain*, comprises low resolution face images and the gallery contains high resolution face images. Gallery represents images in the database and probe represents the query images. Source domain refers to scenario where both gallery and probe images are high resolution images while target domain represents scenario where gallery is always of higher resolution than the probe image. Under these variations, the performance of a biometric system degrades because it is unable to efficiently utilize the knowledge learned in the source domain and there is a scarcity of labeled low resolution data that can be used for training the algorithms. Obtaining sufficient labels for the target data is time consuming, requires human effort, and very expensive. However, there is an abundance of *unlabeled* low resolution data in target domain during testing.

In our preliminary work [18], we made this observation and formulated the problem of cross-resolution face matching where sufficient labeled data is available in source domain and only a few labeled instances are available from the target domain. This research extends the prior work [18] and proposes a generalized *co-transfer learning* (CTL) framework which is a cross-pollination of transfer learning [19] and co-training [20]. The framework integrates transfer learning and co-training in a non-separable manner to efficiently transfer the knowledge from the source domain to the target domain with sequentially available unlabeled instances from the target domain:

- *transfer learning* is used to leverage the knowledge learned in the source domain for efficiently matching low resolution probes with high resolution gallery in the target domain.
- *co-training* is used to enable transfer learning with unlabeled probe instances from the target domain by assigning pseudo-labels to probes.

In face recognition literature, to the best of our knowledge, this is the first work that leverages unlabeled probe instances to facilitate knowledge transfer in an ensemble based algorithm. The performance of the proposed framework is evaluated in a cross-resolution face recognition application and the comparative experiments are performed on four face databases, namely, the CMU Multi-PIE [21], SCface [22], ChokePoint [23], and MBGC v2 video challenge [24] databases. The results are also presented on some real world samples (surveillance images) and recognition is performed using the proposed co-transfer learning algorithm against a large gallery database of 6534 subjects. Finally, the results on still-frontal matching challenge of Point and Shoot Challenge (PaSC) database [25] are also presented. The results show that the proposed algorithm outperforms existing algorithms including FaceVACS which is a commercial face recognition system.

II. Literature Review

The literature review is divided into three parts: (1) cross-resolution face recognition, (2) co-training, and (3) transfer learning.

A. Review of Cross-resolution Face Recognition Literature

In literature, several approaches have been proposed to match cross-resolution face images. As shown in Table I, these algorithms can be classified into two categories: *super-resolution* and *transformation* based approaches. Fig. 3 illustrates the broad categorization and the steps involved in cross-resolution face recognition approaches. Super-resolution based approaches for cross-resolution matching enhance the low quality probe image before recognition. On the other hand, transformation based approaches extract features that are resilient to resolution changes and match cross-resolution face images. Some of the transformation based approaches also perform resolution invariant transformations either in the image space or the feature space for matching.

Super-resolution based approaches: Huang and He [10] proposed to build a coherent subspace between the PCA features of high resolution (HR) and low resolution (LR) images mapped using the radial basis functions for recognition. Baker and Kanade [26] proposed an algorithm to *apriori* learn the spatial distribution of image gradients to enhance the resolution of local features before matching. Chakrabarty *et al.* [27] proposed a learning based method to super-resolve face images with kernel principal component analysis-based prior model. Chang *et al.* [28], [29] formed geometrically similar manifolds using local facial patches in the low and high resolution images. They used training images to estimate the high-resolution embedding and construct a smooth super-resolved image. Yang *et al.* [30] proposed a super resolution

TABLE I
EXISTING ALGORITHMS FOR CROSS-RESOLUTION FACE IMAGE MATCHING.

Approach	Technique	Databases	Gallery/probe resolution
Super-resolution	Coherent features [10]	FERET, UMIST, ORL	72×72 / 12×12
	Multi-modal tensor face [11]	AR, YALE, FERET	56×36 / 14×9
	S2R2 [12]	Multi-PIE, FERET, FRGC v.2	24×24 / 6×6
	Relationship learning [13]	FRGC v.2	64×48 / 28×24
Transformation	LFD [14]	FERET	88×80 / 33×30
	Coupled locality preserving mapping (CLPM) [15]	FERET	72×72 / 12×12
	Synthesis based LR face recognition[16]	CMU-PIE, FRGC v.2	48×40 / 19×16
	MDS [17]	Multi-PIE	48×40 / 12×10

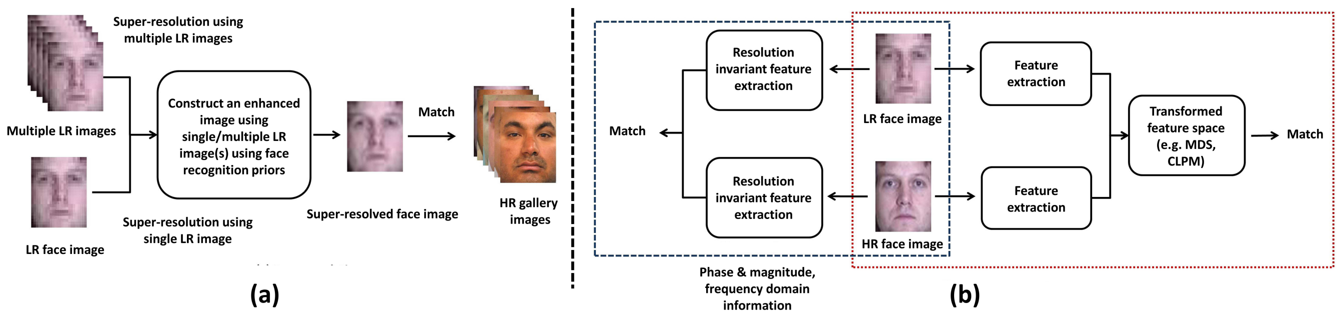


Fig. 3. Broad view of cross-resolution face matching approaches. (a) Super resolution and (b) transformation based approaches.

approach by representing local patches as a sparse linear combination of elements from high resolution images. In addition to these local models, Liu *et al.* [31] integrated a holistic parametric and a local nonparametric model using two-step statistical modeling for face hallucination. It was observed that super-resolution approaches, due to environmental variations and distortions, failed to significantly improve the recognition performance. It is our assertion that the primary objective of super-resolution is to obtain a good visual reconstruction from low resolution face(s), and these algorithms are generally not intended for recognition. However, there are some approaches that simultaneously optimize both super resolution and face recognition. Jia and Gong [11] combined super-resolution and face recognition by computing a maximum likelihood identity parameter vector in high-resolution tensor space for recognition. Further, Hennings-Yeomans *et al.* [12] proposed an approach where facial features were included in a super-resolution method as the prior information for simultaneous reconstruction of super-resolved images. Recently, Zou and Yuen [13] proposed a super-resolution technique based on the relationship between the high-resolution image space and the very low resolution image space. Their technique improved face recognition performance for the very low resolution problem.

Transformation based approaches: Unlike super-resolution, another method to match cross-resolution images is to downsample high resolution images to the level of low resolution images before matching. However, information useful for face recognition such as texture, edges, and other high frequency information is compromised while downsampling the images. To address this problem, Li *et al.* [15] proposed to project both high resolution and low resolution images to a feature space using coupled mappings. Biswas *et al.* [32] proposed a multidimensional scaling approach to simultaneously transform the features from high resolution gallery and low resolution probe images. The Euclidean distance between the transformed feature vectors approximates the distance computed when the probe images were captured at similar resolution as that of the gallery images. Researchers have also studied that the phase and magnitude in frequency domain can be used as a resolution invariant representation for efficiently matching cross-resolution face images. Lei *et al.* [14] proposed a local frequency descriptor based on the magnitude and phase information to match cross-resolution face images in the frequency domain. Shekhar *et al.* [16] proposed a generative approach using the information from high resolution gallery to match low resolution probe images with illumination variations. Lei *et al.* [33] proposed a coupled discriminant analysis for heterogeneous face recognition (matching high vs. low resolution images). To maintain the discriminative power and generalizability of their approach, they utilized multiple samples from different resolutions along with locality information in the kernel space.

B. Review of Co-training Literature

In co-training, as proposed by Blum and Mitchell [20], two classifiers that are trained on separate views (features) co-train each other based on their confidence in predicting the labels. Nonetheless, success of a co-training framework is susceptible to various assumptions. Blum and Mitchell [20] have shown that two classifiers should have sufficient individual accuracy and should be conditionally independent of each other. Later, Abney [46] has shown the weak dependence between the two classifiers can also guarantee successful co-training. Wang and Zhou [47] reported the sufficient and necessary condition for success of a co-training framework. Co-training has been used in several computer vision applications with

TABLE II
SOME REPRESENTATIVE APPROACHES RELATED TO THE PROPOSED ALGORITHM.

Authors	Technique	Application
Zhu *et al.* [34]	Heterogeneous transfer learning using matrix factorization	Classifying image and text data
Quattani *et al.* [35]	Sparse prototype image representation	Recognizing visual categories
Ahmed *et al.* [36]	Hierarchical feed-forward model	Recognizing visual categories
Geng *et al.* [37]	Domain adaptation metric learning	Face recognition & web image annotation
Wang *et al.* [38]	Dyadic knowledge transfer using a non-negative matrix tri-factorization	Computer vision applications
Siyu *et al.* [39]	Subspace transfer learning for kinship verification	Kinship verification using faces
Chen *et al.* [40]	Transferring informative knowledge for learning expression models	Learning person-specific facial expression model
Bhatt *et al.* [41]	Online co-training in SVMs using two independent feature representations	Face Verification
Cao *et al.* [42]	Transfer learning via generative Bayesian model with KL divergence	Face Verification
Ng *et al.* [43]	Co-Transfer Learning using a joint transition probability graph based on co-occurrence of the data	Classifying image and text data
Zhao and Hoi [44]	Ensemble based transfer learning with incremental labeled data	Text classification
Guo and Wang [45]	Domain adaptive input-output kernel learning	Recognizing visual categories
Proposed	Co-transfer learning: Ensemble based transfer leaning using incremental unlabeled target domain data with co-training	Face recognition

very limited exposure in biometrics. However, in biometrics literature, unlabeled data has been used primarily for updating the templates [48], [49], [50]. Poh *et al.* [51] performed a study on the goal of semi-supervised learning where they focused on some of the challenges and research directions for designing adaptive biometric systems. Classifier update using co-training is explored by Bhatt *et al.* [41] where the biometric classifiers are updated using labeled as well as unlabeled instances.

C. Review of Transfer Learning Literature

Transfer learning has been explored in many computer vision applications. Zhu *et al.* [34] proposed a heterogeneous transfer learning framework that utilized annotated images from the web as a bridge to transfer knowledge between text and images using a matrix factorization approach. Quattoni *et al.* [35] proposed a method for learning a sparse prototype image representation for transfer across visual categories. Their approach used a large set of unlabeled data and a kernel function to form a representation. Ahmed *et al.* [36] proposed a hierarchical feed-forward model for visual recognition using transfer learning from pseudo tasks which include a set of pattern matching operations constructed from the data. Geng *et al.* [37] proposed a domain adaptation metric learning by introducing a data dependent regularization to conventional metric learning in the reproducing kernel Hilbert space. This minimized the empirical maximum mean discrepancy between different domains. Wang *et al.* [38] proposed dyadic knowledge transfer which is a non-negative matrix tri-factorization based approach to transfer cross-domain image knowledge for the new computer vision tasks. In face recognition or related domains, transfer learning has been applied to verify kinship using face images through subspace transfer learning [39]. Chen *et al.* [40] also proposed to learn a person-specific facial expression model by transferring the informative knowledge from other people. Their approach allows to learn an accurate person-specific model for a new subject with only a small amount of person specific data. Most of the transfer learning techniques work in offline manner and assume that the data from the target domain is available upfront. Table II also lists some of the closely related approaches to the proposed co-transfer learning algorithm. Cao *et al.* [42] proposed a transfer learning approach for face verification using a simple Bayesian model. Their main idea was to minimize the KL divergence between the source and target domain distributions to enhance the sharing of information. Ng *et al.* [43] proposed a co-transfer learning algorithm using a graph based method to link different feature set using a joint transition probability graph. Their approach is a supervised approach that transfers knowledge across different domains based on the affinities computed using co-occurrence information. Bhatt *et al.* [41] proposed a semi-supervised online co-training approach to update the classifier's decision boundary using labeled as well as unlabeled information for face verification. Zhao *et al.* [44] proposed an online transfer learning (OTL) framework where knowledge is transferred from source domain to target domain classifier within an ensemble in a supervised manner using incremental labeled instances from the target domain. The OTL framework forms the basis of the proposed CTL algorithm. Compared to co-training and transfer learning research directions, the proposed CTL algorithm is different in the sense that it is an incremental semi-supervised approach that uses few labeled and large unlabeled data to transfer knowledge within an ensemble. Co-training is used to transform unlabeled incremental data from the target domain into pseudo labeled data to facilitate transfer learning. In this research, transfer learning and co-training are jointly used to transfer the knowledge learnt in the source domain (with labeled samples) to the target domain (with unlabeled samples), as shown in Fig. 4.

III. CO-TRANSFER LEARNING FRAMEWORK

We, humans, have innate abilities of transferring knowledge between related tasks. It is observed that if the new task is closely related to the previous learning, humans can quickly transfer this knowledge to perform the new task. However, given some prior knowledge in a related task, traditional algorithms are unable to adapt to a new task and have to learn the new task from the beginning. Generally, they do not consider that the two tasks may be related and the knowledge gained in one may be used to learn the new task efficiently in lesser time. *Transfer learning* attempts to mimic this human behavior by transferring the knowledge learned in one or more source tasks and use it for learning the related target task. Several approaches have been proposed for transfer learning and they can be categorized as 1) inductive, 2) transductive, and 3) unsupervised transfer learning. Based on the domain representation, transfer learning approaches can be further categorized into homogeneous and heterogeneous transfer learning. The source and target domains share same feature space in the former whereas feature space is different in the later one. For a more detailed discussion on different transfer learning approaches, readers are directed to [19].

Generally, labeled data in target domain is scarce and obtaining labels for the target data is time consuming and expensive in most real world scenarios; therefore, it is difficult to learn a model for the target data. On the other hand, large amount of unlabeled data, available in the form of probe, can be leveraged to learn the model. There are some existing semi-supervised approaches for face recognition [52], [53], [54], [55] that utilize few labeled and ample amount of unlabeled data for enhancing face recognition performance. Many of these semi-supervised approaches are used for template update such as semi-supervised PCA [49], [56] or LDA [57]. There are few approaches [41], [52] that update/retrain the model with few labeled and large unlabeled data. Mostly, existing semi-supervised algorithms require entire unlabeled data up-front and do not perform well for single sample per subject.

The proposed co-transfer learning algorithm builds on the limitations of existing approaches to address the challenge of single sample per subject and performs transfer learning in online manner with sequential unlabeled data available from the target domain. Transfer learning and co-training are jointly used to *transfer* the knowledge learned in the source domain to the target domain with unlabeled instances, as shown in Fig. 4. Co-training to update the classifiers has been explored by Bhatt *et al.* [41] where biometric classifiers are updated using labeled as well as unlabeled instances. However, to the best of our knowledge, it is the first algorithm that uses transfer learning for face recognition as a semi-supervised approach using few labeled and a large number of unlabeled probe instances.

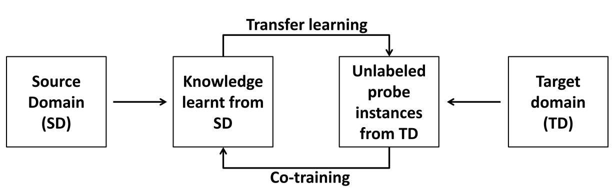


Fig. 4. Illustrating the cross-pollination of transfer learning and co-training for transferring knowledge from source domain to target domain.

The proposed framework is a generalized framework that can be applied to any classifier which allows re-training with incremental data. In this research, we have applied the concept of co-transfer learning to support vector machine (SVM). Re-training the SVM classifier in batch mode is computationally expensive [58] and may not be feasible in real-wold applications. Some approaches have been proposed that allow re-training the SVM classifier using only the previous support vectors and new incremental data points. A method to add or remove one sample at a time to update SVM is proposed in [59] where a solution for $N \pm 1$ samples can be obtained using the N old samples and the sample which is to be added or removed. In the proposed approach, SVM is first trained using an initial training set and a decision hyperplane is obtained. This hyperplane is then updated using the new available instances and the previous support vectors. For more details on updating SVM classifiers with new incremental data, readers are directed to [41], [58], [59].

Transfer Learning: In face recognition, the classifiers such as SVM, are learned using training data (from the source domain) while the performance is evaluated on a separate unseen test data (the target domain) which may have different properties and follow a different distribution compared to the training data. Consider a scenario where there are two classifiers, one trained using the source and another trained using the target domain data. During training, there is a large labeled data in the source domain i.e., for matching HR probe with HR gallery images (source domain) but only a few labeled instances are available in the target domain, i.e., for matching LR probes with HR gallery images. In such a case, the source domain classifier alone may not efficiently classify the test instances because of the variations in data distribution of source and target domains. Since the classifier in target domain is trained using only a few labeled samples, it is not able to efficiently classify the test instances. It has to learn/update its decision boundary with the incremental data available in the target domain. Both the classifiers are individually insufficient to classify the test data from the target domain. Therefore, in the proposed algorithm, an ensemble is built as a weighted combination of the source and target domain classifiers. It efficiently classifies test instances and subsequently transfers the knowledge from the source domain to the target domain as and when the data from the target domain is available. For this, the two classifiers trained on the source and target domains are combined to efficiently classify the unlabeled probe instances.

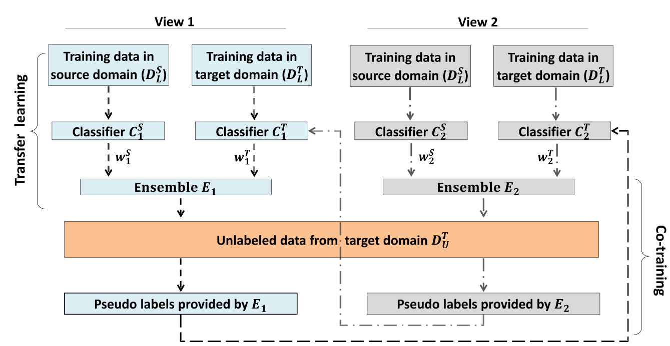


Fig. 5. Block diagram illustrating the steps involved in the proposed co-transfer learning framework.

As shown in Fig. 5, the source domain classifiers (C_j^S) are trained using sufficient HR labeled training data denoted by $D_L^S = \{(\mathbf{u}_1^S, z_1), (\mathbf{u}_2^S, z_2), ..., (\mathbf{u}_n^S, z_n)\}$. Every i^{th} instance, $\mathbf{u}_i$ has two views $\{x_{i,1}, x_{i,2}\}$ for the training label $z_i \in \{-1, +1\}$; here $x_{i,1}$ and $x_{i,2}$ represent the input vectors obtained from two separate views (features). $\{-1\}$ refers to the impostor class where the query and probe images belong to different subjects and $\{+1\}$ refers to the genuine class where the gallery and probe images belong to the same subject. The two views are utilized for co-training (explained later). The target domain classifiers (C_j^T) are initially trained on a few labeled training instances from the target domain represented as D_L^T={$(\mathbf{u}_1^T, z_1), (\mathbf{u}_2^T, z_2), ..., (\mathbf{u}_m^T, z_m)$}. Here, n and m are the number of training instances in the source and target domains respectively, such that $n > m$ and $j = 1, 2$ represents the view (feature). Let a set of r unlabeled probe instances in the target domain be represented as D_U^T={$(\mathbf{u'}_1^T), (\mathbf{u'}_2^T), ..., (\mathbf{u'}_r^T)$}. An ensemble prediction function, denoted as E_j, is constructed for each view. E_j is a weighted combination of the source domain classifier, C_j^S, and the target domain classifier, C_j^T, with $w_{i,j}^S$ and $w_{i,j}^T$ representing the weights of the source domain classifier and target domain classifier for the i^{th} instance of the j^{th} view respectively. For the i^{th} unlabeled probe instance in the j^{th} view, the ensemble function E_j predicts the label, $E_j(x_{i,j}) \rightarrow y_{i,j}$. For the i^{th} instance in the target domain $\mathbf{u'}_i$, class label is predicted by the ensemble as given in Eq. 1.

$$y_{i,j} = sign(w_{i,j}^S \; \Pi(C_j^S(\mathbf{u'}_i)) + w_{i,j}^T \; \Pi(C_j^T(\mathbf{u'}_i)) - \frac{1}{2}) \quad (1)$$

where Π is a normalization function such that $\Pi(x) = max(0, min(1, \frac{x+1}{2}))$. Initially, both the weights are set to 0.5 so that each classifier contributes equally within an ensemble and gradually, they are automatically adjusted to emphasize the contribution from the updated target domain classifiers in an ensemble. As proposed by Zhao and Hoi [44], the two weights are updated dynamically as shown in Eqs. 2 and 3.

$$w_{i+1,j}^S = \frac{w_{i,j}^S h_i(C^S)}{w_{i,j}^S h_i(C^S) + w_{i,j}^T h_i(C_j^T)} \quad (2)$$

$$w_{i+1,j}^T = \frac{w_{i,j}^T h_i(C^T)}{w_{i,j}^S h_i(C^S) + w_{i,j}^T h_i(C^T)} \quad (3)$$

where $w_{i+1,j}^S$ and $w_{i+1,j}^T$ are the updated weights and h_i is defined as:

$$h_i(C) = \exp\{-\eta \; l(\Pi(C_i), \Pi(\hat{y}_i))\}, \quad (4)$$

$\eta = 0.5$, $l(y, \hat{y}) = (y - \hat{y})^2$ is the square loss function, y is the predicted label and $\hat{y}$ is the pseudo label provided by co-training (explained later).

Co-training: As mentioned previously, unlabeled probe instances are available in abundance and can be utilized to update/learn the classifiers in the target domain. However, it is required to obtain the labeled target data. Obtaining labeled training instances from the target domain is difficult, expensive, and requires human effort. In biometrics, there are situations when only a small set of labeled data is available for training while a huge amount of unlabeled data is readily available as probe. This situation is similar to a semi-supervised learning scenario, where co-training [20], [41] has proven beneficial as it can be used to transform unlabeled probe instances into pseudo-labeled training instances. In the proposed co-training approach, a small initial labeled set is available from the target domain for training the classifiers and a large number of unlabeled instances are available as probe. It assumes the availability of two ensemble functions (classifiers), E_1 and E_2, trained on separate views (features) where each ensemble function has sufficient (better than random) accuracy. If the first ensemble confidently predicts genuine label for an instance while the second ensemble predicts impostor label with low confidence, then this particular instance (with pseudo label provided by the first ensemble) is utilized for updating the second ensemble and vice-versa. In this research, the confidence of prediction for an instance on the j^{th} view[1], denoted by α_j, is measured as the distance of that instance from the decision boundary which is computed as shown in Eq. 5.

$$\alpha = \frac{R}{|v|}, \quad (5)$$

where, R is the un-normalized output from the SVM, v is the weight vector for the support vectors and $|v| = v^T v$. For confidently predicting an instance to belong to genuine class, the distance from the decision hyperplane should be greater than the genuine threshold (P_{gen}). Similarly, an instance is confidently predicted as impostor if the distance from the hyperplane is greater than the impostor threshold (P_{imp}). Here P_{gen} refers to genuine threshold when comparing for genuine class and P_{imp} to impostor threshold when comparing for impostor class. Since SVM is used for classification, a genuine threshold is computed as the distance of the farthest support vector of $genuine$ class. Similarly, an impostor threshold is computed as the distance of the farthest support vector of $impostor$ class. Varying the thresholds will change the number of instances on which the co-training is performed. High threshold value implies conservative co-training while smaller value of the threshold leads to aggressive co-training. In this manner, unlabeled probe instances are transformed into pseudo-labeled training instances which are then used to update the ensembles. In an ensemble, knowledge is transferred by updating the decision boundary of the target domain classifier C_j^T using only the new incremental data as proposed in [41].

Co-transfer: In the proposed framework, transfer learning and co-training work concurrently to improve the target domain task with pseudo labels provided by co-training that lead to transfer of knowledge from the source to the target domain. Within each ensemble, the target domain classifier updates its decision boundary [41] with every pseudo-labeled instance obtained during testing. Moreover, the weights corresponding to the source and target domain classifiers are also adjusted dynamically using Eqs. 2 and 3. This scheme avoids the need to learn the target domain classifiers from the beginning and hence, makes the system scalable and computationally efficient. Note that in the co-transfer learning framework, only target domain classifiers are updated with pseudo-labeled instances. The source domain classifiers do not need any update because they are well trained using large amount of labeled data available upfront in the source domain. The proposed *co-transfer learning* framework is summarized in Algorithm 1.

[1] View and features are used interchangeably in the paper.

Algorithm 1 Co-transfer learning

Input: Initial labeled training data D_L^S in the source domain, a few labeled instances D_L^T from the target domain. Unlabeled probe instances D_U^T from target domain (available sequentially).
Iterate: j= 1 **to** 2 (number of views)
Process: Train classifiers C_j^S and C_j^T on j^{th} view of D_L^S and D_L^T respectively to construct ensemble E_j. Compute confidence thresholds P_j for each view.
for $i = 1$ **to** r (number of probe instances) **do**
  Predict labels: $E_j(x_{i,j}) \rightarrow y_{i,j}$; calculate α_j: confidence of prediction
  if $\alpha_1 > P_1$ & $\alpha_2 < P_2$ **then**
    Update C_2^T with pseudo-labeled instance $\{x_{i,2}, y_{i,1}\}$ & recompute w_2^S and w_2^T.
  end if.
  if $\alpha_1 < P_1$ & $\alpha_2 > P_2$ **then**
    Update C_1^T with pseudo-labeled instance $\{x_{i,1}, y_{i,2}\}$ & recompute w_1^S and w_1^T.
  end if.
end for.
end iterate.
Output: Updated classifiers C_1^T, C_2^T and weights w_1^S, w_1^T, w_2^S and w_2^T.

Generally, co-training is performed with two views to co-train the participating classifiers by providing pseudo labeled instances and appending the training set for classifiers trained on each view. However, this can be generalized to multiple views where the final pseudo label is assigned based on the majority vote, similar to [60]. Likewise, the proposed co-transfer learning algorithm can be generalized to accommodate multiple views that may be extracted from other features such as processing different channels in a color image as separate views or adding more resolution invariant features as separate views. However, incorporating more views may increase computational requirements.

Error bounds: To analyze the effectiveness of the proposed co-transfer learning algorithm, we compute the error bounds. Using the square loss function $l^*(y, \hat{y}) = (y - \hat{y})^2$ and the exponential weighting update function, bounds of an ensemble are given as:

$$\sum_{i=1}^{I} l^*(w_i^S \Pi(C_i^S) + w_i^T \Pi(C_i^T), \Pi(\hat{y}_i)) \leq 2ln(2) \quad (6)$$

$$+min\{\sum_{i=1}^{I} l^*((\Pi(C_i^S), \Pi(\hat{y}_i)), \sum_{i=1}^{I} l^*(\Pi(C_i^T), \Pi(\hat{y}_i))\} \quad (7)$$

where I is the number of instances, y_i is the predicted label for the i^{th} instance, and $\hat{y}_i$ is the pseudo label for the i^{th} instance provided by co-training. The above equation is derived by following the proof in [44]. Using this, the error bounds of an ensemble are derived as follows: The error at the i^{th} step is represented as $|w_i^S \Pi(C_i^S) + w_i^T \Pi(C_i^T) - \Pi(\hat{y}_i)| \geq \frac{1}{2}$. Therefore, we have

$$\sum_{i=1}^{I} l^*(w_i^S \Pi(C_i^S) + w_i^T \Pi(C_i^T), \Pi(\hat{y}_i)) = \quad (8)$$

$$\sum_{i=1}^{I} (w_i^S \Pi(C_i^S) + w_i^T \Pi(C_i^T), \Pi(\hat{y}_i))^2 \geq \frac{1}{4} M \quad (9)$$

Combining Eqs. 6 and 8, we have

$$\frac{1}{4} M \leq min\left\{\sum C^S, \sum C^T\right\} + 2ln(2) \quad (10)$$

where $\sum C^S = \sum_{i=1}^{I} l^*(\Pi(C_i^S), \Pi(\hat{y}_i))$ and $\sum C^T = \sum_{i=1}^{I} l^*(\Pi(C_i^T), \Pi(\hat{y}_i))$. For two ensembles, when the final decision classification decision is based on their combination, the error bounds M for the co-transfer learning algorithm are given as:

$$min(M_{E1}, M_{E2}) \leq M \leq max(M_{E1}, M_{E2}) \quad (11)$$

The primary objective of selecting two ensembles is to facilitate co-transfer learning as one ensemble provides pseudo labeled training instances to the other. Therefore, the error bounds of the proposed algorithm will lie between the error bounds of the two participating ensembles as shown in Eq. 11. Note that these error bounds are derived under the assumption that the pseudo labels provided by co-training are correct.

IV. Co-transfer Learning for Cross-resolution Face Recognition

In an operational scenario, training is performed in a controlled environment; whereas during testing, a biometric system encounters data from uncontrolled environment. Co-training is particularly useful for recognizing cross-resolution face images. Fig. 6 shows the block diagram of the proposed co-transfer learning framework for matching cross-resolution face images. First, the source and target domain classifiers are trained on two views (features) and two ensemble functions (E_1 and E_2) are built. One view is the local phase quantization (LPQ)[2] [61] and the second view is the scale invariant feature transform (SIFT)[3] [62]. LPQ operates on the Fourier phase computed locally for a window in every image position and utilizes local phase information extracted using a short-term Fourier transform. In our experiments, same parameters as proposed by Ahonen et al. [61] are used. SIFT [62] is a scale and rotation invariant descriptor that generates a compact representation of an image based on the magnitude, orientation, and spatial vicinity of image gradients. In this research, SIFT descriptor is computed in a dense manner on pre-defined interest points. Both these views are resilient to scale changes and can be effectively used for matching face images with different resolutions. The two features provide diverse information, one encodes the discriminative phase information whereas the other encodes information from the

[2] Source code available at http://www.cse.oulu.fi/CMV/Downloads-/LPQMatlab

[3] Source code available at $http : //labelme.csail.mit.edu - /Release3.0/browserTools/php/matlab_toolbox.php$

image gradients. LPQ and SIFT descriptors are normalized to unit length. A second normalization step is performed by suppressing any component larger than 0.2 down to 0.2 and re-normalizing the vector to unit length. Finally, χ^2 distance is used to compare two corresponding LPQ and SIFT descriptors. For normalizing face images, eye-coordinates are detected using OpenCV's boosted cascade of Haar-like features. Face image is normalized with respect to the horizontal axis and the inter-eye distance is fixed to 100 pixels for a 216×216 image. Low resolution images are also normalized in a similar manner where the inter-eye distance is normalized in proportion to the image resolution[4].

Initial training on labeled data from the source and target domains: The co-transfer learning framework assumes that during training, each subject has high resolution gallery-probe pairs and a few subjects have corresponding low resolution images from the target domain. For a given gallery-probe pair, face images are tessellated into 3×3 patches. LPQ and SIFT descriptors are computed for each local patch and matched using the χ^2 distance measure. Distance scores corresponding to local patches are vectorized to an input vector $\{\mathbf{u}_i, z_i\}$, where $z_i \in \{-1, +1\}$ is the associated label. $\{+1\}$ signifies that the gallery-probe pair belongs to the same individual (i.e. genuine pair) whereas $\{-1\}$ signifies that the gallery-probe pair belongs to images corresponding to different individuals (i.e. impostor pair). Input vectors obtained by matching LPQ descriptors of two high resolution images are utilized for training the source domain SVM classifier (C_1^S) on view 1. On the contrary, the target domain SVM classifiers for view 1 are trained using one high resolution and one low resolution images. The source domain and target domain SVM classifiers are then combined to form an ensemble, E_1. Similarly, the SVM classifiers for view 2 (SIFT) are trained and the ensemble function E_2 is learned.

Co-transfer learning with unlabeled probes from the target domain: Similar to the training phase, for matching a LR probe with a HR gallery image, the images are tessellated into non-overlapping local patches and LPQ and SIFT descriptors are computed for each local patch. LPQ descriptors from the corresponding local patches on the gallery and probe images are matched using χ^2 distance and the distance scores from these local patches are vectorized to form an input vector **u**' for view 1. Similarly, an input vector corresponding to SIFT (view 2) is computed using the χ^2 distance measure. Unlike training, the instances obtained during testing are unlabeled. For every query given to the biometric system, both the ensembles, E_1 and E_2, are used to classify the instance. If one ensemble confidently predicts genuine label for an instance while the other ensemble predicts impostor label with low confidence, then this instance is added as a *labeled* re-training sample for the second ensemble and vice-versa. The target domain SVM classifiers (C^T) in the ensembles are updated with pseudo-labeled probe instances obtained during testing. Further, the weights for both source domain and target domain SVM classifiers are also updated with each pseudo-labeled probe instance, as shown in Eqs. 2 and 3. Thus each ensemble updates the target domain classifier of the other ensemble. The final decision is computed by combining responses from both the ensembles.

[4]For images on which the eye-detection failed because of low resolution, normalization was performed manually.

V. Database and Experimental Protocol

The performance of the proposed co-transfer learning framework is evaluated on four different databases, (1) CMU Multi-PIE [21], (2) SCface [22], (3) ChokePoint [23], and (4) Multiple Biometric Grand Challenge (MBGC) v.2 video challenge database [24]. Fig. 7 show sample images from all four databases used in this research are shown in Fig. 7. The experiments are designed to resemble real world scenario where ample training data is available in source domain to train the classifiers for classifying the high resolution gallery-probe pairs as genuine or impostor. However, only a few low resolution probe and corresponding high resolution gallery images are available for training the classifiers in the target domain. To emulate such conditions, Table III lists the number of high resolution gallery-probe pairs that are used for training the classifiers in source domain and the number of low resolution probe and corresponding high resolution gallery images used for training classifiers in the target domain. The training subjects in target domain are a subset of the training subjects in source domain. Further, co-transfer learning and initial training of source and target domain classifiers are performed on non-overlapping subjects. To evaluate the efficacy of the proposed framework, a joint adapt-and-test [44] strategy is used which allows the data used for performance evaluation to be concurrently used for model adaptation. It is to be noted that the proposed framework is first used to classify an unlabeled probe instance and based on the confidence of prediction, this instance may be used as a pseudo-labeled training instance for updating/re-training the ensemble. Therefore, the classification is always performed on unseen instances which is the case with most of the real world applications.

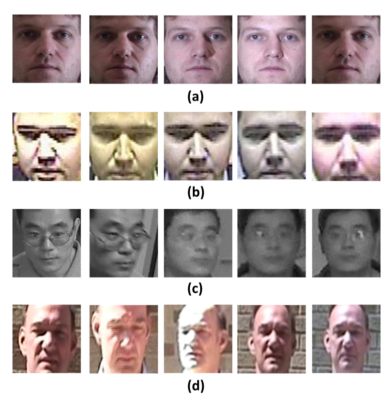


Fig. 7. Sample images from the (a) CMU Multi-PIE, (b) SCface, (c) ChokePoint, and (d) MBGC v.2 video challenge databases.

CMU Multi-PIE [21] database comprises images from 337 individuals captured in four different sessions with varying pose, expression, and illumination. For experiments, a subset

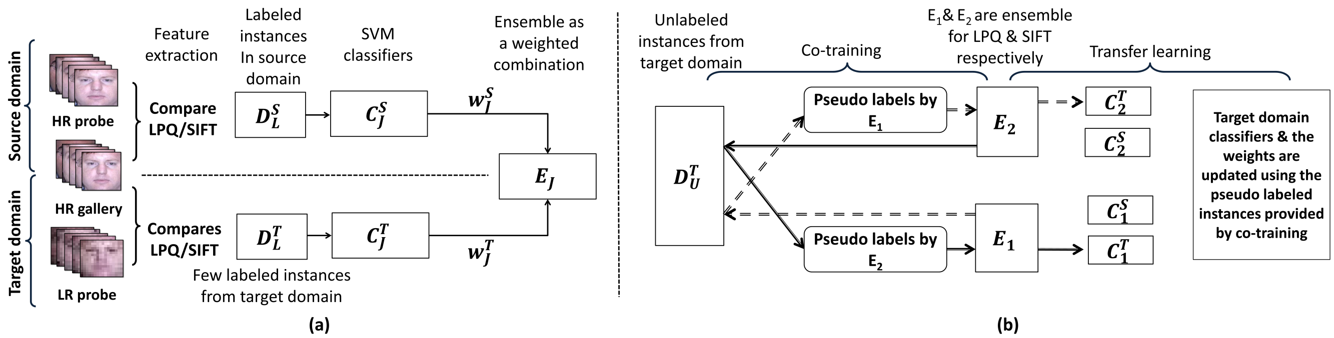


Fig. 6. Block diagram for the co-transfer learning framework for cross-resolution face matching. (a) Illustrates the training process of the source and target domain classifiers to build the ensembles and (b) co-transfer in the target domain with unlabeled probe instances.

TABLE III
EXPERIMENTAL PROTOCOL ON DIFFERENT DATABASES FOR CROSS-RESOLUTION FACE MATCHING. TRAINING SUBJECTS IN THE SOURCE DOMAIN SPECIFIES THE TOTAL NUMBER OF SUBJECTS USED FOR TRAINING DIFFERENT ALGORITHMS. * FOR CHOKEPOINT DATABASE, TRAINING OF SOURCE AND TARGET DOMAIN CLASSIFIERS IS PERFORMED USING THE CMU MULTI-PIE [21] DATABASE. FOR EXPERIMENTS WITH CMU MULTI-PIE, IMAGES ARE SYNTHETICALLY DOWN-SAMPLED TO THE DESIRED RESOLUTION USING LINEAR INTERPOLATION.

Database	No. of subjects in Training		No. of subjects in Testing/Co-transfer learning	Resolution range (pixels)	Covariates (apart from low resolution)
	Source domain	Target domain			
Multi-PIE [21]	100	40	237	216×216 - 16× 16	Illumination
SCface [22]	50	20	80	72×72 - 24×24	Camera distance, pose, & illumination
ChokePoint* [23]	50	20	29	216×216 - 16× 16	Pose, illumination & expression
MBGC v.2 [24]	60	30	87	216×216 - 16× 16	Pose, illumination, walking & talking

pertaining to 337 individuals with frontal pose and neutral expression are selected; however, the gallery and probe images vary in illumination conditions. For each subject, one high resolution image is kept in the gallery and one low resolution image is used as probe.

SCface database is a real-world surveillance database comprising images of 130 individuals captured in uncontrolled indoor environment using multiple surveillance cameras placed at different distances. For each subject, one high resolution image is kept in gallery and five images captured from different cameras are used as probe. SCface database contains low resolution images ranging from 48×48 to 24×24 pixels and experiments are performed without interpolating these images. Therefore in the experimental protocol of the SCface database, gallery and probe images vary from 72×72 to 24×24 pixels.

ChokePoint database is a video database captured under real-world surveillance conditions. Three cameras placed above the portals are used to capture individuals walking through the portal. Images are captured with surveillance cameras in unconstrained environment and include illumination, expression, and pose variations. The database consists of 29 unique subjects and the videos are captured in two portals with a time gap of about one month. Since there are only 29 subjects in the database, training of both source and target domain classifiers is performed using the CMU Multi-PIE database. For each subject in the ChokePoint database, one high resolution image is used as gallery and five images are used as probe.

MBGC v.2 video challenge database used in the experiments contains multiple videos in standard definition (720×480 pixels) and high definition (1440×1080 pixels) format corresponding to 147 subjects are used. The database includes videos where the user is walking or performing some activity. Faces present in these videos have variations due to pose, illumination, and expression. The faces extracted from video frames are partitioned into the gallery and probe data sets (here we ensure that gallery and probe images are from different sessions i.e. from different videos of the person). Gallery consists of single image per user and probe set comprises five images from different sessions.

The application of matching cross-resolution face images is more applicable in an identification (1: N matching) scenario. Following the common protocol in literature [14], [16], [17] the performance of the proposed framework is reported on a closed set identification scenario. Further, to emulate the conditions that the gallery is generally captured under controlled conditions, the experiments are performed with settings such that the resolution of gallery images is always higher than the probe images. Experiments are performed with single image per subject in the gallery. The performance is reported in identification mode with 10 times repeated random sub-sampling (cross-validations) for non-overlapping training-testing partitions. Experiments are performed at different resolutions of gallery and probe images ranging from 216×216 pixels to 16×16 pixels. Face images in the databases are available at different resolutions and are interpolated to the nearest resolution in the experimental protocol using bi-cubic interpolation.

VI. EXPERIMENTAL RESULTS AND ANALYSIS

For cross resolution face matching, the performance of algorithms degrade mainly due to the 1) difference in information content between the high resolution gallery and low resolution probes and 2) limited biometric information in face images at low resolution. The proposed algorithm attempts to address these issues by using the knowledge learned for matching high resolution images from the source domain to efficiently match low resolution images from the target domain. The objective of these experiments is to determine the effectiveness of the proposed algorithm in transferring knowledge from the source domain to target domain for cross resolution face matching. For this, we compare the performance of the proposed algorithm with different algorithms: (1) SIFT with SVM classifier and LPQ with SVM classifier, referred to as SIFT and LPQ in the results, (2) sum-rule score level fusion [63] of two ensembles trained on the initial labeled data from the source and target domains (referred to as 'fusion'), (3) Multidimensional Scaling algorithm (MDS) proposed by Biswas *et al.* [17] for matching low resolution face images, (4) a widely used commercial-off-the-shelf (COTS) face recognition algorithm, FaceVACS, referred to as COTS, (5) three super-resolution techniques, namely super-resolution-1 (SR-1, a standard bicubic interpolation), super-resolution-2 (SR-2[5], a regression based technique proposed by Kim and Kwon [64]), and super-resolution-3 (SR-3[6], a sparse representation based approach proposed by Yang *et al.* [30]), and (6) match score fusion of the proposed algorithm with MDS [17] and COTS using sum-rule [63].

A. Analysis

The results suggest that the proposed approach efficiently matches cross-resolution face images by leveraging knowledge learned in the source domain. It also validates our assertion that co-training enables updating the decision boundary of target domain classifiers with unlabeled probe instances as and when they arrive.

- Cross-pollination of transfer learning and co-training seamlessly transfers the knowledge learned in the source domain for matching cross-resolution face images. Co-training and transfer learning go hand-in-hand as co-training provides pseudo labels for unlabeled test instances which in-turn are used to update the target domain classifiers within each ensemble.
- Updating the weights of the source and target domain classifiers allows to dynamically adjust the contribution from the constituent source and target domain classifiers in an ensemble. Initially, equal weights are assigned to both the classifiers; however with knowledge transfer, weights of classifiers in the target domain become more prominent. Table IV shows the number of instances on which co-transfer learning is performed for different databases. It also shows how the co-transfer learning on unlabeled instances changes the weights of an ensemble so as to better classify the target domain samples. The experiments show that on all four databases combined, co-training provides correct pseudo labels for about 98% of the total instances.
- As more and more pseudo labeled instances are available, the weights for the source and target domain classifiers saturate. Co-transfer learning can converge when the weight saturation occurs i.e., the weights of source domain classifiers become zero, and the emphasis is shifted towards the target domain classifiers. Weight transfer can saturate if and only if (i) the two views are independent (property of co-training); (ii) co-training algorithm yields correct pseudo labels (property of transfer learning); and (iii) large number of samples are available for training. Since, it is challenging to fulfill all three conditions in a real world face recognition scenario, the proposed co-transfer learning algorithm follows the concept of lifelong learning [65] where the classifiers continue to learn and adapt.
- The behavior of the proposed algorithm is further analyzed and Fig. 8(a) illustrates sample cases where the proposed co-transfer learning algorithm correctly recognizes the low resolution probe images. Examples in Fig. 8(b) illustrate cases where the proposed algorithm performs poorly. The poor performance can be attributed to the fact that some of the pseudo labels assigned to unlabeled probe instances may be incorrect leading to *negative transfer*. However, the effect of *negative-transfer* can be minimized by optimally selecting the confidence threshold for co-training. High threshold value implies conservative transfer while smaller value of the threshold leads to aggressive transfer.

[5]Source code is available at authors webpage http://www.mpi-inf.mpg.de/kkim/.

[6]Source code is obtained from http://www.ifp.illinois.edu/jyang29/.

TABLE IV
NUMBER OF INSTANCES ON WHICH CO-TRANSFER LEARNING IS PERFORMED AND HOW THE WEIGHTS WITHIN AN ENSEMBLE SHIFT TO EMPHASIZE THE CONTRIBUTION OF THE TARGET DOMAIN CLASSIFIER.

Database	# pseudo labels		Weights after co-transfer			
	C_1^T	C_2^T	w_1^S	w_1^T	w_2^S	w_2^T
CMU Multi-PIE [21]	5184	4210	0.18	0.82	0.23	0.77
SCface [22]	7346	5268	0.21	0.79	0.27	0.73
ChokePoint [23]	456	540	0.33	0.67	0.36	0.64
MBGC v2 [24]	8136	6874	0.22	0.78	0.24	0.76

Fig. 8. Illustrating sample cases when the proposed approach (a) correctly recognizes and (b) fails to recognize. All the examples are with probe (left image) size 24×24 and gallery (right image) size 72×72.

The subsections below list the performance of individual components of the co-transfer learning algorithm, compare the performance of the proposed algorithm with transformation and super-resolution based approaches, and finally reports the performance on real word surveillance images and Point and Shoot face database [25].

1) Performance of Individual Components: The proposed co-transfer learning algorithm gains from individual components such as co-training, transfer learning and ensembles. To analyze the effect of transfer learning, additional experiments are performed, referred as "HR/LR matching + transfer learning". In this experiment, the source domain comprises HR probe and gallery images. The target domain comprises HR gallery, downsampled HR images as LR probes, and the few labeled examples are available from the target domain. In Tables V-VIII, "HR/LR TL (LPQ)" and "HR/LR TL (SIFT)" refers to experiments on LPQ and SIFT features respectively. "HR/LR TL LPQ + SIFT" refers to match score level fusion. These experiments are trained in a supervised manner unlike the proposed co-transfer learning algorithm which uses semi-supervised learning. In this case, the ensembles on the two views work independent of each other. The target domain classifier and weights for the two components in an ensemble are updated with labeled instances in the target domain and the synthetic data obtained by downsampling the source domain data. Similarly, to analyze the effect of co-training, "HR/LR matching + co-training" experiments are performed, referred to as "HR/LR CT" in Tables V-VIII. In this experiment, large number of instances from the source domain are combined with few labeled instances from the target domain for training the classifier. Here, one classifier is trained on all the data available for initial training. We use two classifiers trained on two separate views of the training data which then co-train each other with the additional unlabeled instances from the target domain. The results in Tables V-VIII report the performance of different components.

- HR/LR + transfer learning on the two views (i.e. SIFT and LPQ) yields better or comparable results compared to the individual ensembles and their fusion. It is observed that HR/LR + transfer learning on LPQ gives better performance as compared to HR/LR + transfer learning on SIFT. However, the proposed CTL algorithm still outperforms HR/LR + transfer learning on individual views as it combines the complimentary information from both the ensembles. HR/LR + transfer learning also outperforms the MDS algorithm.
- The performance of the proposed CTL algorithm is better than the performance of match score level sum rule fusion of HR/LR + transfer learning on SIFT and LPQ. The gain in performance is attributed to the fact that in the proposed co-transfer learning, the two ensembles co-train each other on the target domain instances. In HR/LR + transfer learning on LPQ and SIFT, the two classifiers are trained independent of each other. Moreover, the downsampled instances from the source domain do not greatly facilitate the classifiers to improve the performance on the target domain instances. This validates our assertion that downsampling source domain instances may lead to loss of information which is useful for face recognition.
- The performance of HR/LR + co-training is lower as compared to HR/LR + transfer learning as the few labeled target domain data is over shadowed by large number of labeled source domain data during training. It is a semi-supervised approach where unlabeled data from the target domain has to be transformed into pseudo labeled instances to co-train the classifiers on separate views.

2) Comparison with COTS and Transformation based Approaches: The performance of the proposed co-transfer learning (CTL) algorithm is compared with MDS [17], COTS, individual ensembles of SIFT [62], and LPQ [61], and their fusion. The results are also evaluated by fusing the proposed CTL algorithm with other techniques such as MDS [17] and COTS. Tables V-VIII show the results of the proposed and existing algorithms with different combinations of gallery-probe resolution on the four databases.

The Cumulative Match Characteristics (CMC) curves in Fig. 9 show the performance of different algorithms for matching probe images of resolution 24×24 with gallery images of resolution 72×72. As compared to fusion of two ensembles, the knowledge transfer from the source to target domain improves the accuracy by at least 4-5%. During initial training, since the source and target domain classifiers are trained independently, the knowledge transfer is not available in an ensemble. It is feasible only with pseudo labeled probe instances available in the target domain during testing. Table V shows the results on the CMU Multi-PIE database. The images in the CMU Multi-PIE database are of very high quality and therefore the results on this database may not be representative of cross resolution face matching with surveillance quality databases. However, since the previous research on low resolution face recognition has shown the results on the CMU Multi-PIE database, we are using this database (along with three surveillance databases) to establish the baseline comparison with MDS. The results show that for high resolutions, COTS performs better than the proposed CTL and MDS algorithms. However, the performance of the commercial system reduces significantly on reducing the resolution of probe images. On the contrary, the performance of CTL reduces at a lower rate and it yields better results than COTS when the probe image is of resolution 16×16.

Table VI shows the results on the SCface database [22]. The proposed algorithm yields promising results on the real-world surveillance database and even outperforms COTS by at least 24% on all combinations of gallery and probe resolutions. Since the proposed algorithm uses SIFT and LPQ features that are resilient to pose variations and changes in gray-level intensities due to illumination variations, it inherently addresses the problem of head-pose and illumination variations in the SCface database. Moreover, the knowledge transfer with unlabeled probe instances in the target domain facilitates to efficiently classify the low resolution probes. Tables VII and VIII illustrate the performance on the ChokePoint [23] and MBGC v.2 video challenge [24] databases respectively. On both the databases, the proposed algorithm performs better than the existing algorithms and COTS for all combinations of gallery-probe resolutions (except for gallery 216×216 and probe 72×72, where COTS gives better performance).

From the results shown on the three surveillance databases, it can be inferred that for high resolution gallery-probe pairs, COTS performs better than the proposed algorithm. However, for lower resolutions, the proposed algorithm yields better results. The performance of transformation based approaches such as MDS [17] degrades when the difference in resolution of gallery and probe images increases (i.e. matching gallery images of 216×216 with probe image of resolution 32×32 or

TABLE V

RANK-1 IDENTIFICATION ACCURACY OF THE PROPOSED CTL ALGORITHM AND COMPARISON WITH EXISTING ALGORITHMS AND COMMERCIAL SYSTEM ON THE CMU MULTI-PIE DATABASE [21]. SIFT WITH SVM CLASSIFIER AND LPQ WITH SVM CLASSIFIER ARE REFERRED TO AS **SIFT** AND **LPQ** RESPECTIVELY, **E1** REFERS TO ENSEMBLE 1, **E2** REFERS TO ENSEMBLE 2, **FUSION** REFERS TO SUM RULE FUSION OF TWO ENSEMBLES, **MDS** REFERS TO MULTIDIMENSIONAL SCALING ALGORITHM PROPOSED BY BISWAS *et al.* [17], **COTS** (COMMERCIAL-OFF-THE-SHELF) REFER TO FACEVACS, AND **CTL** IS USED FOR THE PROPOSED CO-TRANSFER LEARNING ALGORITHM. **HR/LR TL (LPQ)** REFERS TO HR/LR MATCHING + TRANSFER LEARNING VIA LPQ FEATURES AND **HR/LR TL (SIFT)** REFERS TO HR/LR MATCHING + TRANSFER LEARNING VIA SIFT FEATURES, **HR/LR TL LPQ + SIFT** REFERS TO MATCH SCORE LEVEL FUSION OF THE TWO APPROACHES, AND HR/LR MATCHING + CO-TRAINING EXPERIMENTS ARE REFERRED TO AS **HR/LR CT**.

Resolution		Algorithm													
Gallery	**Probe**	**LPQ**	**SIFT**	**E1**	**E2**	**Fusion**	**MDS**	**CTL**	**HR/ LR TL (LPQ)**	**HR/ LR TL (SIFT)**	**HR/LR TL LPQ+ SIFT**	**HR/ LR CT**	**COTS**	**CTL+ MDS**	**CTL+ COTS**
216×216	72×72	66.3	61.7	72.4	68.1	76.2	77.8	81.0	72.5	76.6	79.2	69.6	99.5	80.2	**99.8**
	48×48	63.6	58.2	70.6	67.3	74.5	75.2	79.7	71.2	74.4	77.4	68.4	98.1	79.4	**99.3**
	32×32	45.4	41.8	53.2	47.4	58.7	61.3	65.3	53.4	56.3	61.8	49.2	97.4	63.7	**98.5**
	24×24	22.2	21.4	29.5	26.8	32.9	33.4	37.7	27.8	30.4	34.4	28.5	54.5	35.6	**58.2**
	16×16	10.8	9.6	16.7	13.3	18.1	20.2	23.6	13.6	17.5	20.5	15.8	10.9	22.1	**24.8**
72×72	48×48	73.8	71.4	79.4	76.3	86.1	89.2	92.3	78.2	85.4	89.6	79.2	98.2	92.7	**99.1**
	32×32	62.8	49.8	69.1	55.2	79.4	81.5	84.1	73.4	78.3	81.7	58.5	96.3	84.3	**97.4**
	24×24	56.8	52.6	61.8	59.4	70.3	75.7	77.4	64.3	69.2	73.3	62.4	64.5	78.5	**80.1**
	16×16	50.2	47.4	56.7	52.1	66.2	68.9	72.4	60.8	66.5	69.6	55.6	11.5	72.8	**76.1**
48×48	32×32	44.2	42.5	50.3	47.8	55.2	58.7	61.8	51.6	55.2	58.2	49.8	96.8	60.5	**97.1**
	24×24	42.6	39.8	48.6	44.5	51.7	54.9	57.1	46.4	50.6	54.4	48.5	75.9	55.8	**78.5**
	16×16	20.6	18.2	26.2	22.3	29.9	31.3	32.9	23.8	28.6	30.5	25.2	6.4	39.4	**43.2**
32×32	24×24	37.6	30.1	41.2	30.4	44.8	40.9	45.7	38.6	44.4	45.3	33.4	78.4	45.4	**80.6**
	16×16	22.1	16.8	24.3	17.2	27.0	25.1	28.1	22.8	26.6	27.4	19.6	5.4	29.8	**30.0**
24×24	16×16	30.8	26.4	35.6	30.2	42.1	38.1	43.2	36.5	41.8	42.5	33.5	16.3	44.6	**47.8**

TABLE VI

RANK-1 IDENTIFICATION ACCURACY OF THE PROPOSED CTL ALGORITHM AND COMPARISON WITH EXISTING ALGORITHMS AND COMMERCIAL SYSTEM ON THE SCFACE DATABASE [22].

Resolution		Algorithm													
Gallery	**Probe**	**LPQ**	**SIFT**	**E1**	**E2**	**Fusion**	**MDS**	**CTL**	**HR/ LR TL (LPQ)**	**HR/ LR TL (SIFT)**	**HR/LR TL LPQ+ SIFT**	**HR/ LR CT**	**COTS**	**CTL+ MDS**	**CTL+ COTS**
72×72	48×48	58.4	55.8	63.2	60.4	74.4	76.1	79.4	75.8	65.6	77.2	62.8	35.7	80.4	**83.4**
	32×32	53.4	52.3	58.1	57.8	67.4	70.4	72.8	69.1	61.3	71.6	60.2	18.5	73.7	**76.2**
	24×24	48.1	43.5	52.6	49.1	60.2	64.8	66.4	61.6	54.8	64.5	52.3	10.3	67.6	**70.1**
48×48	32×32	36.2	32.6	40.2	36.5	45.8	47.9	50.0	46.8	42.7	48.8	38.6	23.8	50.6	**54.3**
	24×24	25.6	24.2	30.2	28.3	35.6	38.1	40.3	36.8	33.6	38.3	30.8	14.5	39.5	**45.1**
32×32	24×24	22.5	17.3	26.4	21.3	29.7	31.2	33.1	31.6	29.4	32.0	24.4	8.4	33.9	**36.2**

TABLE VII

RANK-1 IDENTIFICATION ACCURACY OF THE PROPOSED CTL ALGORITHM AND COMPARISON WITH EXISTING ALGORITHMS AND COMMERCIAL SYSTEM ON THE CHOKEPOINT DATABASE [23].

Resolution		Algorithm													
Gallery	**Probe**	**LPQ**	**SIFT**	**E1**	**E2**	**Fusion**	**MDS**	**CTL**	**HR/ LR TL (LPQ)**	**HR/ LR TL (SIFT)**	**HR/LR TL LPQ+ SIFT**	**HR/ LR CT**	**COTS**	**CTL+ MDS**	**CTL+ COTS**
216×216	72×72	32.2	28.6	36.3	32.5	39.8	41.6	44.6	41.2	38.2	42.7	34.6	46.2	43.2	**50.9**
	48×48	23.1	22.1	29.6	28.1	31.5	33.8	38.4	33.0	31.5	35.6	30.4	33.7	36.8	**42.3**
	32×32	21.8	21.8	27.3	25.7	30.6	32.5	35.5	32.2	29.4	33.4	28.1	20.4	34.1	**39.5**
	24×24	18.4	16.2	23.2	20.7	28.4	29.1	32.4	29.8	26.8	30.8	23.0	10.3	31.7	**35.1**
	16×16	9.6	8.2	14.7	11.2	15.6	17.8	20.2	17.1	16.6	18.3	12.8	6.04	19.3	**23.4**
72×72	48×48	42.4	36.1	48.4	42.6	50.5	50.9	53.7	52.2	50.4	53.4	45.3	22.7	53.1	**56.4**
	32×32	32.6	31.8	37.6	35.7	39.5	41.6	43.8	41.2	39.2	42.0	38.1	12.7	42.6	**47.2**
	24×24	25.4	23.6	30.5	28.9	31.6	32.4	36.1	33.0	32.6	34.2	30.8	9.5	34.8	**39.5**
	16×16	21.4	19.6	26.2	23.8	28.1	28.7	31.6	29.8	28.4	30.3	25.6	7.6	30.4	**35.2**
48×48	32×32	35.4	32.6	41.2	37.6	44.7	45.4	48.2	46.6	43.4	47.1	40.4	18.5	47.8	**50.9**
	24×24	23.2	20.4	27.4	24.8	29.5	30.2	33.1	31.6	29.1	32.8	27.1	11.8	32.6	**37.2**
	16×16	17.6	14.5	21.8	19.6	24.1	26.3	28.3	25.8	23.6	26.5	22.7	4.7	27.5	**31.6**
32×32	24×24	20.4	14.8	23.4	18.7	24.3	28.6	31.6	26.2	25.6	29.4	21.3	16.4	30.8	**35.4**
	16×16	14.6	9.6	17.3	13.4	19.6	21.9	23.1	21.1	19.2	21.8	15.6	3.5	22.5	**26.0**
24×24	16×16	19.4	15.6	22.7	18.6	25.8	28.7	30.5	27.4	24.2	29.1	20.8	13.5	31.4	**35.8**

TABLE VIII
RANK-1 IDENTIFICATION ACCURACY OF THE PROPOSED CTL ALGORITHM AND COMPARISON WITH EXISTING ALGORITHMS AND COMMERCIAL SYSTEM ON THE MBGC V.2 VIDEO CHALLENGE DATABASE [24].

Resolution		Algorithm													
Gallery	Probe	LPQ	SIFT	E1	E2	Fusion	MDS	CTL	HR/ LR TL (LPQ)	HR/ LR TL (SIFT)	HR/LR TL LPQ+ SIFT	HR/ LR CT	COTS	CTL+ MDS	CTL+ COTS
216×216	72×72	27.2	25.4	30.8	28.2	33.4	36.5	40.7	32.6	35.3	37.0	30.4	44.3	39.2	**47.3**
	48×48	22.6	24.8	26.2	23.7	29.3	30.8	33.5	28.4	31.8	32.5	25.5	31.4	32.7	**36.8**
	32×32	20.8	17.2	23.6	20.9	26.1	28.4	32.6	25.7	28.4	30.1	22.7	18.5	31.4	**35.2**
	24×24	17.6	15.4	21.5	18.6	23.7	25.3	28.1	23.5	25.2	26.8	20.8	9.8	26.9	**29.5**
	16×16	9.6	8.8	12.5	10.1	14.8	16.8	19.5	14.2	16.5	18.2	12.4	5.7	18.7	**20.8**
72×72	48×48	38.2	33.6	43.1	39.7	45.3	46.8	49.3	45.8	47.8	48.1	42.1	21.2	48.6	**50.7**
	32×32	29.2	26.4	33.4	29.1	35.9	38.3	41.9	35.1	38.1	40.8	31.5	11.4	40.5	**45.2**
	24×24	23.6	20.4	26.3	22.5	28.7	29.8	33.2	28.2	31.2	32.0	24.4	8.7	31.5	**36.9**
	16×16	19.6	16.8	22.7	19.4	24.8	26.5	29.5	24.8	26.6	27.5	21.3	6.2	28.1	**33.5**
48×48	32×32	33.6	31.2	38.4	33.1	40.5	44.3	47.4	40.7	42.7	45.6	35.8	17.2	46.8	**48.7**
	24×24	21.4	20.2	24.6	21.3	25.8	27.6	30.3	26.6	27.6	28.4	23.5	10.2	29.4	**33.5**
	16×16	16.5	14.8	18.2	15.7	20.3	24.1	26.5	20.4	22.5	24.1	17.4	4.2	26.1	**27.9**
32×32	24×24	17.8	12.6	20.7	15.6	22.3	27.1	28.6	22.6	24.2	26.2	17.7	14.8	29.2	**33.2**
	16×16	12.6	8.6	14.6	10.9	16.3	19.8	21.3	16.2	18.4	19.8	12.3	3.1	20.6	**23.9**
24×24	16×16	18.4	14.8	20.4	16.1	23.5	27.3	28.2	22.5	25.6	27.9	18.0	11.9	29.4	**31.8**

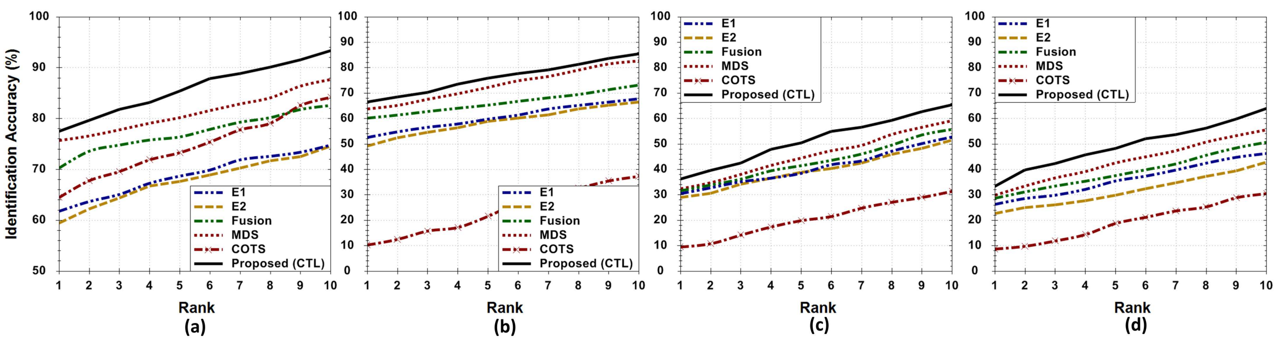


Fig. 9. CMC curves showing the performance for matching 24×24 probe images with 72×72 gallery images on the a) CMU Multi-PIE, b) SCface, c) ChokePoint, and d) MBGC v.2 video challenge databases.

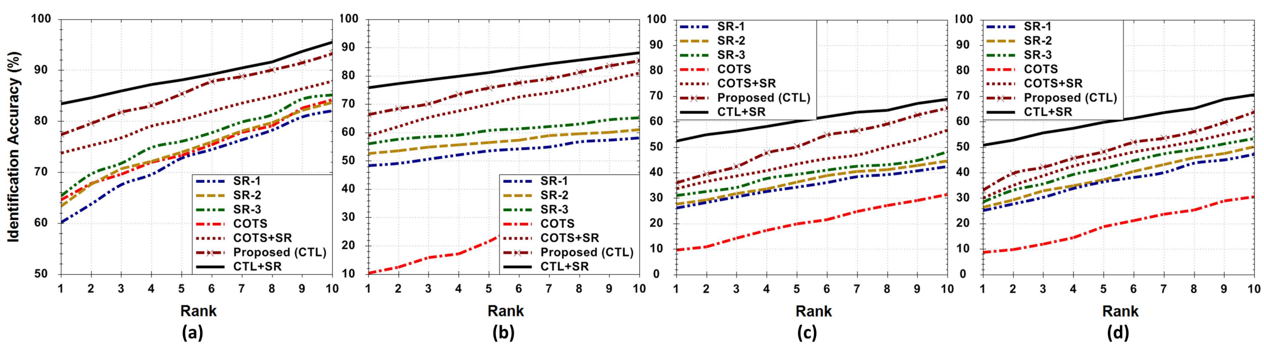


Fig. 10. CMC curves comparing the performance of the proposed algorithm with three super resolution techniques on the (a) CMU Multi-PIE, (b) SCface, (c) ChokePoint, and (d) MBGC v.2 video challenge databases. Probe images of 24×24 pixels are super-resolved by a magnification factor of three to match the gallery resolution of 72×72 pixels.

lower). The transformations learned for such wide variations in gallery-probe resolution may not be precise and thus degrade the performance. Experimental results in Tables V-VIII also show that sum-rule fusion [63] of the proposed algorithm with COTS further enhances the performance of cross-resolution face matching. This improvement in performance may be attributed to the combined effect of COTS and CTL. COTS efficiently addresses the difference in the information content at higher resolutions, while CTL addresses the problem of limited biometric information at low resolution images. On the contrary, sum-rule fusion of the proposed CTL with MDS [17] slightly degrades the performance as it may not efficiently accommodate for large difference in information content between the gallery-probe pairs.

For experiments on the CMU Multi-PIE database, there is only a single gallery and a single probe image from the target domain. The proposed co-transfer learning can efficiently transfer the knowledge to better learn the target domain decision boundary using only a single gallery and single probe image (completely unseen training and testing). However, this condition is slightly relaxed for experiments with other three databases which contain multiple probes per subject in the target domain and the classifiers may be trained on subjects with these multiple probes in an incremental manner. It is to be noted that the probe images are first used for recognition (testing) and then used to update the weights in an incremental semi-supervised manner. Therefore, the probe images are incorporated in training only after they have classified - thereby maintaining the unseen nature of training and testing databases.

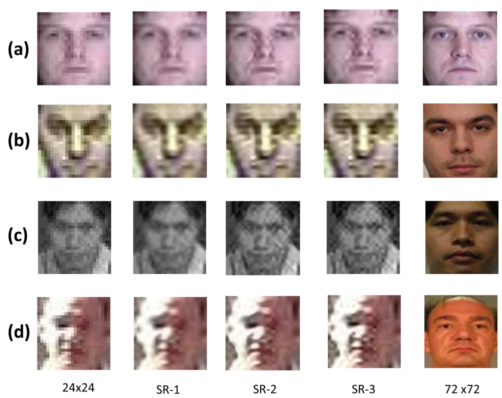


Fig. 11. Enhanced images obtained using three super-resolution techniques (SR-1,SR-2, and SR-3). The leftmost column represents low resolution (24×24) images and the rightmost column represents the original high resolution images (72×72) from the (a) CMU Multi-PIE, (b) SCface, (c) ChokePoint, and (d) MBGC v.2 video challenge databases.

3) Comparison with Super-resolution Approaches: In this section, the performance of the proposed co-transfer learning algorithm is compared with three super-resolution techniques proposed in literature. For evaluating the effectiveness of super-resolution techniques for matching low and high resolution face images, it is used as a pre-processing step to enhance the quality of low resolution face images before matching. Some examples of enhanced images are shown in Fig. 11. The enhanced image is matched with high resolution gallery image using different algorithms. The LPQ and SIFT features are extracted from the super-resolved images and the performance is computed after sum-rule fusion [63] of LPQ and SIFT match scores computed using the χ^2 distance metric. For evaluating the performance with the proposed technique and COTS, super-resolution based on sparse representation (SR-3) is applied on the probe images and then feature extracting and matching are performed using the CTL algorithm (referred to as "CTL+SR") and COTS (referred to as "COTS+SR"). The target domain thus includes enhanced images obtained using super-resolution. It is to be noted that transfer learning is still applicable as super-resolution introduces several artifacts that may affect the biometric information in a face image and leads to variations in data distribution (of features or match scores) between the source and target domains. The classifiers in target domain are now trained to match the enhanced probe images with HR gallery. For the experiments, super-resolution is performed with a magnification factor of three to match probe images of size 24×24 with 72×72 gallery images. CMC curves in Fig. 10 show that the proposed co-transfer learning algorithm outperforms all three super-resolution techniques by at least 11% on the CMU Multi-PIE database, 10% on the SCface database, and 4% on the ChokePoint and MBGC v.2 video challenge databases. Further, as shown in Fig. 10, enhancing probe images using super-resolution boosts the performance of both CTL and COTS. It is observed that super-resolution minimizes the difference in the resolutions of gallery and probe images. However, it does not enhances the biometric information in face images. Therefore, the performance gain is constrained by limited biometric information in low resolution face images.

B. Performance on Real World Cases

Recently, Klontz and Jain [66] have investigated the opportunity for face recognition algorithms to facilitate law enforcement agencies in identifying individuals from the crime scene CCTV footage during the Boston bombings incident. Inspired by their study, the performance of the proposed co-transfer learning algorithm is also evaluated on some real world examples pertaining to cross-resolution face matching. In our experiments, some real world examples are collected from different sources on the internet which includes two individuals from Boston bombing [66], [67], four individuals from London bombing [1] and one individual from Mumbai terrorist attack [2]. Fig. 12 shows the low resolution probes and corresponding gallery images considered in the experiment. In this additional experiment for evaluating the performance with these seven real world examples, we appended these images to the SCface database for co-transfer learning. The experiments are performed with gallery image resolution of 72×72 pixels and query image resolution of 32×32 pixels. Each individual has one image in the gallery and one or more low resolution images as probe. Further, an extended gallery of 6534 individuals is created by using frontal images acquired from a law enforcement agency and appending it to the gallery of the SCface database. The performance of the proposed co-transfer learning algorithm is also compared with COTS for matching 15 probe images corresponding to these 7 real world cases. The results in Table IX show that the proposed algorithm consistently retrieves the correct match at a lower rank than COTS[7] on all the cases. The results validate our initial assertion that the proposed co-transfer learning algorithm can efficiently be coupled with surveillance systems to assist law enforcement agencies.

[7] Since, the eye region is occluded in some of the probe images, COTS is not able to process such cases (represented as NP - Not Processed).

TABLE IX
RESULTS FOR MATCHING REAL WORLD EXAMPLES AGAINST A LARGE SCALE GALLERY OF 6534 INDIVIDUALS. VALUES IN THE TABLE REPRESENTS THE RANK AT WHICH THE CORRECT IDENTITY IS RETRIEVED. NP REPRESENTS THE CASES WHICH ARE NOT PROCESSED BY THE COTS.

Algorithm	Probe Id														
	1a	1b	2a	2b	2c	3a	3b	4a	5a	5b	6a	6b	6c	7a	7b
CTL	7	29	8	17	1	15	5	19	17	1	1	10	18	2	4
COTS	NP	NP	11	26	3	28	9	22	NP	1	1	14	NP	4	NP

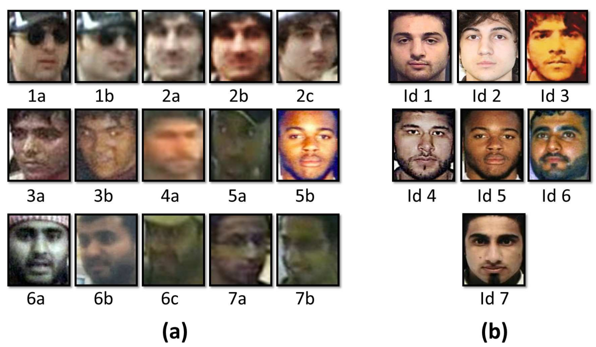


Fig. 12. Real world cases for cross-resolution face matching: (a) low resolution probe images and (b) corresponding gallery images.

C. Performance on Point and Shoot Still Challenge Database

The performance of the proposed algorithm is also evaluated on the still matching challenge of Point and Shoot Challenge (PaSC) [25] database. The experiments are performed on the frontal part of the still face database and the predefined protocol of PaSC is followed except the resolution of images. In PaSC, the size of both gallery and probe images is same; however, the proposed algorithm requires the two to be of different resolutions. Therefore, we have set the probe image size to 24×24 whereas the gallery images are of original size 72×72. Under this environment, the proposed algorithm yields the verification accuracy of $\sim$47% at 1% false accept rate and the COTS yields a significantly lower accuracy of $\sim$33%. It is to be noted that these results cannot be compared with the ones provided in [25] because the resolution of the probe images is different from our experiments.

VII. CONCLUSION AND FUTURE WORK

The paper introduces a co-transfer learning framework which seamlessly combines the co-training and transfer learning paradigms for efficient cross-resolution face matching. During training, the proposed framework learns to match high resolution face images in the source domain. This knowledge is then transferred from the source domain to the target domain to match low resolution probes with high resolution gallery. The proposed framework builds ensembles from the weighted combination of source and target domain classifiers on two separate views. Two ensembles trained on separate views transform the unlabeled probe instances into pseudo-labeled instances using co-training. These pseudo-labeled instances are utilized for updating the decision boundary of the target domain classifier, thus, transferring knowledge from the source domain to the target domain. Further, dynamically updating the weights assigned to each classifier facilitates gradual shift of knowledge from the source to target domain. The amalgamation of transfer learning and co-training helps to transfer knowledge from the source to target domain with probe instances as and when they arrive. Comprehensive analysis, including comparison with existing cross-resolution face matching algorithms, super-resolution techniques, and a commercial face recognition system, is performed for different gallery-probe resolutions ranging from 216×216 to 16×16 pixels. The proposed co-transfer learning framework provides significant improvement for cross-resolution face matching on different surveillance quality face databases. As future research directions, this research can be extended for addressing cross resolution face recognition under face age and weight, disguise, and plastic surgery variations. It can further be extended for cross resolution face recognition in videos where several of frames can be of very poor quality and resolution [68].

REFERENCES

[1] "http://www.cbc.ca/news/background/london_bombing/investigation_timeline.html", (last accessed: January, 5, 2013).
[2] "http://www.hindustantimes.com/india-news/newdelhi/who-s-keeping-watch/article1-908391.aspx", (last accessed: January, 5, 2013).
[3] "http://www.lawisgreek.com/can-surveillance-cameras-preventdeter-terrorist-acts", (last accessed: January, 5, 2013).
[4] S. Li and A. Jain, *Handbook of face recognition*, Springer, 2005.
[5] M. Singh, S. Nagpal, R. Singh, and M. Vatsa, "On recognizing face images with weight and age variations", *IEEE Access*, vol. 2, pp. 822–830, 2014.
[6] R. Singh, M. Vatsa, and A. Noore, "Face recognition with disguise and single gallery images", *Image Vision Computing*, vol. 27, pp. 245–257, 2009.
[7] T. Dhamecha, R. Singh, M. Vatsa, and A. Kumar, "Recognizing disguised faces: Human and machine evaluation", *PLoS ONE*, vol. 9, no. 7, pp. e99212, 2014.
[8] R. Singh, M. Vatsa, H.S. Bhatt, S. Bharadwaj, A. Noore, and S.S. Nooreyezdan, "Plastic surgery: A new dimension to face recognition", *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 441 –448, sep. 2010.
[9] H.S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, "Recognizing surgically altered face images using multiobjective evolutionary algorithm", *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 89–100, 2013.
[10] H. Huang and H. He, "Super-resolution method for face recognition using nonlinear mappings on coherent features", *IEEE Transactions on Neural Networks*, vol. 22, no. 1, pp. 121–130, 2011.
[11] K. Jia and S. Gong, "Multi-modal tensor face for simultaneous super-resolution and recognition", in *Proceedings of International Conference on Computer Vision*, 2005, pp. 1683–1690.
[12] P. H. Hennings-Yeomans, S. Baker, and V. Bhagavatula, "Simultaneous super-resolution and feature extraction for recognition of low-resolution faces", in *Proceedings of Computer Vision and Pattern Recognition*, 2008, pp. 1–8.
[13] W. W. W. Zou and P. C. Yuen, "Very low resolution face recognition problem", *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 327–340, 2012.
[14] Z. Lei, T. Ahonen, M. Pietikainen, and S. Li, "Local frequency descriptor for low-resolution face recognition", in *Proceedings of International Conference on Automatic Face Gesture Recognition and Workshops*, 2011, pp. 161–166.

[15] B. Li, H. Chang, S. Shan, and X. Chen, “Low-resolution face recognition via coupled locality preserving mappings”, *IEEE Signal Processing Letters*, vol. 17, no. 1, pp. 20–23, 2010.

[16] S. Shekhar, V. M. Patel, and R. Chellappa, “Synthesis-based recognition of low resolution faces”, in *International Joint Conference on Biometrics*, 2011, pp. 1–6.

[17] S. Biswas, K. W. Bowyer, and P. J. Flynn, “Multidimensional scaling for matching low-resolution face images”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 2019–2030, 2012.

[18] H. Bhatt, M. Singh, R. Vatsa, and N. Ratha, “Matching cross-resolution face images using co-transfer learning”, in *Proceedings of International Conference on Image Processing*, 2012, pp. 1–4.

[19] S. J. Pan and Q. Yang, “A survey on transfer learning”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[20] A. Blum and T. Mitchell, “Combining labeled and unlabeled data with co-training”, in *Proceedings of Conference on Learning Theory*, 1998, pp. 92–100.

[21] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, “Multi-PIE”, *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[22] M. Grgic, K. Delac, and S. Grgic, “SCface- surveillance cameras face database”, *Multimedia Tools and Applications*, vol. 51, no. 3, pp. 863–879, 2011.

[23] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, “Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition”, in *Proceedings of Computer Vision and Pattern Recognition Workshops*, 2011, pp. 74–81.

[24] “MBGC v2: available at http://www.nist.gov/itl/iad/ig/mbgc.cfm”.

[25] J.R. Beveridge, P.J. Phillips, D.S. Bolme, B.A Draper, G.H. Givens, Yui Man Lui, M.N. Teli, Hao Zhang, W.T. Scruggs, K.W. Bowyer, P.J. Flynn, and Su Cheng, “The challenge of face recognition from digital point-and-shoot cameras”, in *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems,*, 2013.

[26] S. Baker and T. Kanade, “Limits on super-resolution and how to break them”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.

[27] A. Chakrabarti, A.N. Rajagopalan, and R. Chellappa, “Super-resolution of face images using kernel pca-based prior”, *IEEE Transactions on Multimedia*, vol. 9, no. 4, pp. 888–892, 2007.

[28] H. Chang, D. Yeung, and Y. Xiong, “Super-resolution through neighbor embedding”, in *Proceedings of Computer Vision and Pattern Recognition*, 2004, pp. 275–282.

[29] B. Li, H. Chang, S. Shan, and X. Chen, “Locality preserving constraints for super-resolution with neighbor embedding”, in *Proceedings of International Conference on Image Processing*, 2009, pp. 1189–1192.

[30] J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation”, *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[31] C. Liu, H. Shum, and W. Freeman, “Face hallucination: Theory and practice”, *International Journal of Computer Vision*, vol. 75, no. 1, pp. 115–134, 2007.

[32] S. Biswas, G. Aggarwal, and P. J. Flynn, “Pose-robust recognition of low-resolution face images”, in *Proceedings of Computer Vision and Pattern Recognition*, 2011, pp. 601–608.

[33] Z. Lei, S. Liao, A. K. Jain, and S. Z. Li, “Coupled discriminant analysis for heterogeneous face recognition”, *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1707–1716, 2012.

[34] Y. Zhu, Y. Chen, Z. Lu, S. J. Pan, G. Xue, Y. Yu, and Q. Yang, “Heterogeneous transfer learning for image classification.”, in *Proceedings of AAAI Conference on Artificial Intelligence*, 2011, pp. 1304–1309.

[35] A. Quattoni, M. Collins, and T. Darrell, “Transfer learning for image classification with sparse prototype representations”, in *Proceedings of Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[36] A. Ahmed, K. Yu, W Xu, Y. Gong, and E. Xing, “Training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks”, in *Proceedings of European Conference on Computer Vision*, 2008, pp. 69–82.

[37] B. Geng, D. Tao, and C. Xu, “Daml: Domain adaptation metric learning”, *IEEE Transactions on Image Processing*, vol. 20, no. 10, pp. 2980–2989, 2011.

[38] H. Wang, F. Nie, H. Huang, and C. Ding, “Dyadic transfer learning for cross-domain image classification”, in *Proceedings of International Conference on Computer Vision*, 2011, pp. 551–556.

[39] X. Siyu, S. Ming, and F. Yun, “Kinship verification through transfer learning”, in *Proceedings of International Joint Conference on Artificial Intelligence*, 2011, pp. 2539–2544.

[40] J. Chen, X. Liu, P. Tu, and A. Aragones, “Person-specific expression recognition with transfer learning”, in *Proceedings of International Conference on Image Processing*, 2012, pp. 2621–2624.

[41] H. S. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, A. Ross, and A. Noore, “On co-training online biometric classifiers”, in *Proceedings of International Joint Conference on Biometrics*, 2011, pp. 1–6.

[42] X. Cao, D. Wipf, F. Wen, G. Duan, and J. Sun, “A practical transfer learning algorithm for face verification”, in *Proceedings of International Conference on Computer Vision*, 2013, pp. 3208–3215.

[43] M. K. Ng, Q. Wu, and Y. Ye, “Co-transfer learning via joint transition probability graph based method”, in *Proceedings of International Workshop on Cross Domain Knowledge Discovery in Web and Social Network Mining*, 2012, pp. 1–9.

[44] P. Zhao and S. Hoi, “OTL: A framework of online transfer learning”, in *Proceedings of International Conference on Machine Learning*, 2010, pp. 1231–1238.

[45] Z. Guo and Z. J. Wang, “Cross-domain object recognition via input-output kernel analysis”, *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3108–3119, Aug 2013.

[46] S. Abney, “Bootstrapping”, in *Proceedings of the Association for Computational Linguistics*, 2002, pp. 360–367.

[47] W. Wang and Z. H. Zhou, “A new analysis of co-training”, in *Proceedings of International Conference on Machine Learning*, 2010, pp. 1135–1142.

[48] X. Jiang and W. Ser, “Online fingerprint template improvement”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1121–1126, 2002.

[49] F. Roli, L. Didaci, and G. Marcialis, “Adaptive biometric systems that can improve with use”, in *Proceedings of Advances in Biometrics: Sensors, Systems and Algorithms*, 2008, pp. 447–471.

[50] C. Ryu, H. Kim, and A. K. Jain, “Template adaptation based fingerprint verification”, in *Proceedings of International Conference on Pattern Recognition*, 2006, pp. 582–585.

[51] N. Poh, R. Wong, J. Kittler, and F. Roli, “Challenges and research directions for adaptive biometric recognition systems”, in *Proceedings of International Conference on Advances in Biometrics*, 2009, pp. 753–764.

[52] D. Rim, K. Hassan, and C. J. Pal, “Semi supervised learning for wild faces and video”, in *Proceedings of British Machine Vision Conference*, 2011, pp. 1–12.

[53] D. Cai, X. He, and J.i Han, “Semi-supervised discriminant analysis”, in *International Conference on Computer Vision*, 2007, pp. 1–7.

[54] R. Gross, L. Sweeney, F. Torre, and S. Baker, “Semi-supervised learning of multi-factor models for face de-identification”, in *Proceedings of Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[55] Y. Zhang and D. Yeung, “Semi-supervised discriminant analysis using robust path-based similarity”, in *Proceedings of Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[56] F. Roli and G. Marcialis, “Semi-supervised PCA-based face recognition using self-training”, in *Structural, Syntactic, and Statistical Pattern Recognition*, pp. 560–568. 2006.

[57] X. Zhao, N. Evans, and J. Dugelay, “Semi-supervised face recognition with LDA self-training”, in *Proceedings of International Conference on Image Processing*, 2011, pp. 3041–3044.

[58] R. Singh, M. Vatsa, A. Ross, and A. Noore, “Biometric classifier update using online learning: A case study in near infrared face verification”, *Image and Vision Computing*, vol. 28, no. 7, pp. 1098–1105, 2010.

[59] G. Cauwenberghs and T. Poggio, “Incremental and decremental support vector machine learning”, in *Proceedings of Advances in Neural Information Processing Systems*, 2000, pp. 409–415.

[60] Z. H. Zhou and M. Li, “Tri-training: Exploiting unlabeled data using three classifiers”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 11, pp. 1529–1541, 2005.

[61] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila, “Recognition of blurred faces using local phase quantization”, in *Proceedings of International Conference on Pattern Recognition*, 2008, pp. 1–4.

[62] D. Lowe, “Distinctive image features from scale-invariant keypoints”, *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[63] A. Ross and A. Jain, “Information fusion in biometrics”, *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115–2125, 2003.

[64] K. I. Kim and Y. Kwon, “Single-image super-resolution using sparse regression and natural image prior”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1127–1133, 2010.

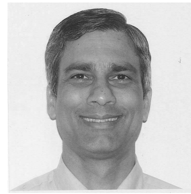
[65] S. Thrun, *Lifelong Learning Algorithms*, Kluwer Academic Publishers, 1998.

[66] J. C. Klontz and A. K. Jain, "A case study on unconstrained facial recognition using the boston marathon bombings suspects", Tech. Rep., Michigan State University, 2013.
[67] "http://www.reuters.com/article/2013/04/23/us-usa- explosions-boston-injuries-idusbre93m0lw20130423".
[68] G. Goswami, R. Bhardwaj, R. Singh, and M. Vatsa, "MDLFace: Memorability augmented deep learning for video face recognition", in *IEEE/IAPR Proceedings of International Joint Conference on Biometrics*, 2014, pp. 1–7.

Himanshu S. Bhatt received the Bachelor in Technology degree in information technology in 2009 from the Jaypee Institute of Information Technology, Noida, India. He has completed his Ph.D. degree from the Indraprastha Institute of Information Technology (IIIT) Delhi, India in 2014 and joined Xerox Research Center India. His research interests include image processing, machine learning and their applications in biometrics. He is a recipient of IBM PhD fellowship 2011-13, best poster awards in IEEE BTAS 2010 and IJCB 2011.

Richa Singh received the M.S. and Ph.D. degrees in computer science in 2005 and 2008, respectively from the West Virginia University, Morgantown, USA. She is currently an Assistant Professor at the Indraprastha Institute of Information Technology (IIIT) Delhi, India. Her research has been funded by the UIDAI and DIT, India. She is a recipient of FAST award by DST, India. Her areas of interest are biometrics, pattern recognition, and machine learning. She has more than 140 publications in refereed journals, book chapters, and conferences. She is also an editorial board member of Information Fusion, Elsevier and EURASIP Journal of Image and Video Processing, Springer. Dr. Singh is a senior member of the IEEE and a member of the Association for Computing Machinery. She is the recipient of several best paper and best poster awards in international conferences.

Mayank Vatsa received the M.S. and Ph.D. degrees in computer science in 2005 and 2008, respectively from the West Virginia University, Morgantown, USA. He is currently an Assistant Professor at the Indraprastha Institute of Information Technology (IIIT) Delhi, India. He has more than 140 publications in refereed journals, book chapters, and conferences. His research has been funded by the UIDAI and DIT. He is the recipient of FAST award by DST, India. His areas of interest are biometrics, image processing, forensics, and information fusion. Dr. Vatsa is a Senior member of the IEEE, Computer Society and Association for Computing Machinery. He is the recipient of several best paper and best poster awards in international conferences. He is also an area editor of IEEE Biometric Compendium, area chair of Information Fusion, Elsevier, and PC Co-Chair of International Conference on Biometrics 2013 and International Joint Conference on Biometrics 2014.



Nalini K. Ratha is a Research Staff Member at IBM Thomas J. Watson Research Center, Yorktown Heights, NY where he is the team leader for the biometrics-based authentication research. He has over 20 years of experience in the industry working in the area of pattern recognition, computer vision and image processing. He received his B. Tech. in Electrical Engineering from Indian Institute of Technology, Kanpur, M.Tech. degree in Computer Science and Engineering also from Indian Institute of Technology, Kanpur and Ph. D. in Computer Science from Michigan State University. He has authored more than 80 research papers in the area of biometrics and has been co-chair of several leading biometrics conferences and served on the editorial boards of IEEE Trans. on PAMI, IEEE Trans. on SMC-B, IEEE Trans. on Image Processing and Pattern Recognition journal. He has co-authored a popular book on biometrics entitled "Guide to Biometrics" and also co-edited two books entitled "Automatic Fingerprint Recognition Systems" and "Advances in Biometrics: Sensors, Algorithms and Systems". He has offered tutorials on biometrics technology at leading IEEE conferences and also teaches courses on biometrics and security. He is Fellow of IEEE, Fellow of IAPR and a senior member for ACM. His research interests include biometrics, pattern recognition and computer vision. He is an adjunct professor at IIIT Delhi, had been an adjunct for several year at Cooper Union and NYU-Poly. During 2011-2012 he was the president of the IEEE Biometrics Council. At IBM, he has received several awards including a Research Division Award, Outstanding Innovation Award and Outstanding Technical Accomplishment Award along with several patent achievement awards.