

# Multi-source deep transfer learning for cross-sensor biometrics

Chetak Kandaswamy<sup>1,2,3</sup> · João C. Monteiro<sup>2</sup> · Luís M. Silva<sup>1,3</sup> · Jaime S. Cardoso<sup>2</sup>

Received: 16 October 2015 / Accepted: 8 April 2016  
© The Natural Computing Applications Forum 2016

**Abstract** Deep transfer learning emerged as a new paradigm in machine learning in which a deep model is trained on a source task and the knowledge acquired is then totally or partially transferred to help in solving a target task. In this paper, we apply the source–target–source methodology, both in its original form and an extended multi-source version, to the problem of cross-sensor biometric recognition. We tested the proposed methodology on the publicly available CSIP image database, achieving state-of-the-art results in a wide variety of cross-sensor scenarios.

**Keywords** Transfer learning · Deep neural networks · Source–target–source · Optimization · Cross-sensor biometrics

## 1 Introduction

Transfer learning is an approach in which the knowledge acquired in solving one task is used to solve a new target task without having to perform the whole training procedure. It is anticipated that new tasks and concepts are learned more quickly and accurately by exploiting past knowledge. Such approach gains an increased interest given that the learning of each task in isolation represents

an expensive process, requiring large amounts of both time and data. With this idea in mind, the focus of transfer learning falls on the leveraging of information from existing sources to train new models with increased efficiency.

Transfer learning has found a series of practical applications in a vast array of research fields, such as text classification from one writing style to another [1], cross-domain video concept detection [2], customers sentiment classification over time [3], biomedical labeling for genes, proteins and biological entities [4] or sensor-based location estimation [5].

Another possible application of transfer learning techniques is the extension of the biometric recognition problem to multiple sensor scenarios. In information technology, *biometrics* refers to the quantitative measure and analysis of human anatomical or behavioral characteristics, such as DNA, fingerprints, eye retinas and irises, voice patterns, facial patterns and hand measurements, for authentication purposes.

With the increasing popularity and availability of mobile devices, capable of performing the whole biometric recognition framework, from data acquisition to final decision, a new obstacle is presented to the development of such systems: the need to adapt to the wide variety of available sensors and the resulting heterogeneity with regards to image quality. The question of whether or not sensors from different manufacturers show a high degree of interoperability allowing, for example, for an individual to be enrolled in a single system and, then, be successfully recognized in a vast variety of alternative devices, is of growing importance in the research field of biometrics. With this formulation in mind, it is trivial to understand how the principles of transfer learning may be adopted for this rising challenge.

---

✉ Chetak Kandaswamy  
chetak.kand@gmail.com

<sup>1</sup> i3S and INEB, Universidade do Porto, Porto, Portugal

<sup>2</sup> INESC TEC and Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

<sup>3</sup> Departamento de Matemática, Universidade de Aveiro, Aveiro, Portugal

With recent studies showing that cross-sensor matching, where the test instances are verified using data enrolled with a different sensor, often lead to reduced performance, we attempt to overcome this challenge by making use of transfer learning principles and, thus, achieve state-of-the-art performance for a large variety of acquisition scenarios. For that purpose, we explore and extend the source–target–source (STS) approach, first proposed in [6], while applying it to the specific challenge of cross-sensor periocular recognition. STS is a recent alternative that has shown increased performance in computer vision object recognition tasks, as well as a significant gain in processing speed.

The remainder of the paper is organized as follows: Sect. 2 details the evolution of both transfer learning techniques and cross-sensor biometrics, while outlining some of the most relevant recent works in the field; Sect. 3 presents the transfer learning techniques that were implemented to tackle the cross-sensor biometric recognition problem, with increased focus on the proposed multi-source STS (MS-STS) approach; Sect. 4 outlines the main results observed for performance assessment carried out in a series of experimental setups; finally, the main conclusions and suggestions for future work can be found in Sect. 5.

## 2 Related work

The issue of learning from few training instances has motivated several works on classification in recent years. The NIPS-95 workshop on *Learning to Learn* unveiled the need for machine learning methods that retain and reuse previously learned knowledge, thus initiating the development of a vast array of *transfer learning* (TL) algorithms.

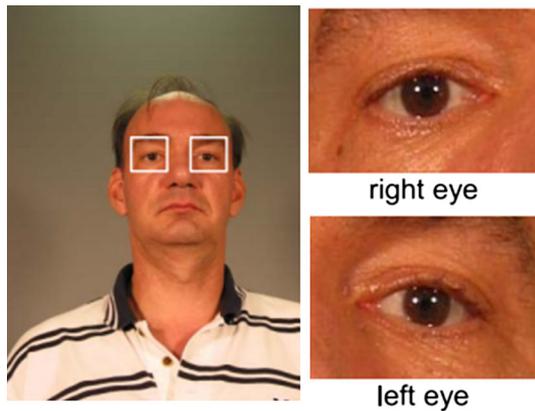
In [7], the author introduced the notion of *multitask learning* where the knowledge from a task is improved by using the information contained in the training instances of other related tasks. This methodology enables the learning of multiple tasks in parallel, by taking advantage of shared representations, thus simultaneously improving generalization for all learnt tasks. The TL algorithm known as *lifelong learning* [8] is based on the assumption that a learner faces multiple learning problems during its lifetime. Thus, when learning the  $n$ -th task, a learner can reuse knowledge gathered in the previous  $n - 1$  tasks to boost the generalization ability. The *cross-domain learning* or *domain adaptation* (see [9–12]) TL algorithm in which a machine learns to perform a task on training instances

drawn from a *source* problem and then performs the *same task* on a *target* problem, whose instances are drawn from a *related distribution*. Domain adaptation expects that the closer the distributions are, the better the features trained on the source problem will perform on the target problem.

Deep transfer learning (DTL) is a framework that combines deep learning models with the transfer learning algorithms (see [13–15]) in which deep learning networks are used for training the source problem, and then, layer-by-layer feature transference is performed to solve a target problem in either a supervised [13] or an unsupervised [15] setting. The advantage of DTL is that it offers a far greater flexibility in extracting high-level features that are transferred from a source to a target problem while, unlike the classical approach, not being affected by expert’s bias [16].

The practical problem of *cross-sensor* biometrics has also been the focus of many works in recent years, highly motivated by the growing variety and availability of mobile sensors. The most commonly found works concern mostly iris recognition. Connaughton et al. [17] performed a comparison between three commercially available iris cameras, with the aim of assessing the interoperability between them and the impact of some state-of-the-art recognition algorithms in both single- and cross-sensor scenarios. The authors arrived at some fulcral conclusions, namely the fact that the relative performance of a given algorithm in a variety of single-sensor scenarios does not relate reliably to the performance of the same algorithm when tested in cross-sensor scenarios. Furthermore, performance observed for all cross-sensor scenarios was consistently worse than their single-sensor counterparts. This idea is also corroborated by the conclusions of Monteiro et al.’s work [18], where two state-of-the-art recognition algorithms fail to cope with the variations introduced to their input data when faced with cross-sensor scenarios.

Another recent work in the field of iris recognition, proposed by Pillai et al. [19], attempted to adapt iris instances acquired with one sensor to the characteristics of a new sensor, in an attempt to mitigate the performance drop commonly observed in cross-sensor scenarios. Both Santos et al. [20] and Jilela and Ross [21] propose methods based on information extracted from the *periocular* region. It is common to describe the periocular region as the region in the immediate vicinity of the eye, as depicted in Fig. 1, whose potential as a biometric trait can be motivated as a representation in between face and iris recognition. The periocular region has been shown to present increased performance when only degraded facial data or low-quality iris images are made available, and even in mobile application scenarios, it does not require rigid capture or



**Fig. 1** Example of periocular regions from both eyes, extracted from a face image [22]

complex imaging systems, thereby making it easy to acquire even by an inexperienced user [18]. While Santos et al. propose a framework based on multiple descriptors to work on periocular data on multiple mobile sensors, Jilela and Ross attempt to match iris and face images from the same individual, acquired with distinct sensors, using periocular traits to help in the recognition process.

With the marked advantages of periocular recognition over its iris and face counterparts becoming more widely accepted and researched, especially when unconstrained acquisition settings are considered, the present work will focus on exploring transfer learning alternatives to periocular recognition in order to attenuate the problems commonly associated with cross-sensor scenarios.

### 3 Deep transfer learning frameworks

Traditionally, the goal of transfer learning (TL) is to transfer the knowledge (learning) obtained from a *source* domain  $S$  to one or more *target* domains  $T$ , in order to efficiently develop an effective hypothesis for a new task, domain or distribution [11]. A variety of divergence levels may exist across different pairs of source domain and target domain data. Brute-forcing knowledge from the source domain into the target domain, irrespective of their divergence, may cause a certain performance degeneration, or, in even worse cases, break the original data consistency in the target domain [23].

With this in mind, a series of specific transfer learning techniques have been designed to meaningfully perform knowledge transfer between the source and target problem’s domains. Such knowledge transfer techniques are based on

the nonnegative matrix trifactorization framework, with the transfer learning phase being performed via dimensionality reduction [23]. The most widely used methods transfer not only features but also parameters to the target domain. The knowledge learnt from a problem, or a set of problems, is then reused to help in solving the new problem(s) more effectively. Inside this broad definition of TL, various methodologies that have been previously explored in the context of deep neural networks are outlined next.

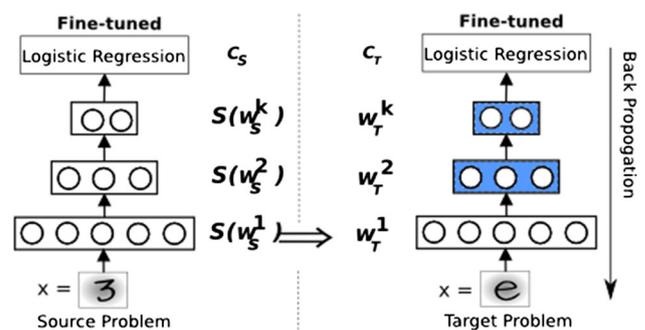
#### 3.1 Established frameworks

Given a dataset  $D = \{(x_i, y_i)\}_{i=1}^N$  drawn from input space  $X$  and a set of labels  $Y$ , a classifier is any function  $f(x) : X \rightarrow Y$  that maps instances  $x_i \in X$  to labels. Our classifier is a deep network with  $K$  layers:  $K - 1$  hidden layers and one output classifier layer (logistic regression). The deep network is thus composed of a set of features  $W = \{w^1, w^2, \dots, w^K\}$  where  $w^k$  is an  $M_{k-1} \times M_k$  weight matrix,  $M_k$  is the number of neurons in the  $k$ th layer and  $M_0$  is the number of inputs of the network. We define the baseline classifier (BL) as the one trained directly to solve the target problem without any type of transference, thus starting from a random initialization of the weight matrices. Classifier performance measures such as prediction accuracy or computation time are measured on a test set  $X_{test}$ .

##### 3.1.1 Transfer learning supervised (TLs)

We assume that the “source” dataset  $D_S$  with input space  $X_S$  and a set of labels  $Y_S$  is drawn from a distribution  $P_S(X)$ , and the “target” dataset  $D_T$  with input space  $X_T$  and a set of labels  $Y_T$  is drawn from a distribution  $P_T(X)$ . Such  $P_S(X)$  and  $P_T(X)$  may be either equal or different.

In TLs, we transfer fine-tuned source problem features to the target network. Figure 2 depicts such transference approach between the first layers of both source and target



**Fig. 2** TLs for first-layer feature transference

networks. As a first step, we randomly initialize each layer of the network using a uniform distribution [10]:

$$\mathbf{w}_S^k = \mathcal{U}\left[\frac{-\sqrt{6}}{\sqrt{M_{k-1} + M_k}}, \frac{\sqrt{6}}{\sqrt{M_{k-1} + M_k}}\right] \quad (1)$$

Next, we apply greedy layerwise pre-training  $U(\cdot)$  until the  $K - 1^{\text{th}}$  hidden layer using unlabeled samples from the source data as represented below:

$$U(\mathbf{w}_S^1, \dots, \mathbf{w}_S^{K-1}) \quad (2)$$

After pre-training, we fine-tune  $S(\cdot)$  these unsupervised features with the source problem labeled instances as given in Eq. (3):

$$S(U(\mathbf{w}_S^1, \dots, \mathbf{w}_S^{K-1}), \mathbf{w}_S^K) \quad (3)$$

Feature transference is then applied by mapping the fine-tuned source features to the target network. In the case due to a different number of classes between the problems, the logistic regression layer from the source problem cannot be reused. A new randomly initialized logistic layer is then used for the target problem. Finally, we fine-tune this entire deep network as a multilayer perceptron using back propagation.

### 3.1.2 Source–target–source (STS)

In our previous work, we proposed a source–target–source (STS) approach [6]. As previously discussed, the main idea of transfer learning is that the knowledge (features) learnt in a source problem may provide a good initialization for the learning task in a target problem, serving as an improvement over random initialization of the learning parameters for the target domain. In [6], we proposed an iterative learning between both domains. The intuition is that, like in typical metaheuristics in optimization (i.e., tabu search and simulated annealing), moving the learning process from one domain to the other will ‘shake’ the current local optimal solution, allowing us to keep exploring the space of solutions and, ideally, reaching a better solution in the process. The solutions reached in each iteration are kept, and only the global best solution achieved is considered. Further technical details and preliminary results using this approach may be found in [6].

## 3.2 Proposed multi-source source–target–source framework

In the present work, we extend the STS transfer learning methodology presented in the last section, by reusing

knowledge learnt by a model trained on multiple sources. As previously mentioned in Sect. 2, two different approaches have attempted to account for the reuse of multiple sources for TL: (a) lifelong learning [8] and (b) multitask learning [7]. Both approaches are based on specific TL scenarios and assume that the data and the tasks are related.

Both previously mentioned approaches suffer from serious limitations. For example, if two tasks are negatively correlated, the learning process will cause degradation of the generalization performance of both tasks. In order to avoid such issues, a strong task selection is required in order to constrain the application of such methods to a limited set of positively correlated problems. The multi-source source–target–source (MS-STs) approach improves generalization performance over multiple problems, with no need for prior task selection.

The MS-STs approach is briefly explained in the pseudocode presented in Algorithm 1. Consider a pool, *Pool*, containing multiple datasets from a similar problem or particular application,  $Pool = \{D_A, D_B, \dots, D_Z\}$  drawn from  $P_A(X), P_B(X), \dots, P_Z(X)$ , where  $Z$  is number of datasets. We select a deep neural network architecture and initialize the weights of each layer of the network using a uniform distribution under the limits defined in step 1 of Algorithm 1. Initializing the weights through this method narrows down the gradient search parameter, thus speeding up the training of the network [10]. Heuristically, we set the maximum number of cycles to 10. This value may suffer variations according to the nature of the problem. In step 2, we select a desired target dataset  $D_{dT}$  from the *Pool* for which we intend to have the best overall accuracy. During each of  $H$  cycles, a target dataset  $D_T$  is selected among the pool of datasets, for which we apply the deep transfer learning approach discussed in Sect. 3.1 as well as the source–target–source approach as discussed in [6] by selecting layers to transfer and/or to lock in the new network.

The new network is trained and tested as a regular deep network for the selected  $D_T$ . A list of best accuracies for each cycle is maintained for every dataset in the pool. If the current cycle test accuracy for the desired target dataset  $D_{dT}$  is greater than the average of the top test accuracies, for the desired target dataset  $D_{dT}$ , we break the cycle and store the final weights of the network. By training and testing serially on multiple datasets, we improve the domain generalization property of the approach, while by focusing on the desired target dataset, we aim to improve the domain specialization property.

**Algorithm 1** Pseudocode for MS-STS

1: **Randomly initialize** the weights of every layer of the network using uniform distribution:

$$\mathbf{w}^k \sim \mathcal{U} \left[ \frac{-\sqrt{6}}{\sqrt{M_{k-1} + M_k}}, \frac{\sqrt{6}}{\sqrt{M_{k-1} + M_k}} \right]$$

2: select a desired target dataset  $D_{dT}$  from the  $Pool = \{D_A, D_B, \dots, D_Z\}$ .  
 3: **for**  $h$  in  $H$  cycles **do**  
 4:   **for**  $p$  in  $Pool$  **do**  
 5:      $D_T = p$  {set  $p$  as target dataset}  
    {transfer features to new network as discussed in the transfer learning approach [6] and select which of the layers to transfer and which of the layers to lock, out of  $K$  layers}  
 6:     **for**  $k$  in  $K$  **do**  
 7:        $\mathbf{w}_S^k \Rightarrow \mathbf{w}_T^k$  {transfer selected layers}  
 8:        $\mathbf{w}_S^k \Downarrow \mathbf{w}_T^k$  {lock selected layers}  
 9:     **end for**  
 10:     train and test the new network with  $D_T$   
 11:     update the test accuracy list  
 12:     **if** accuracy of  $D_{dT} > \text{avg}(\text{top 3 best test accuracy in } D_{dT} \text{ accuracy list})$  **then**  
 13:       **BREAK**  
 14:     **end if**{Continue MS-STs step 3 till global optima is reached for the  $D_{dT}$ .}  
 15:   **end for**  
 16: **end for**

It is necessary to clarify exactly what is meant by MS-STs, as all of the above-mentioned methods also take advantage of multiple sources to train the network. MS-STs proposes to extend the established STs methodology with multiple sources instead of a single source. The intuition is that by providing multiple initialization points for exploring the space, we may reach a solution closer to the optimal. This search may result in an increase of the computational cost, but if a trade-off can be found between classification performance, system robustness and computational complexity, this might not end up as a significant limitation.

## 4 Cross-sensor recognition results and discussion

In the present work, we explore the approaches outlined in the previous section as an alternative to tackle the cross-sensor biometric recognition problem. This problem can be understood as the problem of successfully performing biometric recognition on a specific image acquisition device without the need of performing a new enrollment phase for the new device specifically. This interpretation can be easily extrapolated to the domain of the aforementioned approaches if both devices are understood as the target (where recognition is to be performed) and the

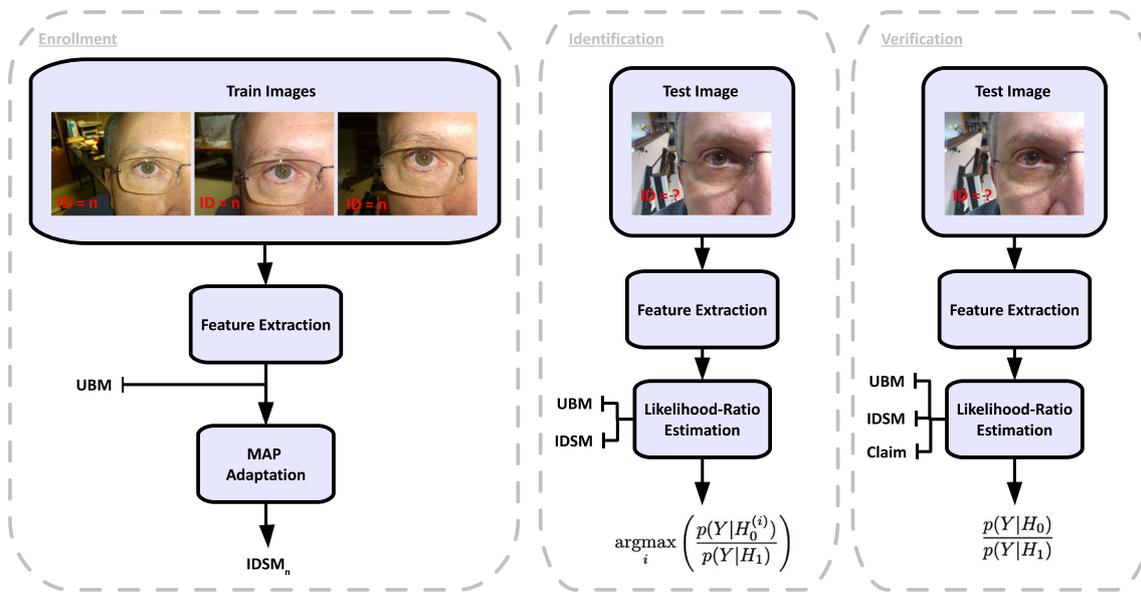
source (where enrollment was carried out). In the following sections, we outline the experimental setups designed to assess the performance of the proposed methodologies in the specific practical problem of cross-sensor periocular recognition.

We start by detailing a baseline algorithm, first proposed by Monteiro et al. [24], that has presented state-of-the-art performance for multiple single-sensor scenarios, as well as a commonly used feature representation technique—GMM supervectors—which will be explored for stacked denoising autoencoders (SDA) approaches. We then present the experimental setup under which each of the tested methodologies was assessed as well as the performance metrics chosen for such process. Finally, we present the most significant results as well as a detailed discussion concerning the relative performance of each method for each of the proposed challenges.

### 4.1 Tested methodologies

#### 4.1.1 GMM-universal background model (GMM-UBM)

The GMM-UBM algorithm for periocular recognition, first proposed by Monteiro et al. [24], is schematically represented in Fig. 3. During the enrollment, a set of  $I$  models describing the unique statistical distribution of biometric features for each individual  $i \in \{1, \dots, I\}$  is trained by



**Fig. 3** Schematic representation of the GMM-UBM periocular recognition algorithm proposed by Monteiro et al. [24]

maximum a posteriori (MAP) adaptation of an universal background model (UBM). The UBM is a representation of the variability that the chosen biometric trait presents in the universe of all individuals. MAP adaptation works as a specialization of the UBM based on each individual's biometric data. The idea of MAP adaptation of the UBM was first proposed by Reynolds [25], for speaker verification. The tuning of the UBM parameters in a maximum a posteriori sense, using individual-specific biometric data, provides a tight coupling between the individual models and the UBM, resulting in better performance and faster scoring than uncoupled methods, as well as a robust and precise parameter estimation, even when only a small amount of data is available.

The recognition stage is carried out through the projection of the features extracted from an unknown sample onto both the UBM and the individual-specific models (IDSM) of interest. A likelihood ratio between both projections outputs the final recognition score. Depending on the functioning mode of the system—verification or identification—decision is carried out by thresholding or maximum likelihood ratio, respectively. The use of a likelihood ratio score with an universal reference works as a normalization step, mapping the likelihood values in accordance with their global projection. Without such step, finding a global optimal value for the decision threshold would be a far more complex process.

Gaussian mixture models (GMM) were chosen to model both the UBM and the individual-specific models (IDSM). From the most common interpretations, GMMs are seen as capable of representing broad hidden classes, reflective of the unique structural arrangements observed in the analyzed

biometric. The original work was proposed using SIFT key-point descriptors as the only features, but a more recent version [26] proposed a score-level fusion of multiple descriptors (SIFT, HOG, LBP and GIST), resulting in improved performance.

The original work was designed with single-sensor recognition in mind, i.e., the source and target data are the same. In the present work, we also assess the performance in cross-sensor scenarios, where training of models and classification are carried out on distinct data sources. Some preliminary results for such setup have already been reported in a follow-up work by the original authors [26]. The present work will more thoroughly analyze and discuss such results, as well as presenting a comparative analysis with alternative approaches.

#### 4.1.2 GMM supervectors (SV-SDA)

With the previous methodology, recognition was carried out through a likelihood ratio between a target IDSM and the UBM. Recently, a significant amount of works have explored the use of an alternative GMM representation—GMM supervectors—as the input for classification algorithms, with some promising results being reported in the literature [27]. Supervector notation consists on concatenating in a single vector all the parameters describing a GMM (weights, means and covariance matrices). For example, the mean values of the UBM can be concatenated to form a single mean supervector,  $mu$ , given by  $mu = [\mu_{T1}, \mu_{T2}, \dots, \mu_{Tj}]$ , where  $j$  is the total number of mixtures in the UBM [28]. A similar representation can be extracted for the IDSM parameters. In the SV-SDA

approach, we describe each training image  $t$  belonging to subject  $i$ ,  $Im_{t,i}$ , by its supervector representation, obtained by MAP adaptation of the UBM parameters using the feature data extracted solely from  $Im_{t,i}$ . SIFT is used for feature description.

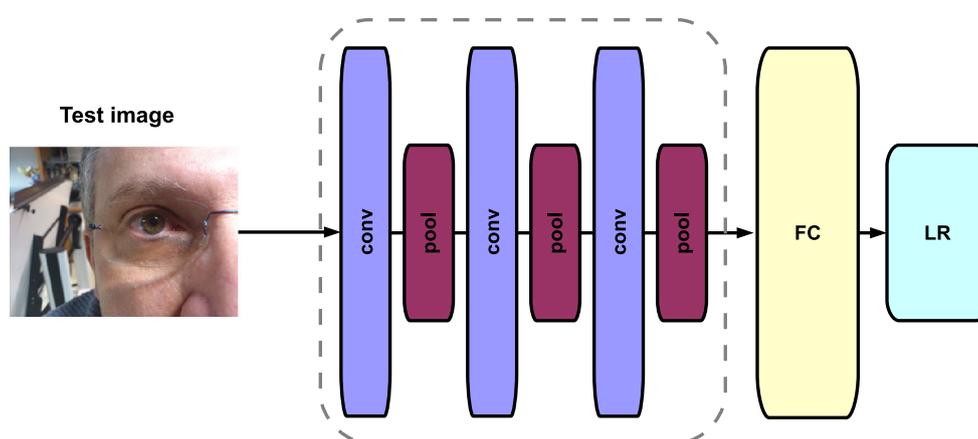
We then perform training, validation and classification using stacked denoising autoencoders (SDA) for both TLs and STS approaches. SDAs are multiple layer networks where each individual layer is trained as a denoising autoencoder (dA). The training of a SDA model is comprised of two stages: an *unsupervised pre-training* (PT) stage followed by a *supervised fine-tuning* (FT) stage. During PT the network is generated by stacking multiple dA one on top of each other, thus learning *unsupervised features*, represented as a vector  $U(w)$ , of optimal weights and biases. Then a logistic regression layer is added on top and the whole network is fine-tuned, in a supervised way, learning a set of  $K$  *supervised features*  $w = (w^1, \dots, w^K)$ , where  $K$  is the number of layers [29].

#### 4.1.3 CNN

A methodology based on Convolutional Neural Networks (CNN) was also carried out, for TLs, STS and MS-STS approaches, using raw pixel intensity values. CNN take advantage of the fact that the input consists of images to constrain the neural network (NN) architecture in a more sensible way. In particular, unlike regular NNs, the layers of a CNN have neurons arranged in three dimensions: width, height and depth. (Note that the word depth here refers to the third dimension of an activation volume, not to the depth of a full Neural Network, which can refer to the total number of layers in a network.) The CNN architecture reduces the full image into a single vector of class scores, arranged along the depth dimension.

CNNs exploit *spatially local correlation* by enforcing a local connectivity pattern between neurons of adjacent layers. This architecture thus ensures that the learnt neurons produce the strongest response to a spatially local input pattern. Also, *sharing weights* increases the invariance of learnt filters by replicating each of them across the entire visual field. These replicated filters share the same parameterization (weight vector and bias) and form a feature map. Replicating units in this way allows for features to be detected regardless of their position in the visual field. Additionally, weight sharing increases learning efficiency by greatly reducing the number of free parameters being learnt. The constraints on the model enable CNNs to achieve better generalization on image-based problems.

We used three main types of layers to build CNN architectures: convolutional layers (conv), pooling layers (pool) and fully connected layers (FC). We will stack these layers to form a full CNN architecture with a logistic regression classifier (LR). The architecture of our 5-layer CNN model is composed of [Conv – Pool]  $\times$  3 – FC – LR. We first crop the image to a fixed dimension of  $200 \times 120$  and then convert the image to gray scale to work as the input. Convolution with 30 different first-layer filters, each of size  $12 \times 12$ , using a stride of 1 in both  $x$  and  $y$ , is then carried out. The resulting feature maps are then pooled in (maximum within  $2 \times 2$  regions, using stride 1) to give 30 different  $94 \times 54$  element feature maps. Similar operations are repeated with 60 and 90 different layer filters in second and third layers, respectively. The fourth layer is fully connected, taking features from the top convolutional layer as input in vector form. The final layer is a  $c$ -way logistic regression classifier,  $c$  being the number of classes. All filters and feature maps are square in shape. A schematic representation of the developed structure may be observed in Fig. 4.



**Fig. 4** Schematic representation of the developed CNN architecture

## 4.2 Tested dataset

The methodologies outlined in the previous sections were assessed on the cross-sensor iris and periocular (CSIP) dataset. The CSIP database, created for the assessment of the algorithm proposed by Santos et al. [20], is a recent and publicly available dataset, designed with the main goal of gathering periocular images from a representative group of participants, acquired using a variety of mobile sensors under a set of variable acquisition conditions. Given the heterogeneity of the camera sensors and lens setups of consumer mobile devices, ten different setups were used during the dataset acquisition stage: four different devices, some of which had both frontal and rear cameras, and LED flash. This variety of sensors confers a strong appeal to the CSIP database regarding its potential use for the assessment of algorithms under a highly heterogeneous set of conditions. A summary of the details concerning each of such setups may be observed in Table 1, while a visual example of an image for each subset of the same individual is depicted in Fig. 5. Each participant was imaged using all of the presented setups.

To simulate the variable noise associated with on-the-go recognition, participants were not imaged at a single

location, but instead, they were enrolled at multiple sites with artificial, natural, and mixed illumination conditions. In total, 50 participants were enrolled, all Caucasian and mostly males (82 %), with ages ranging between 21 and 62 years ( $\mu \pm \sigma = 31.18 \pm 3.15$  years). For each periocular image acquired by the mobile devices, a binary iris segmentation mask was also produced. The masks were obtained automatically using the state-of-the-art iris segmentation approach proposed by Tan et al. [30], which is particularly suitable for uncontrolled acquisition conditions, as demonstrated by its first-place ranking at Noisy Iris Challenge Evaluation—Part 1 (NICE.I) [31].

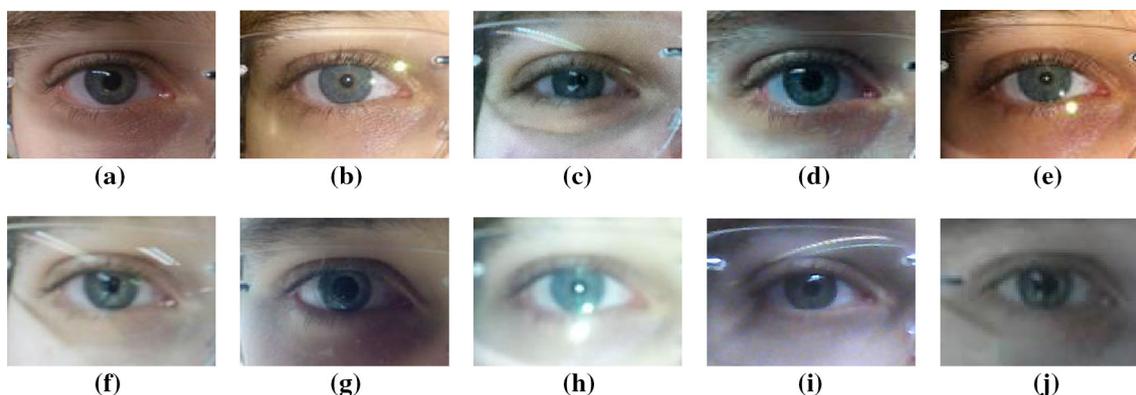
## 4.3 Experimental setup

### 4.3.1 Image preprocessing

Images from the CSIP database were converted to gray scale and resized so as to present a fixed number of pixels, necessary for the implementation of all the approaches based on the CNN methodology. Resizing was carried out in such a way that geometrical proportions were kept from the original images.

**Table 1** Technical details concerning the acquisitions setups used for each subset of the CSIP database

Setup ID	AR0	AR1	BF0	BR0	BR1	CF0	CR0	CR1	DF0	DR0
Device	A		B			C		D		
Manufacturer	Sony Ericsson		Apple			ThL		Huawei		
Model	Xperia Arc S		iPhone 4			W200		U8510		
O.S.	Android 2.3.4		iOS 7.1			Android 4.2.1		Android 4.3.3		
Camera	Rear		Frontal	Rear		Frontal	Rear		Frontal	Rear
Resolution	3264	2448	640 480	2592	1936	2592 1920	3264	2448	640 480	2048 1536
Flash	No	Yes	No	No	Yes	No	No	Yes	No	No



**Fig. 5** Examples of images from each subset of the CSIP database. From a–j, respectively: AR0, AR1, BF0, BR0, BR1, CF0, CR0, CR1, DF0 and DR0

### 4.3.2 Data partitioning

In order to achieve a fair and meaningful comparison between the tested methodologies, a common experimental setup was designed. The set of all images of the CSIP dataset was divided as follows: 50 % of the images per individual and per subset were kept for model training, 25 % were chosen for validation of the trained models, and the remaining 25 % were used to assess performance. Train, validation and test subsets were randomly selected, and all experiments were cross-validated ten times.

### 4.3.3 Evaluation metrics

Performance was evaluated only for identification problems, where, given a biometric sample from an unknown source, the  $p$  most probable identities are assessed. For such problems, the most commonly found performance metric is the *rank-1 recognition rate*, which refers to the ratio of correctly assessed identities, when  $p = 1$ .

## 4.4 Cross-sensor recognition performance

The main results obtained for the experimental setups detailed in the last section are summarized in Tables 2, 3, 4, 5, 6, 7 and 8. Discussion of these results will be carried out, from this point onwards, starting with the BL and TLs approaches, followed by an analysis on how the STS strategy may improve performance in cross-sensor scenarios and, finally, on the effect that multiple sources of information may present.

### 4.4.1 Baseline and transfer learning

The baseline results for each tested methodology (GMM-UBM, SV-SDA and CNN) are presented in the diagonal

**Table 2** Rank-1 recognition rates, in %, observed for the GMM-UBM algorithm for all possible cross-sensor scenarios in the CSIP database

Source	Target								
	AR0	AR1	BF0	BR0	BR1	CF0	CR0	CR1	DR0
AR0	94.4	57.8	34.1	76.8	45.9	41.8	52.8	40.7	65.6
AR1	64.6	97.1	26.5	47.9	83.6	25.5	35.4	66.8	33.7
BF0	33.1	23.1	78.2	21.1	19.4	30.8	24.0	19.8	16.0
BR0	67.4	39.6	19.7	92.4	54.2	36.5	42.3	34.3	67.7
BR1	31.8	62.3	12.0	48.1	95.5	28.3	25.6	52.5	35.3
CF0	36.4	29.1	34.7	36.5	30.8	89.8	55.8	39.8	46.3
CR0	59.5	30.2	24.4	58.1	36.4	59.3	80.3	45.7	71.9
CR1	42.6	64.9	21.2	47.8	70.8	47.5	50.5	90.0	49.0
DR0	41.3	18.0	17.4	53.0	23.1	30.3	39.8	24.8	88.7

**Table 3** Rank-1 recognition rates, in %, observed for the SV-SDA methodology and TLs approach for all possible cross-sensor scenarios in the CSIP database

Source	Target								
	AR0	AR1	BF0	BR0	BR1	CF0	CR0	CR1	DR0
AR0	38.0	<b>81.8</b>	39.3	<b>33.4</b>	79.3	14.1	<b>27.7</b>	52.5	20.0
AR1	35.4	76.7	22.7	24.6	77.3	15.9	20.9	<b>61.6</b>	17.6
BF0	<b>43.8</b>	81.4	41.4	27.3	80.1	<b>22.3</b>	17.8	59.7	21.5
BR0	35.2	79.0	35.8	25.6	77.3	14.6	21.8	45.9	22.1
BR1	32.4	78.4	37.7	25.1	82.1	22.1	18.0	55.6	23.0
CF0	32.4	79.8	36.4	24.1	75.3	12.1	16.3	50.6	<b>24.6</b>
CR0	41.5	79.0	36.6	24.8	81.6	16.1	23.5	51.7	23.9
CR1	28.7	79.0	36.6	22.1	78.0	16.4	18.2	57.9	22.4
DR0	28.9	80.4	39.0	29.0	72.2	15.0	22.0	52.5	19.7

**Table 4** Rank-1 recognition rates, in %, observed for the CNN methodology and TLs approach for all possible cross-sensor scenarios in the CSIP database

Source	Target								
	AR0	AR1	BF0	BR0	BR1	CF0	CR0	CR1	DR0
AR0	61.5	84.2	<b>54.0</b>	<b>69.0</b>	91.0	<b>64.4</b>	63.9	92.0	<b>65.3</b>
AR1	60.8	82.1	52.0	66.0	88.5	58.1	63.9	92.0	56.7
BF0	<b>65.9</b>	84.5	50.0	68.5	89.5	62.6	67.4	92.3	<b>65.3</b>
BR0	63.0	82.6	52.4	63.5	88.0	61.8	67.0	92.0	56.7
BR1	63.0	85.0	52.4	60.5	85.5	60.4	64.3	90.3	57.3
CF0	64.4	85.0	50.8	63.5	<b>91.5</b>	54.8	66.1	93.0	58.0
CR0	62.2	<b>86.8</b>	51.2	66.5	90.5	60.7	67.9	93.3	62.7
CR1	61.4	82.9	50.4	66.5	89.5	57.8	65.2	88.0	57.3
DR0	60.0	83.2	53.2	62.0	86.5	55.2	<b>68.7</b>	90.0	53.3

**Table 5** Rank-1 recognition rates, in %, observed for the SV-SDA methodology and a single cycle of the STS approach for all possible cross-sensor scenarios in the CSIP database

Source	Target								
	AR0	AR1	BF0	BR0	BR1	CF0	CR0	CR1	DR0
AR0	38.0	89.1	<b>54.4</b>	<b>31.1</b>	90.6	16.0	<b>27.7</b>	59.6	27.3
AR1	40.8	76.7	42.1	29.2	92.2	<b>17.0</b>	25.4	<b>68.9</b>	<b>32.0</b>
BF0	42.3	<b>91.6</b>	41.4	30.3	90.8	16.5	22.8	63.9	29.3
BR0	<b>43.1</b>	88.7	48.2	25.5	90.6	14.0	27.9	62.5	25.7
BR1	41.8	90.0	50.3	28.4	82.1	16.8	23.5	64.1	26.0
CF0	39.2	88.0	47.4	30.0	89.4	12.0	25.4	63.9	25.0
CR0	42.6	88.0	50.9	30.0	<b>93.1</b>	16.5	23.5	66.1	26.7
CR1	37.4	89.6	50.0	28.9	90.0	18.3	25.1	57.8	26.0
DR0	41.8	88.4	53.2	30.2	90.0	16.8	25.6	61.6	19.7

**Table 6** Rank-1 recognition rates, in %, observed for the SV-SDA methodology and the STS approach for all possible cross-sensor scenarios in the CSIP database

Source	Target									
	AR0	AR1	BF0	BR0	BR1	CF0	CR0	CR1	DR0	
AR0	38.0	91.3	<b>54.4</b>	<b>38.7</b>	90.6	19.0	<b>30.9</b>	59.6	<b>32.7</b>	
AR1	40.8	76.7	42.1	29.7	93.9	<b>19.8</b>	25.4	<b>71.6</b>	32.0	
BF0	<b>48.5</b>	<b>91.6</b>	41.4	32.2	90.8	<b>19.8</b>	23.5	65.9	31.7	
BR0	43.1	88.7	48.2	25.5	90.6	17.3	27.9	62.5	31.7	
BR1	43.1	90.0	50.3	29.5	82.1	19.0	26.1	64.3	30.3	
CF0	40.0	88.0	47.4	33.8	89.4	12.0	26.1	63.9	28.7	
CR0	44.4	88.0	50.9	32.2	<b>93.1</b>	<b>19.8</b>	23.5	66.1	29.3	
CR1	37.4	90.0	50.6	28.9	92.2	18.3	25.1	57.8	30.0	
DR0	43.3	88.4	53.2	31.9	90.0	16.8	29.1	61.6	19.7	

**Table 7** Rank-1 recognition rates, in %, observed for the CNN methodology and a single cycle of the STS approach for all possible cross-sensor scenarios in the CSIP database

Source	Target									
	AR0	AR1	BF0	BR0	BR1	CF0	CR0	CR1	DR0	
AR0	61.5	84.5	<b>55.2</b>	<b>70.0</b>	87.5	63.0	68.3	92.3	59.3	
AR1	64.4	82.1	50.4	66.5	88.5	59.3	68.7	91.0	60.0	
BF0	63.7	84.0	50.0	67.5	88.0	<b>64.8</b>	66.1	92.0	57.3	
BR0	62.2	84.0	54.0	63.5	<b>89.5</b>	64.1	67.0	92.0	61.3	
BR1	64.4	84.2	52.4	65.0	85.5	60.4	<b>70.9</b>	<b>93.3</b>	54.7	
CF0	61.5	84.5	53.6	64.0	88.0	54.8	69.6	90.7	62.0	
CR0	63.3	84.5	52.0	67.5	88.5	63.0	67.9	90.0	62.0	
CR1	<b>65.2</b>	85.3	52.8	68.5	86.5	64.1	68.7	88.0	<b>64.0</b>	
DR0	63.7	<b>85.5</b>	50.4	65.5	89.0	62.6	64.8	91.7	53.3	

**Table 8** Rank-1 recognition rates, in %, observed for the CNN methodology and STS approach for all possible cross-sensor scenarios in the CSIP database

Source	Target									
	AR0	AR1	BF0	BR0	BR1	CF0	CR0	CR1	DR0	
AR0	61.5	88.7	<b>58.4</b>	72.5	91.0	70.0	67.8	92.3	70.0	
AR1	71.9	82.1	55.2	69.5	<b>92.5</b>	66.7	70.9	92.0	66.7	
BF0	<b>73.3</b>	90.3	50.0	71.5	91.5	<b>70.7</b>	69.1	92.0	71.3	
BR0	72.6	89.0	57.6	63.5	90.5	69.6	70.4	92.7	<b>72.0</b>	
BR1	71.5	90.8	54.8	69.0	85.5	64.4	<b>72.2</b>	<b>93.3</b>	64.7	
CF0	72.6	90.5	57.2	72.5	91.5	54.8	70.0	92.3	69.3	
CR0	72.2	90.8	<b>58.4</b>	<b>73.5</b>	91.0	68.9	67.9	92.3	69.3	
CR1	70.7	<b>92.1</b>	56.4	72.5	90.5	65.9	70.4	88.0	68.0	
DR0	71.1	89.7	57.2	70.0	91.0	67.0	68.7	92.3	53.3	

values of Tables 2, 3 and 4, respectively. By the sole analysis of these results some conclusions may already be drawn. First of all, it is easily discernible how the GMM-

UBM methodology, specifically designed to solve the single-sensor periocular recognition problem, outperforms both alternatives in such conditions. Even for the CSIP subsets that, in theory, offer the least challenging conditions (*AR1*, *BR1* and *CR1*), the performance drop observed is non-negligible. Taking *AR1* as a specific example a relative performance drop of 26.6 and 18.3 % is observed in the SV-SDA and CNN methodologies, respectively. This effect is, however, reversed when cross-sensor scenarios are taken into consideration.

If we consider a single target dataset, it is readily observable that the variance in performance is a lot less pronounced for the CNN and SV-SDA methodologies than for GMM-UBM. Furthermore, it also notorious how the significantly better single-sensor scenario results of the GMM-UBM are severely degraded when a more complex challenge is presented to the algorithm. A trivial conclusion can be taken from such observations: Even though the GMM-UBM presents the best baseline results, as expected from an algorithm tailored for that specific challenge, the application of transfer learning to both CNN and SV-SDA methodologies results in a considerably lower variance in the performance values observed for a single target dataset, regardless of the chosen source. A valid deduction, following such conclusions, is that improving the baseline performance of such methodologies will also result in an increased cross-sensor performance. Given that the challenge of cross-sensor biometric recognition is mostly concerned with the performance loss observed in such cases, the global behavior of the tested methodologies seems to, at least, motivate further research seeking to improve the baseline performance.

In the next section, we explore the effect that the source–target–source approach presents over the simpler TLs alternative.

#### 4.4.2 Source–target–source

As detailed in Sect. 3, we propose a cyclic source–target–source (STS) approach for classification using the CNN and SV-SDA methodologies. Tables 5 and 6 present the STS results observed for the SV-SDA methodology for a single cycle ( $STS_1$ ) and for a total of ten cycles, respectively. Analogous results for the CNN methodology may be observed in Tables 7 and 8. For a simpler analysis the baseline results are kept on the diagonal of each table, as in the last section.

The first observation to be taken from the analysis of the aforementioned tables is how even a single cycle of STS can significantly improve some of the baseline results. The *CF0* baseline for example, presents a relative improvement of 18.2 % for the CNN methodology, and most of the observed results already exceed those observed for the

simpler TLs approach. This improvement is even more discernible when multiple STS cycles are carried out. The results presented in Tables 6 and 8 depict this behavior. Here, and taking the same *CF0* baseline result as referred before, the performance, comparing to the baseline, is increased to 29.0 %. It is interesting to note how the stability observed in the last section, when a single target dataset is considered, is also observed in this approach, with the addition of significantly increased performance. The same conclusion can, thus, be achieved: If a stronger baseline performance is achieved, STS approaches to classification seem to present the capability of both improving the baseline performance and guaranteeing the maintenance of such performance when different acquisition scenarios are considered.

Another consideration to be taken from the analysis of these results is how significantly worse the results using supervector-based SDA classification are when compared with their CNN counterparts. This observation can also be made from the analysis of Tables 2, 3 and 4 from last section. The most obvious explanation concerns the fact that the supervector representation based on the GMM modeling of SIFT keypoint descriptors might not present enough discriminative information for accurate SDA classification, except in some specific cases. For example, the datasets composed by higher quality images (*AR1* and *BR1*) present considerably better performance, even surpassing the performance of their CNN counterparts. These results show that some discriminative power exists, even though it seems severely compromised when the quality of the input images decreases. Regardless of that, the STS behaviors described above still remain relevant for the SDA methodology, and may earn some further research regarding the use of more adequate feature representation techniques.

As a final approach we also explored the effect of using information from multiple sources in order to improve the performance of the cyclic STS algorithm. The main results and observations regarding this approach will be outlined in the next section, in an attempt to summarize all the results and observations obtained in the present work.

#### 4.4.3 Multiple source STS

Figure 6 summarizes both the results obtained for the STS algorithm using multiple sources (MS-STS) and the most relevant results presented in the last sections. The main goal of MS-STS is to achieve a high degree of domain generalization, in order to allow the trained classifiers to perform well for the widest possible variety of scenarios. For the multiple source examples, we chose the flash subsets (*AR1*, *BR1* and *CR1*) as the sources and all other no-flash datasets as the targets. This choice can be

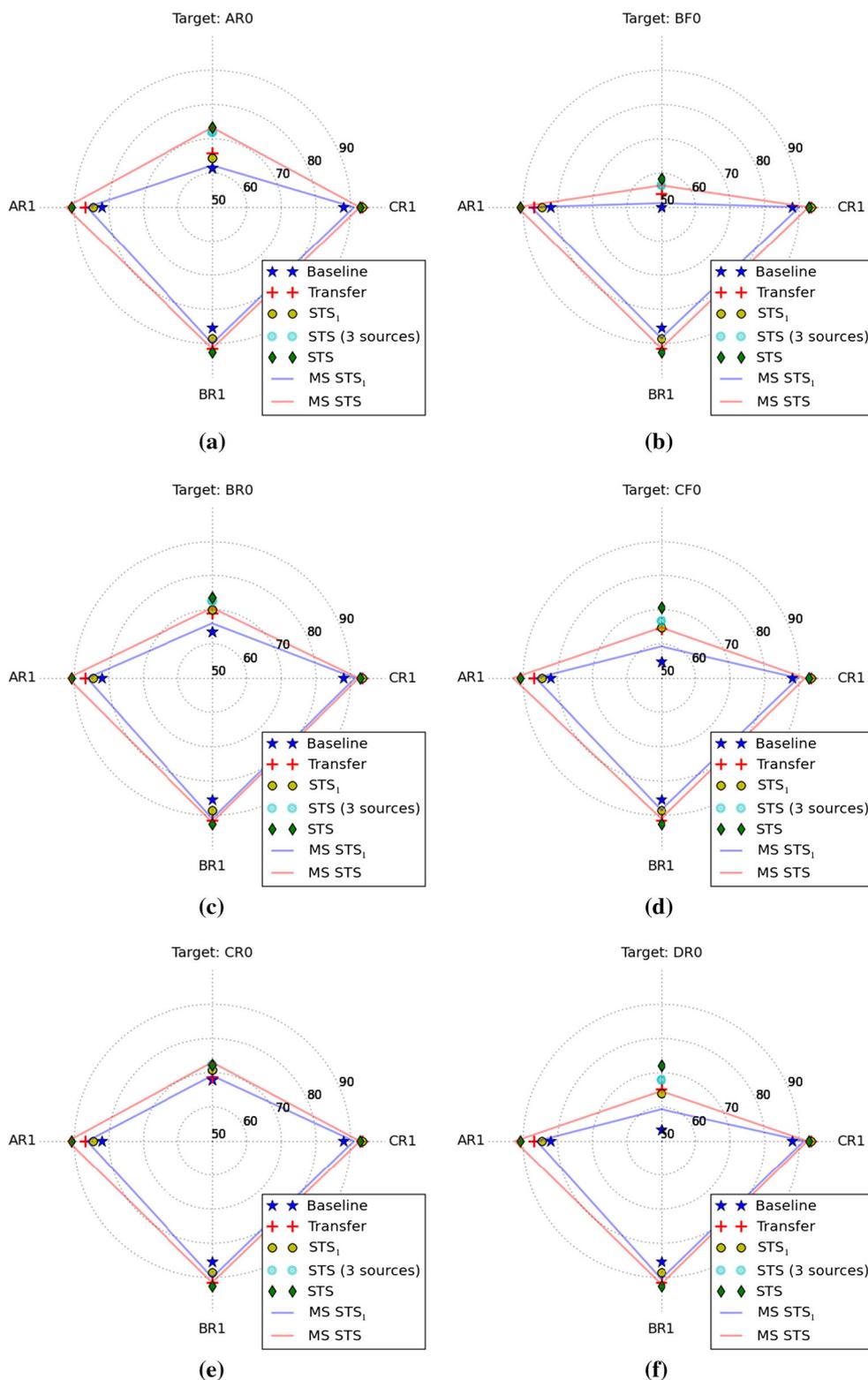
motivated by the fact that the three flash datasets consistently presented the best absolute performance among all the experiments that we carried out. Such observation seems to indicate that the intrinsic discriminative power of such datasets might be higher than the remaining alternatives, thus conferring them, at least in theory, a marked advantage as choice for source datasets. We also chose to work only with the CNN methodology, as the vast majority of the results observed in the last section seemed to point toward its better fit for the problem at hand.

So as to better visualize and understand the effect of the MS-STS approach over the approaches presented in the last sections, we decided to present the results in the radial plot representation that can be observed in the six images from Fig. 6. For each image, a series of features can be observed:

- *Source and Target Datasets*: Each of the axis of the radial plot represents the rank-1 recognition rate (in %) for the chosen target (positive vertical axis) as well all the three source datasets.
- *BL, TLs, STS<sub>1</sub> and STS*: The BL label represents the baseline performance as already presented in the diagonal values of Tables 2, 3 and 4. TLs and STS<sub>1</sub>/STS, on the other hand, represent the best results for each of the 4 depicted datasets (3 sources and 1 target), for each of their individual TLs and STS<sub>1</sub>/STS experiments (bold values in Tables 3, 4, 5, 6, 7, 8). STS<sub>1</sub> represents the performance after a single cycle of the STS approach, whereas STS refers to the best performance observed after ten cycles.
- *STS (3 sources)*: This label depicts the best performance observed for the target dataset, for its individual STS<sub>1</sub>/STS experiment, considering only *AR1*, *BR1* and *CR1* as a possible source. We chose to include this label in order to achieve the fairest possible comparison between the MS-STS performance and the optimal single-source experiment.
- *MS-STS<sub>1</sub> and MS-STS*: The two polygons, in blue and red, respectively, represent the first cycle and optimal performances, after ten repetitions of the whole multi-source cyclic process, for each of the source and target datasets.

By the analysis of the plots, independent of the chosen target dataset, a few interesting conclusions can be drawn. First, there seems to be no significant performance change, regarding the target dataset, between the MS-STS (after ten cycles) approach and the analogous results for STS using only the best single source from the [*AR1*, *BR1*, *CR1*] set of sources. What the MS-STS offers is a way of achieving this optimal performance without the need of an empirical choice of the best source subset, thus conferring a more robust nature to the whole process. This is also the main

**Fig. 6** Graphical representation of the MS-STS rank-1 recognition rates obtained for all the no-flash subsets of the CSIP database using all the flash datasets as sources, plotted against the respective BL, TLs and STS results



advantage of the multiple source approach when compared to the optimal STS performance obtained when considering all eight possible sources for a specific target: As the only way to achieve the best individual performance for a given

target dataset is to extensively test all possible sources and, then, choosing the best, the real-world applicability of an approach based on STS will be limited by the amount of available data sources. By using the proposed multiple

source approach, we can achieve, with high confidence, a performance for the chosen target similar to the individual best observed among all the chosen sources. This observation, however, does not compensate the fact that by manually choosing a single optimal source, the performance observed for the chosen target dataset is consistently better or, in the worst case, in a similar range to the one observed for MS-STs. Further research is needed in order to optimize the choice of source datasets so as to reduce this performance gap.

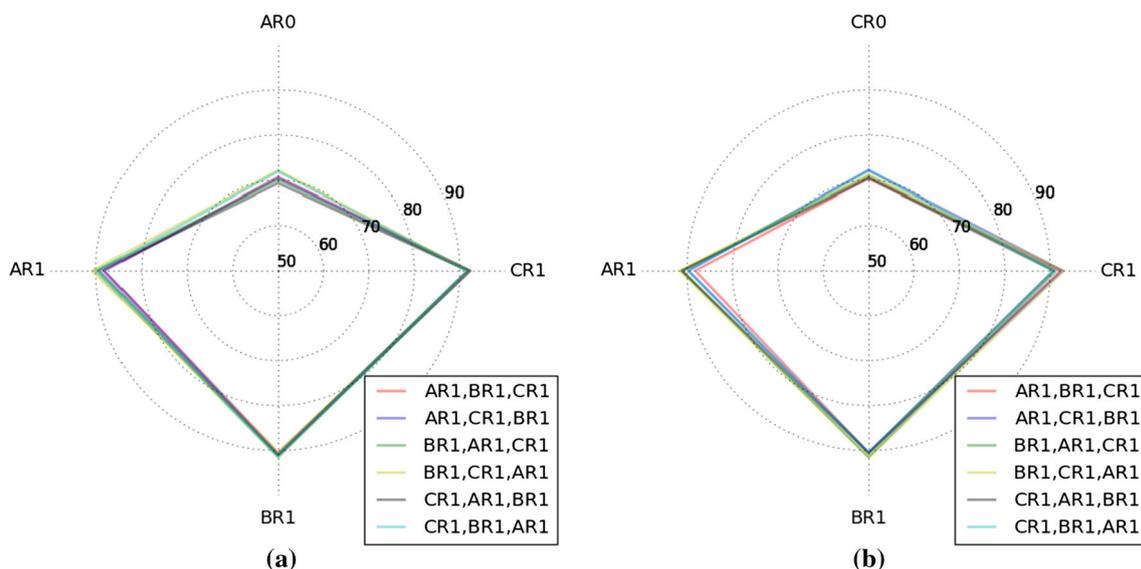
Another interesting observation concerns the effect of the order in which the sources  $[AR1, BR1, CR1]$  are considered during the cyclic evolution of the MS-STs process. In order to assess whether this order had any discernible effect over the observed performance, we chose to run, for each target dataset, a set of six variants of the original results, changing the order in which the three sources are organized during a single cycle:  $[[A, B, C], [A, C, B], \dots, [C, B, A]]$ . The performance plots from Fig. 7 seem to point to the conclusion that the performance in all four datasets converges to a set of values in very similar ranges, regardless of the chosen organization of source datasets along the MS-STs pipeline. This observation leads to the conclusion that, if the best sources are found, there is no need to optimize their order. Whereas, the presented example was considerably simple, with a very small number of sources, and in a practical application, there is no guarantee that the number of combinations becomes unfeasible for a brute force optimization step of their organization. The observed results seem to indicate that this optimization process might be less relevant, especially in scenarios such as the tested,

where all sources present a relatively similar nature (flash illumination in this specific case). It is still unclear, due to the preliminary nature of this study how increasing variability in the source dataset conditions would affect these observations. The focus of future research should, thus, fall on the optimal choice of sources so that the most complete domain generalization is achieved. With this in mind, the aim of future work would be to accurately and intelligibly perform classification under highly variable scenarios, especially using more heterogeneous sets of source information.

## 5 Conclusions and future work

In the present work, we proposed an extended version of the source–target–source approach to deep transfer learning, making use of multiple sources of information. We successfully applied the developed algorithm to the specific problem of cross-sensor biometrics, a recent field of research that aims to mitigate the performance drop observed when training and testing acquisition conditions are considerably heterogeneous.

We observed that, when compared to a state-of-the-art algorithm designed for single-sensor scenarios, the proposed STs and MS-STs approaches revealed a worse baseline performance, but managed to present a very interesting cross-sensor stability regardless of the nature of the data used in the training process. It is trivial to deduce that an improvement in the baseline performance of any of the proposed methodologies—CNN or SV-SDA—would, necessarily, result in a stable increase of performance in all



**Fig. 7** Graphical representation of the MS-STs rank-1 recognition rates obtained for all the six possible orders of the chosen source datasets. Results concern to **a** ARO and **b** CRO as targets

cross-sensor scenarios. Some ideas to achieve such improvement would necessarily consist on exploring alternatives to the SIFT description chosen for the super-vector generation, or on the development of ensemble or joint strategies capable of making the most of the pros of both GMM-UBM (or any other state-of-the-art single-sensor methodology) and STS strategies to simultaneously achieve good baseline and cross-sensor performance. Achieving a tight coupling between both methodologies will, most certainly, represent a very significant step in the field of cross-sensor biometrics.

Concerning the CNN methodology, it must be noted that, traditionally, this strategy is explored when large datasets are available, so as to achieve the most robust modeling possible. In the current work, we used only two training instances from each source dataset, thus, theoretically, limiting the potential of achieving good results for the problem at hand. With this observation in mind, we can conclude that testing the proposed approaches on a larger cross-sensor periocular dataset would probably result both in higher baseline and higher cross-sensor performances. One must note, however, that the availability of large amounts of data to perform the enrollment step is not guaranteed in real-life applications. This limitation should, therefore, be overcome in the long run if this strategy is expected to be implemented in more practical solutions. Another focus of future work for CNN would be using information from all three RGB channels instead of gray scale transformation used on the present work.

Regarding the MS-STs, we may conclude that even though the optimal STS performance managed to outperform its multi-source counterpart in almost all scenarios, the reasoning for this behavior is both expected and negligible for practical applications. As we are manually choosing the best-performing source when presenting the STS results, whereas in the MS-STs, we are fixing the same set of sources for all experiments, it is expected that optimal performance is not achieved in the situations where the best-performing single source is not included on the set of chosen sources. From a practical point of view, testing the universe of all possible sources to empirically choose the best one is not a viable possibility. The focus of future work should therefore fall on the automatic choice of the fittest sources to achieve the highest degree of domain generalization during the MS-STs learning process and, thus, cause the convergence of the MS-STs performance to the best possible STS result. As a future work, some testing is still needed to assess MS-STs using SV-SDA with instance weighting and pre-training the network with unlabeled biometric data.

**Acknowledgments** This work was financed by FEDER funds through the Programa Operacional Factores de Competitividade—

COMPETE and by Portuguese funds through FCT—Fundação para a Ciência e Tecnologia in the framework of the project PTDC/EIA-EIA/119004/2010. The second author would like to thank Fundação para a Ciência e Tecnologia (FCT)—Portugal the financial support for the PhD grant SFRH/BD/87392/2012.

## References

- Dai W, Xue GR, Yang Q, Yu Y (2007) Co-clustering based classification for out-of-domain documents. In: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp 210–219
- Yang J, Yan R, Hauptmann AG (2007) Cross-domain video concept detection using adaptive svms. In: Proceedings of the 15th international conference on Multimedia. ACM, pp 188–197
- Blitzer J, Dredze M, Pereira F et al (2007) Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification. *ACL* 7:440–447
- Dahlmeier D, Ng HT (2010) Domain adaptation for semantic role labeling in the biomedical domain. *Bioinformatics* 26(8): 1098–1104
- Pan SJ, Shen D, Yang Q, Kwok JT (2008) Transferring localization models across space. In: *AAAI*, pp 1383–1388
- Kandaswamy C, Silva LM, Cardoso JS (2015) Source-target-source classification using stacked denoising autoencoders. In: *Pattern recognition and image analysis*. Springer, pp 39–47
- Caruana R (1997) Multitask learning. *Mach Learn* 28(1):41–75
- Thrun S (1998) Lifelong learning algorithms. In: *Learning to learn*. Springer, pp 181–209
- Daume III H, Marcu D (2006) Domain adaptation for statistical classifiers. *J Artif Intell Res* 26:101–126
- Glorot X, Bordes A, Bengio Y (2011) Domain adaptation for large-scale sentiment classification: a deep learning approach. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pp 513–520
- Ben-David S, Blitzer J, Crammer K, Kulesza A, Pereira F, Vaughan JW (2010) A theory of learning from different domains. *Mach Learn* 79(1–2):151–175
- Bruzzone L, Marconcini M (2010) Domain adaptation problems: a DASVM classification technique and a circular validation strategy. *IEEE Trans Pattern Anal Mach Intell* 32(5):770–787
- Ciresan D, Meier U, Schmidhuber J (2012) Multi-column deep neural networks for image classification. In: *2012 IEEE conference on computer vision and pattern recognition (CVPR)*. IEEE, pp 3642–3649
- Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? In: *Advances in neural information processing systems*, pp 3320–3328
- Kandaswamy C, Silva LM, Alexandre LA, Santos JM, de Sá JM (2014) Improving deep neural network performance by reusing features trained with transductive transference. In: *Artificial neural networks and machine learning—ICANN 2014*. Springer, pp 265–272
- Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 35(8):1798–1828
- Connaughton R, Sgroi A, Bowyer KW, Flynn P (2011) A cross-sensor evaluation of three commercial iris cameras for iris biometrics. In: *IEEE computer society conference on computer vision and pattern recognition workshops (CVPRW)*. IEEE, pp 90–97
- Monteiro JC, Esteves R, Santos G, Fiadeiro PT, Lobo J, Cardoso JS (2015) A comparative analysis of two approaches to periocular recognition in mobile scenarios. In: *Advances in visual*

- computing, vol. 9475 of Lecture Notes in Computer Science. Springer International Publishing, pp 268–280
19. Pillai J, Puertas M, Chellappa R (2014) Cross-sensor iris recognition through Kernel learning. *IEEE Trans Pattern Anal Mach Intell* 36:73–85
  20. Santos G, Grancho E, Bernardo MV, Fiadeiro PT (2015) Fusing iris and periocular information for cross-sensor recognition. *Pattern Recogn Lett* 57:52–59
  21. Jillela R, Ross A (2014) Matching face against iris images using periocular information. In: 2014 IEEE international conference on image processing (ICIP). IEEE, pp 4997–5001
  22. Woodard DL, Pundlik S, Miller P, Jillela R, Ross A (2010) On the fusion of periocular and iris biometrics in non-ideal imagery. In: 2010 20th international conference on pattern recognition (ICPR). IEEE, pp 201–204
  23. Shao L, Zhu F, Li X (2015) Transfer learning for visual categorization: a survey. *IEEE Trans Neural Netw Learn Syst* 26(5):1019–1034
  24. Monteiro JC, Cardoso JS (2015) Periocular recognition under unconstrained settings with universal background models. In: Proceedings of the international conference on bio-inspired systems and signal processing (BIOSIGNALS)
  25. Reynolds DA, Quatieri TF, Dunn RB (2000) Speaker verification using adapted Gaussian mixture models. *Digit Signal Process* 10(1):19–41
  26. JCM et al (2015) Periocular recognition in mobile scenarios: a comparative analysis of two approaches. In: Proceedings of 11th international symposium on visual computing (ISVC)
  27. Campbell WM, Sturim DE, Reynolds DA (2006) Support vector machines using gmm supervectors for speaker verification. *IEEE Signal Process Lett* 13(5):308–311
  28. Ge Z, McCool C, Sanderson C, Corke P (2015) Modelling local deep convolutional neural network features to improve fine-grained image classification. arXiv preprint arXiv:1502.07802
  29. Kandaswamy C, Silva LM, Alexandre L, Sousa R, Santos JM, de Sá JM et al (2014) Improving transfer learning accuracy by reusing stacked denoising autoencoders. In: 2014 IEEE international conference on systems, man and cybernetics (SMC). IEEE, pp 1380–1387
  30. Tan T, He Z, Sun Z (2010) Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. *Image Vis Comput* 28(2):223–230
  31. Proença H, Alexandre L (2007) The NICE. I: noisy iris challenge evaluation—part I. In: First IEEE international conference on biometrics: theory, applications, and systems. IEEE, pp 1–4