

Instance-weighted Transfer Learning of Active Appearance Models

Daniel Haase, Erik Rodner, and Joachim Denzler

Computer Vision Group, Friedrich Schiller University of Jena, Germany
Ernst-Abbe-Platz 2-4, 07743 Jena, Germany

{daniel.haase, erik.rodner, joachim.denzler}@uni-jena.de

Abstract

There has been a lot of work on face modeling, analysis, and landmark detection, with Active Appearance Models being one of the most successful techniques. A major drawback of these models is the large number of detailed annotated training examples needed for learning. Therefore, we present a transfer learning method that is able to learn from related training data using an instance-weighted transfer technique. Our method is derived using a generalization of importance sampling and in contrast to previous work we explicitly try to tackle the transfer already during learning instead of adapting the fitting process. In our studied application of face landmark detection, we efficiently transfer facial expressions from other human individuals and are thus able to learn a precise face Active Appearance Model only from neutral faces of a single individual. Our approach is evaluated on two common face datasets and outperforms previous transfer methods.

1. Introduction and Motivation

Since their first introduction in [3, 9, 4], *Active Appearance Models* (AAMs) and related concepts of deformable models (*e.g.*, [10]) have become a vital part of the computer vision community. In general, AAMs are statistical generative models that describe deformable objects in images using a shape component and a shape-free texture component. Given a set of training images with annotated landmark positions, an AAM can easily be constructed, where, at its heart, *principal component analysis* (PCA) is used to parameterize shape and texture. A trained model can be fitted to new images efficiently and accurately, revealing landmark positions, model instance parameters, *et cetera*. AAMs have successfully been applied to a wide variety of scenarios, for instance medical imaging (*e.g.*, [15]). However, by far the most prominent application area of AAMs is face modeling and analysis, where it is used for tracking (*e.g.*, [34]), face recognition (*e.g.*, [9]), emotion classification (*e.g.*, [19, 21]),

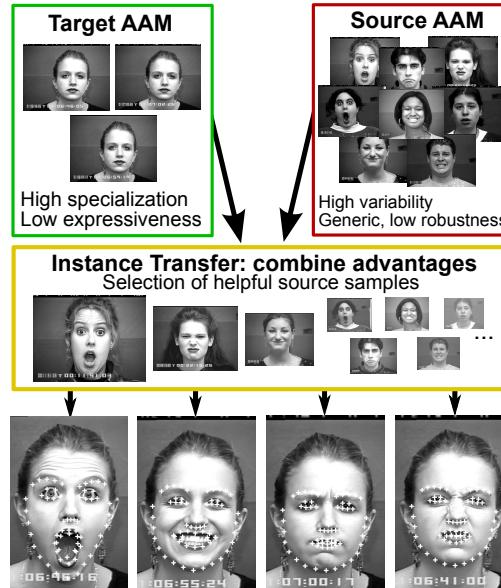


Figure 1. Overview of the proposed method and its application to face landmark detection. (Images: © Jeffrey Cohn, CK+ dataset)

and many other tasks.

One of the main disadvantages, as for instance stated in [12, 5, 25, 27, 20, 30], is their weak generalization ability when learned with only a few training examples that do not cover the complete range of possible variations in the data. For the face scenario, this includes cases of unseen pose, expression, illumination, or identity, and thus has a major impact on practical applications. As a result, numerous extensions of standard AAMs have been proposed to improve the fitting quality under such conditions. The majority of such AAM extensions can loosely be categorized based on how they tackle the problem, with the most common strategies being: (1) improvement of the actual fitting procedure by changing the factors involved in the optimization (*e.g.*, [22, 12, 13, 5]); and (2) usage of more robust feature representations, *e.g.* to obtain invariance with respect to illumination [23], occlusions [30], or non-linear shape deformations [14].

It is important to note that in either of the two cases, the improved generalization ability is gained by improving the way the model’s existing knowledge is used, and no knowledge about unseen cases and variations is added to the model itself. In this paper, we present an entirely contrary approach towards an improved AAM generalization ability: we explicitly increase the expressiveness of an AAM by adding new knowledge obtained from related but different training data. Considering the example of illumination changes in face images, this means that existing approaches aim to use illumination-invariant features of robust fitting metrics, while our approach is to add knowledge about unseen illumination conditions to an existing AAM. In machine learning, and specifically in classification, regression, and clustering tasks, such a strategy can be established by *transfer learning* [24], an area which has gained a lot of attention in recent years. For AAMs, however, the advances in this field have been largely ignored so far.

Similar to existing transfer learning scenarios, we assume that we have a specific target AAM with low expressiveness and thus weak generalization ability (*e.g.* only one illumination condition, no facial expressions, *etc.*), and a generic source AAM with high expressiveness. As described in [12], the high expressiveness of the source model leads to a reduced fitting robustness. By transferring knowledge from associated source data to the target AAM, we wish to keep the advantages of both models and improve the fitting performance on target-specific data. Of course, we could simply train a new model on the union of source and target AAM training samples (to which we refer to as “source \cup target” in the following referring to one of the main baselines for transfer learning [6]). However, this would completely ignore the difference between the underlying source and target distributions. On the contrary, our approach is based on a probabilistic formulation of the AAM transfer problem, which allows us to define and derive a direct solution incorporating the difference in the data distributions. As will be shown in the experiments, this approach is generally more favorable than the trivial solution “source \cup target.” In Fig. 1, the key intention of our approach is visualized for the example of face AAMs.

After discussing related work in the following section, we briefly present relevant parts of AAMs in Sect. 3. In Sect. 4, we introduce our probabilistic approach for AAM transfer, and additionally describe two heuristic baseline AAM transfer methods. Experiments and evaluations of the respective approaches are given in Sect. 5. Sect. 6 concludes this paper and points out possible future work.

2. Related Work

There is a huge body of literature on transfer learning techniques and we limit the following literature review only to specific papers either related to our instance transfer tech-

nique or used in the context of Active Appearance Models.

Very broadly speaking, transfer learning is all about exploiting data coming from different distributions during learning. In our case, training data from two distributions is given: source and target examples. Following the survey of Pan and Yang [24], transfer learning methods can be categorized into *instance transfer*, *feature transfer*, *parameter transfer*, and *relational transfer*. As will be shown in Sect. 4, the *instance transfer* approach is of particular relevance for our AAM scenario. In [16] and [17], for example, a technique is studied based on a probabilistic problem formulation and importance sampling for correcting sample selection bias. The authors of [11] show how to construct easier tasks by finding source examples similar to target examples and adding them to the target training set. After that a feature transfer technique is utilized to learn a kernel for the target task. In contrast, we propose a generalization of the instance weighting technique of [16], where we allow each source example to adaptively influence the transfer depending on their distance to the target example. This avoids hard decisions on the similarity as done in [11].

All of the previous methods have been studied only for the case of classification and regression. In our case, we are confronted with a structured prediction problem, because the landmark positions we would like to estimate have a complex dependency structure that needs to be tackled. Transfer learning for structured prediction problems has not been studied before [8] to the best of our knowledge with [31] being a rare exception that solely focuses on applications in computational biology [32]. However, several previous methods can be extended to structured learning, with the approach proposed in this paper being one example how to accomplish this.

For transfer learning from an AAM-specific point of view, a related application is the transfer of expressions between two face model instances. In [28], this problem is solved by finding a linear transformation between source and model parameter space. For models which do not share similar variation components, this may lead to under-articulated expressions in the target space. This problem is tackled in [7] by including a full variation transfer from source to target into the solution.

Approaches which increase the expressiveness of an AAM are very rare. One example is the *Online Appearance Model* [27], where the texture component is constantly being updated via incremental PCA during model fitting to account for illumination changes. Similarly, *Adaptive AAMs* [20] feature a generic and a subject-specific texture component, where the latter is again being updated during fitting. However, both methods add knowledge to the model only at fitting time, while we wish to increase the expressiveness of the model in a more general manner and before fitting. Additionally, both approaches update only the texture com-

ponent, which, for instance, excludes the possibility to add new facial expressions to existing AAMs.

Another line of research are sparse methods for single-sample face recognition [35].

3. Active Appearance Models

This section gives a brief overview of Active Appearance Models [3, 9, 4, 22] and focuses on the learning process and its assumptions. An AAM is trained from a set $\{(\mathbf{I}_n, \mathbf{s}_n) \mid n \in \{1, \dots, N\}\}$ of N annotated samples, each consisting of an image \mathbf{I}_n and L corresponding landmark coordinates $\mathbf{s}_n = (x_{n,1}, y_{n,1}, \dots, x_{n,L}, y_{n,L}) \in \mathbb{R}^{2L}$. As first step, a shape model is built by applying PCA on the matrix $\mathbf{S} = (\mathbf{s}_1, \dots, \mathbf{s}_N) \in \mathbb{R}^{2L \times N}$ of the sample landmarks. The resulting shape model is fully characterized by the *mean shape* $\boldsymbol{\mu} = \frac{1}{N} \sum_{n=1}^N \mathbf{s}_n \in \mathbb{R}^{2L}$, the matrix $\Phi \in \mathbb{R}^{2L \times K}$ of $K \leq \text{rank}(\mathbf{S})$ orthonormal *shape basis vectors*, and the vector $\boldsymbol{\lambda} \in \mathbb{R}^K$ of corresponding eigenvalues. An arbitrary shape sample \mathbf{s}^* can then be represented by

$$\hat{\mathbf{s}}^* = \Phi \mathbf{p}_{\mathbf{s}^*} + \boldsymbol{\mu}, \quad \text{with } \mathbf{p}_{\mathbf{s}^*} = \Phi^\top (\mathbf{s}^* - \boldsymbol{\mu}) \quad (1)$$

being the *shape parameters* of \mathbf{s}^* . The second step of AAM training consists of building a shape-free texture model. Here, each training image \mathbf{I}_n is warped from its given shape configuration \mathbf{s}_n to the mean shape $\boldsymbol{\mu}$ and vectorized to form the texture vector $\mathbf{t}_n \in \mathbb{R}^M$. Afterwards, PCA is applied on the matrix $\mathbf{T} = (\mathbf{t}_1, \dots, \mathbf{t}_N) \in \mathbb{R}^{M \times N}$ of sample textures. As for the shape model, the resulting texture model is fully characterized by the *mean texture* $\boldsymbol{\nu}$, the orthonormal *texture basis vectors* Ψ and the corresponding eigenvalues κ . An arbitrary texture sample \mathbf{t}^* is represented by

$$\hat{\mathbf{t}}^* = \Psi \mathbf{q}_{\mathbf{t}^*} + \boldsymbol{\nu}, \quad \text{with } \mathbf{q}_{\mathbf{t}^*} = \Psi^\top (\mathbf{t}^* - \boldsymbol{\nu}) \quad (2)$$

being the *texture parameters* of \mathbf{t}^* . An optional third step is to combine shape and texture models using another PCA [4]. As will be shown in Sect. 4, however, we base our model transfer on independent AAMs, where shape and texture components are not combined [22].

To fit an existing AAM to a given input image \mathbf{I}^* , model parameters $\mathbf{p}_{\mathbf{s}^*}$ and $\mathbf{q}_{\mathbf{t}^*}$ need to be found that minimize the error between the corresponding model instance and \mathbf{I}^* . For independent AAMs, the *inverse compositional/project-out* algorithm [22] is an efficient gradient descent based approach and is used in this work. The necessary pre-computations for this method are based on $\boldsymbol{\mu}$, Φ , $\boldsymbol{\lambda}$, $\boldsymbol{\nu}$, and Ψ only and thus can be run as a last step of AAM training.

4. Model Transfer for AAMs

As shown in [12], the performance of AAMs heavily depends on the variations included in the training set. In general, performing PCA with only a few given examples,

leads to severe overfitting especially with high-dimensional data, such as images. Therefore, we demonstrate in the following how to exploit training data from another but related distribution. In particular, we describe possible approaches to knowledge transfer from a given source AAM $\mathcal{A}_S = (\boldsymbol{\mu}_S, \Phi_S, \boldsymbol{\lambda}_S, \boldsymbol{\nu}_S, \Psi_S, \kappa_S)$ to improve the quality of a given target AAM $\mathcal{A}_T = (\boldsymbol{\mu}_T, \Phi_T, \boldsymbol{\lambda}_T, \boldsymbol{\nu}_T, \Psi_T, \kappa_T)$ in the target domain. We denote the version of the target AAM \mathcal{A}_T after the knowledge transfer by $\mathcal{A}^* = (\boldsymbol{\mu}^*, \Phi^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*, \Psi^*, \kappa^*)$. We begin by describing two straightforward heuristic approaches in Sect. 4.1, which—together with the trivial solutions *source only*, *target only*, and *source \cup target* [6]—will serve as a baseline for our instance transfer approach presented in Sect. 4.2.

4.1. Heuristic Transfer

Full Source Variation. As discussed in [7], a straightforward attempt to transfer a source AAM \mathcal{A}_S into the target domain is to use target AAM mean components $(\boldsymbol{\mu}_T, \boldsymbol{\nu}_T)$ and source AAM variation components $(\Phi_S, \boldsymbol{\lambda}_S, \Psi_S, \kappa_S)$ for the transferred model \mathcal{A}^* . Because source texture vectors use a different reference shape (namely $\boldsymbol{\mu}_S$) than the target texture vectors ($\boldsymbol{\mu}_T$), we need to transform the former into the reference shape of the latter. We denote this operation by $\mathcal{W}_{\boldsymbol{\mu}_S \rightarrow \boldsymbol{\mu}_T}(\Psi_S)$, and it can be easily performed similar to the image warping during the AAM training step (see [22] for details). As a result, we have

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}_T, \quad \Phi^* = \Phi_S, \quad \boldsymbol{\lambda}^* = \boldsymbol{\lambda}_S \quad (3a)$$

$$\boldsymbol{\nu}^* = \boldsymbol{\nu}_T, \quad \Psi^* = \mathcal{W}_{\boldsymbol{\mu}_S \rightarrow \boldsymbol{\mu}_T}(\Psi_S), \quad \kappa^* = \kappa_S. \quad (3b)$$

While this transfer approach requires almost no computational effort, it implicitly assumes that the data distributions of the source and target domain differ only in their expected values. For the example of faces, this simple approximation usually leads to visual artifacts in the transferred model [7].

Subspace Transfer. Inspired by [1] and [28], a second AAM transfer method is to unite the subspaces spanned by the source and target basis vectors. For the case of the shape model, target and source subspaces $V_T, V_S \subseteq \mathbb{R}^{2L}$ are given by $V_T = \text{span}(\Phi_T)$ and $V_S = \text{span}(\Phi_S)$. To obtain a new set of orthonormal shape basis vectors Φ^* which spans the subspace V^* with $V_T, V_S \in V^*$, we simply have to find the QR decomposition $(Q_{(\Phi_T, \Phi_S)}, R_{(\Phi_T, \Phi_S)})$ of the matrix (Φ_T, Φ_S) . The transferred shape model of \mathcal{A}^* with

$$\begin{aligned} \boldsymbol{\mu}^* &= \boldsymbol{\mu}_T, & \Phi^* &= Q_{(\Phi_T, \Phi_S)}, \\ \boldsymbol{\lambda}^* &= \text{diag}(R_{(\Phi_T, \Phi_S)}) \cdot (\boldsymbol{\lambda}_T^\top, \boldsymbol{\lambda}_S^\top)^\top \end{aligned} \quad (4)$$

then contains the full variation of the target model plus the part of the source model variation which can not be represented by the target model. The eigenvalues $\boldsymbol{\lambda}^*$ of the new model are simply the concatenated target and source

eigenvalues scaled by the length of the orthogonalized basis vectors. All operations for the texture model are identical to the shape case, but again the reference shape of the source texture vectors has to be transformed beforehand via $\mathcal{W}_{\mu_S \rightarrow \mu_T}(\Psi_S)$.

The main disadvantage of this method is that it is solely based on the already estimated covariance matrix of the source model. It therefore neglects cases, when only a subset of the source examples are related to the target domain. For example in our face model application, it is important to transfer information and learned variations only from similar individuals, as will be shown in Sect. 5.

4.2. An Instance Transfer Approach

In the following, we present a new instance transfer approach for AAMs. Instead of a heuristic model, our method is based on a probabilistic formulation of the transfer problem and a generalization of importance sampling, which allows us to incorporate information from both source and target samples directly. Due to the similarity between an AAM's shape and texture component (*cf.* Eq. 1 and Eq. 2), we only present the derivation in detail for the shape component. For the texture component, only differences to the shape case are discussed.

Shape Component. Given N_S training shape samples $s_{1,S}, \dots, s_{N_S,S}$ from the source domain S and N_T training samples $s_{1,T}, \dots, s_{1,N_T}$ from the target domain T , our goal is to estimate the mean shape μ^* , the matrix Φ^* of K orthonormal shape basis vectors, and their corresponding variances λ^* . As shown for the example of face images in [28], the estimation of the target mean shape μ_T is a non-critical operation, and gives reasonable results even for small sample sizes N_T . We therefore use the target mean μ_T as estimation for μ^* , and our main focus is to find Φ^* and, by association, λ^* . For standard AAMs—*i.e.* when only training data from one domain is available—PCA is employed for this task, as it minimizes the reconstruction error of the input samples. To derive our transfer method, we take a step back and study the case of an infinite number of training samples, which allows to minimize the *expected reconstruction error* instead taking the whole input distribution into account. Finding the orthonormal basis $\Phi^* \in \mathbb{R}^{2L \times K}$ with the smallest expected reconstruction error *in the target domain* can be formalized as follows:

$$\Phi^* = \underset{\Phi: \Phi^\top \Phi = \mathbb{1}}{\operatorname{argmin}} \mathbb{E}_{s \sim p_T} \left(\underbrace{\|s - \Phi \Phi^\top s\|^2}_{\epsilon(\Phi, s)} \right) \quad (5)$$

with

$$\xi = \mathbb{E}_{s \sim p_T} (\epsilon(\Phi, s)) = \int \epsilon(\Phi, s) p_T(s) ds. \quad (6)$$

Here, s is an arbitrary *mean centered* sample from either the source or the target domain, and p_T is the probability

distribution of shapes in the target domain. To relate this formulation to the source domain, we use a transformation of the integral in Eq. 6 known from importance sampling (*e.g.*, [16]) and generalize it to incorporate a weight factor $0 \leq \alpha \leq 1$:

$$\begin{aligned} \xi &= \int \epsilon(\Phi, s) \left(\underbrace{\alpha + (1 - \alpha) \frac{p_S(s)}{p_T(s)}}_{=1} \right) p_T(s) ds \\ &= \alpha \int \epsilon(\Phi, s) p_T(s) ds + (1 - \alpha) \int \epsilon(\Phi, s) \frac{p_T(s)}{p_S(s)} p_S(s) ds \\ &= \alpha \mathbb{E}_{s \sim p_T} (\epsilon(\Phi, s)) + (1 - \alpha) \mathbb{E}_{s \sim p_S} \left(\epsilon(\Phi, s) \frac{p_T(s)}{p_S(s)} \right). \end{aligned}$$

This trick allows us to express the reconstruction error in the target domain in terms of the target and the source data distribution. The parameter α can be used to trade-off the individual estimates. We are now able to go back to the case of a finite number of given examples by replacing the expected values in above equation by their respective empirical counterparts. We use the target samples for the target domain estimation and the source samples for the source domain estimation, and finally have

$$\begin{aligned} \Phi^* &= \underset{\Phi: \Phi^\top \Phi = \mathbb{1}}{\operatorname{argmin}} \left(\frac{\alpha}{N_T} \sum_{n_T=1}^{N_T} \epsilon(\Phi, s_{n_T}) \right. \\ &\quad \left. + \frac{1 - \alpha}{N_S} \sum_{n_S=1}^{N_S} \epsilon(\Phi, s_{n_S}) \underbrace{\frac{p_T(s_{n_S})}{p_S(s_{n_S})}}_{=w_{n_S}} \right). \quad (8) \end{aligned}$$

The weights w_{n_S} have an important role, because they control the individual influence of each source example on the estimation and thus on the transfer. If S_T and S_S denote the data matrices of mean centered target and source samples, it can be easily shown that the solution of Eq. 8 is given by applying PCA to the new data matrix $S^* = (S_T W_T^{\frac{1}{2}}, S_S W_S^{\frac{1}{2}})$ with the corresponding weight matrices

$$W_T = \frac{\alpha}{N_T} \mathbb{1} \quad \text{and} \quad W_S = \frac{1 - \alpha}{N_S} \operatorname{diag}(w_1, \dots, w_{N_S})$$

where $\mathbb{1}$ denotes the identity matrix. The PCA of S^* corresponds to the eigendecomposition $V \Lambda V^{-1}$ of the *weighted* combination of the *weighted* target and source covariance matrices, *i.e.*

$$V \Lambda V^{-1} = S^* S^{*\top} = S_T W_T S_T^\top + S_S W_S S_S^\top. \quad (9)$$

The resulting shape basis vectors and eigenvalues are $\Phi^* = V$ and $\lambda^* = \operatorname{diag}(\Lambda)$. For the special case of $w_1 = \dots = w_{N_S} = 1$, this result is identical to *smart PCA* [33], a method derived using the *probabilistic PCA* framework [29]. Consequently, as for smart PCA, $\alpha \in [0, 1]$ can be interpreted

as regularization parameter which governs how much we trust the target and source samples. For $\alpha = 1$, the resulting AAM is identical to the target AAM. Additionally, it can easily be verified that the “full source variation” transfer heuristic presented in Sect. 4.1 is also just a special case of our approach for $\alpha = 0$ and $w_1 = \dots = w_{N_S} = 1$. Both methods are unable to perform an instance-based transfer and are thus heavily restricted.

Texture Component. Aside from the aforementioned need to transform source texture vectors into the reference shape of the target textures via $\mathcal{W}_{\mu_S \rightarrow \mu_T}$ (*cf.* Sect. 4.1), theoretically all steps are identical to the shape equivalents. Practically, however, the dimension M of the texture space is usually much larger than the dimension $2L$ of the shape space or the sample size N . For the commonly used IMM face dataset [26], for example, we have $M \approx 30,000$, $2L = 116$, and $N = 240$. As for the texture component of standard AAMs, it is therefore advisable to obtain the principal components via eigendecomposition of $\mathbf{T}^{*\top} \mathbf{T}^*$ instead of $\mathbf{T}^* \mathbf{T}^{*\top}$, where \mathbf{T}^* denotes the texture counterpart of \mathbf{S}^* (see [2, Sect. 4.5.3] for more details).

Estimation of Sample Weights. The estimation of the source sample weights $w_{n_S} = \frac{p_T(s_{n_S})}{p_S(s_{n_S})}$ for $n_S \in \{1, \dots, N_S\}$ is a crucial step of instance transfer approaches (see [24] for examples). Below we first derive a straightforward, theoretically oriented approximation for the case of AAMs. Afterwards, we present a more heuristic approach which addresses drawbacks of the former. Again, we will only handle the shape component explicitly, as the texture component can be treated in the very same manner.

As described in Sect. 3, for AAMs an arbitrary shape sample s can be represented by its shape parameters $\mathbf{p}_s = \Phi^\top(s - \mu)$. We can model the shape space density $p(s)$ using the parameter space density $p(\mathbf{p}_s)$. From the definition of PCA it follows that \mathbf{p}_s is mean free with covariance $\text{diag}(\boldsymbol{\lambda})$, and with a normal assumption we have $\mathbf{p}_s \sim \mathcal{N}(\mathbf{0}, \text{diag}(\boldsymbol{\lambda}))$. This model relates to a degenerate normal distribution with low-rank covariance matrix for s and is helpful to tackle high dimensional data. Taking all the model assumptions into account, one way to approximate the source sample weights for AAM transfer is therefore

$$\frac{p_T(s_{n_S})}{p_S(s_{n_S})} = \frac{\mathcal{N}(\Phi_T^\top(s_{n_S} - \mu_T); \mathbf{0}, \text{diag}(\boldsymbol{\lambda}_T))}{\mathcal{N}(\Phi_S^\top(s_{n_S} - \mu_S); \mathbf{0}, \text{diag}(\boldsymbol{\lambda}_S))}. \quad (10)$$

In practice, however, such a direct approach can have several drawbacks [16]. First of all, the normal assumption does not hold for typical scenarios such as face images. Secondly, it might very well be the case that only few target samples are given—even just one target sample is possible. In this case, the estimation of the target density $p_T(s_{n_S})$ is an ill-posed problem and is likely to distort the resulting sample weight w_{n_S} .

Following the argumentation of [16], we thus propose a heuristic alternative to Eq. 10. We interpret w_1, \dots, w_{N_S} as general weights to select favored and unfavored source samples for the transfer. For AAMs, our notion of a useful source sample is twofold: On the one hand we want to reward innovation, *i.e.* source samples which carry information that is not covered in the target model. On the other hand, we want to prevent negative transfer from source samples which, in the eyes of the target model, are very dissimilar from what was previously seen. Both *innovation* and *target space dissimilarity* can easily be computed in the case of AAMs: To quantify innovation, we compute the reconstruction error of the source sample using the target model. The larger the error, the more information is present which can not be represented by the model. For the dissimilarity, we use the difference between the projection $\Phi_T \Phi_T^\top(s_{n_S} - \mu_T) + \mu_T$ of s_{n_S} into the target model space and the mean shape μ_T of the target model. For the final weight, both factors are combined by dividing the innovation term with the target space dissimilarity estimate.

The practical quality of the different weighting schemes is analyzed in detail in the experimental section.

Applicability and Limitations. Our approach assumes that shape and texture component of an AAM can be treated separately, which limits its application to independent AAMs as in [22]. However, as the fitting method presented in [22] and its countless derivations (*e.g.*, [30]) is probably the most popular variant used for AAMs in the last decade, this is not an overly restricting factor. Considering the necessary amount of samples in the target and source domain, our approach is very flexible as will be shown in the experimental section.

5. Experiments and Results

Because the vast majority of AAM applications is related to face analysis, we base the evaluation of the presented transfer approaches on two popular face benchmark datasets, namely IMM [26] and CK+ [18, 21]. Both datasets feature a multitude of different individuals, each performing a set of standardized head movements and expressions. Ground-truth landmark positions are given for typical facial landmarks covering the eyes, mouth, jaw, *etc.*, but are not compatible between the datasets. Details and example images of the utilized datasets are given in Table 1.

To assess the transfer ability in this facial expression context, we constructed a test scenario where we have (i) a single-person target AAM with very limited expressiveness, and (ii) a generic multi-person source model covering all typical facial actions. Specifically, we assume the target model to be trained on a very small number of samples, all of which showing a neutral facial expression and frontal head pose. As a result, the target AAM models are not able to detect other face expressions, such as smiling. The source model,

Table 1. Overview of the face datasets used for our AAM transfer experiments. We used only neutral images of one person as target model, while the source model is trained on a large variety of persons and variations. © Jeffrey Cohn for images of CK+.

Dataset	IMM [26]	CK+ [18, 21]
Persons	40	123
Landmarks	58	68
Annotated Images per Person	6	2–22
Variations	Neutral, head pose, expression (smile and random), illumination	Neutral, expression (anger, disgust, fear, happiness, etc.)
Typical Target Training Images		
Typical Source Training Images		

Table 2. Quantitative results of different AAM transfer methods on the IMM [26] and CK+ [18, 21] face datasets. * : the differences are significant.

AAM Transfer Method	Median Geom. Error (px)	
	IMM	CK+
Source only [6]	5.02	5.78
Target only [6]	4.14	5.24
Source \cup target [6]	4.80	4.96
Full source variation [7]	3.64	4.10
Subspace transfer [1, 28]	3.05*	4.38
Ours (instance transfer)	2.93*	3.88

on the contrary, is able to handle all common expressions, poses, and illumination conditions, while representing no specific identity. We ensured that no image of the target person was included in the corresponding source model. After the transfer, we evaluate the Euclidean landmark errors with respect to the ground truth positions (“geometric error”) for all available images of the target person. Due to the reason that for the CK+ dataset, every second annotated image shows a neutral expression, all neutral images were excluded from the evaluation to prevent a bias. In the ideal transfer case, we expect to see a resulting AAM to be good at representing both the identity as well as expressions, poses, and illumination conditions of the target person, and thus should perform better than the generic source model [12].

Proof-of-Concept For a proof of concept, we tested all presented transfer approaches (*cf.* Sect. 4) against the commonly used baseline methods “source only”, “target only”, and “source \cup target” [6]. Additionally, we included results of the special case of constant source sample weights of our instance transfer approach, which is identical to “smart PCA” ([33], *cf.* Sect. 4). Results of the qualitative evalua-

for the IMM and the CK+ dataset are shown in Table 2.

First of all, it can be stated that the baseline methods perform worst for both datasets highlighting the importance of transfer learning techniques. The best result is obtained by our instance transfer approach with weights calculated as proposed in Sect. 4.2, which reduces the median geometric error of the source \cup target model by 40% for IMM and 20% for the CK+ dataset. The performance of the two heuristic transfer methods “full source variation” [7] and “subspace transfer” (inspired by [1, 28]) lies between the baseline methods and our approach. While the former performs better on the IMM dataset, the latter yields a smaller median error on the CK+ dataset.

Though only trained on few neutral target samples, the “target only” model performs best for all baseline methods. This effect can be explained by the fact that even for unseen expressions, poses, *etc.*, a certain amount of landmarks can still be explained well by the neutral model (*e.g.* the nose landmarks for the “smiling” expression).

As a conclusion, we can state that it is favorable to perform any type of AAM transfer instead of using the baseline methods. Secondly, there is a significant difference between previously published heuristic transfer methods and our proposed instance transfer approach ($p < 0.001$ checked with a paired U-test). Exemplary qualitative results for the performance of the described AAM transfer approaches are shown in Fig. 2.

Source Sample Weighting Schemes The estimation of the source sample weights w_{ns} plays a central role for our instance transfer approach. We presented three different weighting schemes: a probabilistically motivated estimation, a heuristic approach, and the special case of constant weights which degenerates to the smart PCA method [33]. Similarly to the evaluation in the previous section, we tested the performance of these three weighting schemes on the IMM dataset. Surprisingly, the probabilistic estimation yields by far the worst results, with a median geometric error of 3.72 pixels. This performance is comparable to the two heuristic transfer approaches (*cf.* Table 2), and indicates that the density model for the source and target distribution was not properly estimated (*cf.* Eq. 10). For the constant weighting scheme, we obtained a median error of 2.97 pixels, which is very close to the one of heuristic weighting with 2.93 pixels (*cf.* Table 2). We therefore conclude that the heuristic, we introduced for the source sample weight estimation, is reasonable.

Shape and Texture Components While the general importance of an AAM’s shape and texture component is discussed in [12], here we are particularly interested in finding the contribution of each of these components to the model transfer quality. As baseline, we use our instance transfer approach with heuristic weights, which gives the best results for all tested methods (*cf.* Table 2). For the IMM dataset, the baseline method has a median geometric error of 2.93

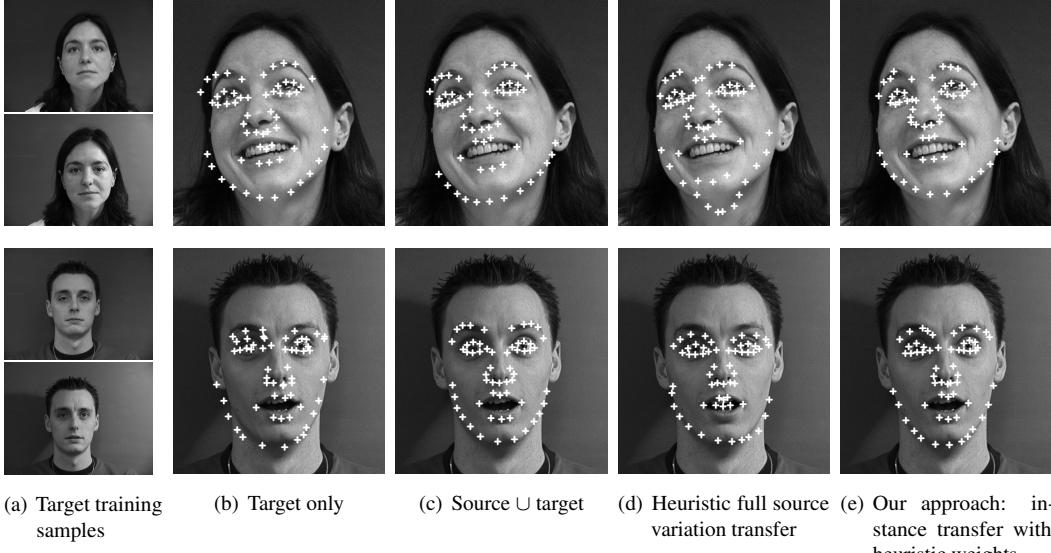


Figure 2. Qualitative results for different AAM transfer approaches for two example cases of the IMM face dataset.

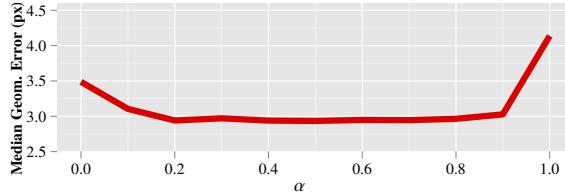


Figure 3. Influence of the regularization parameter α (Sect. 4.2) on our instance transfer approach. For $\alpha = 0$ our method degenerates to the full source variation heuristic [7]. For $\alpha = 1$, the resulting model is identical to the target AAM. Aside from the two extreme cases, this evaluation shows that the model transfer performance is not sensitive to the actual choice of the regularization factor.

pixels. When transferring only the shape component of the AAM with the identical experimental setup, we get a median error of 3.39 pixels. For the opposite case of a texture-only transfer, the corresponding median error is 3.81 pixels.

As to be expected, both single component transfers perform substantially worse than the joint transfer of both components. Their performance is comparable to the proposed heuristic transfer approaches (*cf.* Table 2), which implies that even the transfer of only one model component is superior to methods that do not transfer any knowledge at all. Interestingly, the error for shape-only transfer is comparable to the texture-only transfer, with a small advantage for the former. This means that the shape component is slightly more important for the generalization ability of AAMs and confirms similar findings of [12].

Regularization Parameter α For the derivation of our instance transfer approach, we introduced a regularization parameter $\alpha \in [0, 1]$ which governs how much we trust the data from the target and source domain (Sect. 4.2). For $\alpha = 0$, no target variation knowledge is used for the transfer, and (for additionally constant source sample weights, *cf.* Sect. 4.2), our method degenerates to the full source vari-

ation heuristic [7]. The other extreme, $\alpha = 1$, results in a model which is identical to the target AAM. To learn how this parameter influences the transfer of AAMs, we evaluated identical scenarios for different values of α . Again, our instance transfer approach with heuristic source sample weights and $\alpha = 0.5$ serves as the baseline model, as it gave the best results so far.

The resulting median errors are shown in Fig. 3. For the extreme cases $\alpha = 0$ and $\alpha = 1$ the transfer quality quickly decays and reaches the performance of the “full source variation” method and the “target only” model. However, for $\alpha \in [0.2, 0.9]$, the resulting transfer quality remains almost unchanged. We therefore conclude that for the AAM transfer, the new variation components do not have to be excessively prominent (*e.g.* in terms of large eigenvalues).

Computational Aspects We used our own C++ implementation of AAMs with the inverse compositional/project-out fitting algorithm [22]. As times needed for the fitting pre-computations and actual model fitting are similar for all transfer approaches, we only state computation times needed for the model transfer. Naturally, transfer is fastest for the “full source variation” technique, as no major computations have to be performed. On average, a time of 59 ms was needed for the whole transfer process. Similarly, the “subspace transfer” only has to perform vector projections and matrix concatenations, and needed 92 ms on average. The presented instance transfer method took about 8.8 s per transfer, mainly because weighted covariance matrices have to be estimated for source and target samples.

Further Analyses For further analyses of our approach, please refer to the supplementary material provided at <http://www.inf-cv.uni-jena.de/aam-transfer>.

6. Conclusions

We presented a transfer learning approach for Active Appearance Models, which generalizes as well as outperforms previous transfer learning methods and is additionally easy to implement and integrate into existing AAMs. Our method is based on integrating training data from an additional source training dataset by estimating example-specific weights that influence the degree of transfer and selects similar but still informative examples. We evaluated our method on two important face datasets and studied several aspects of it and the resulting impact on facial landmark detection accuracy.

In future work, we plan to extend our approach towards new challenging application areas for deformable part models, such as object detection [10]. Furthermore, an important goal will be to integrate unlabeled data for a more accurate instance weight estimation.

Acknowledgements

This research was supported by grant DE 735/8-1 of the German Research Foundation (DFG).

References

- [1] R. Anderson, B. Stenger, V. Wan, and R. Cipolla. Expressive visual text-to-speech using active appearance models. In *CVPR*, pages 3382–3389, 2013. [3](#), [6](#)
- [2] I. N. Bronshtein, K. A. Semendyayev, G. Musiol, and H. Mühlig. *Handbook of Mathematics*. Springer, 2007. [5](#)
- [3] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *ECCV*, pages 484–498, 1998. [1](#), [3](#)
- [4] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *PAMI*, 23(6):681–685, 2001. [1](#), [3](#)
- [5] T. F. Cootes and C. J. Taylor. An algorithm for tuning an active appearance model to new data. In *BMVC*, pages 919–928, 2006. [1](#)
- [6] H. Daumé. Frustratingly easy domain adaptation. In *ACL*, 2007. [2](#), [3](#), [6](#)
- [7] M. de la Hunty, A. Asthana, and R. Goecke. Linear facial expression transfer with active appearance models. In *ICPR*, pages 3789–3792, 2010. [2](#), [3](#), [6](#), [7](#)
- [8] T. G. Dietterich, P. Domingos, L. Getoor, S. Muggleton, and P. Tadepalli. Structured machine learning: the next ten years. *Machine Learning*, 73(1):3–23, 2008. [2](#)
- [9] G. J. Edwards, T. F. Cootes, and C. J. Taylor. Face recognition using active appearance models. In *ECCV*, volume 1407, pages 581–595, 1998. [1](#), [3](#)
- [10] P. F. Felzenszwalb, R. B. Girshick, D. A. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *PAMI*, 32(9):1627–1645, 2010. [1](#), [8](#)
- [11] B. Gong, K. Grauman, and F. Sha. Connecting the dots with landmarks: Discrim. learning domain-invariant features for unsup. domain adaptation. In *ICML*, 2013. [2](#)
- [12] R. Gross, I. Matthews, and S. Baker. Generic vs. person specific active appearance models. *Image and Vision Computing*, 23(12):1080–1093, 2005. [1](#), [2](#), [3](#), [6](#), [7](#)
- [13] R. Gross, I. Matthews, and S. Baker. Active appearance models with occlusion. *Image and Vision Computing*, 24(6):593–604, 2006. Face Processing in Video Sequences. [1](#)
- [14] O. C. Hamsici and A. M. Martínez. Active appearance models with rotation invariant kernels. In *ICCV*, pages 1003–1009, 2009. [1](#)
- [15] J. Hansegård, S. Urheim, K. Lunde, and S. I. Rabben. Constrained active appearance models for segmentation of triplane echocardiograms. *TMI*, 26(10):1391–1400, 2007. [1](#)
- [16] J. Huang, A. J. Smola, A. Gretton, K. M. Borgwardt, and B. Schölkopf. Correcting sample selection bias by unlabeled data. In *NIPS*, pages 601–608, 2006. [2](#), [4](#), [5](#)
- [17] J. Jiang and C. Zhai. Instance weighting for domain adaptation in NLP. In *ACL*, pages 264–271, 2007. [2](#)
- [18] T. Kanade, Y. Tian, and J. F. Cohn. Comprehensive database for facial expression analysis. In *Autom. Face and Gesture Recogn.*, pages 46–53, 2000. [5](#), [6](#)
- [19] A. Lanitis, C. J. Taylor, and T. F. Cootes. Automatic interpretation and coding of face images using flexible models. *PAMI*, 19(7):743–756, 1997. [1](#)
- [20] X. Liu. Video-based face model fitting using adaptive active appearance model. *IVC*, 28(7):1162–1172, 2010. [1](#), [2](#)
- [21] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *CVPR Workshops*, pages 94–101, 2010. [1](#), [5](#), [6](#)
- [22] I. Matthews and S. Baker. Active appearance models revisited. *IJCV*, 60(2):135–164, 2004. [1](#), [3](#), [5](#), [7](#)
- [23] R. Navarathna, S. Sridharan, and S. Lucey. Fourier active appearance models. In *ICCV*, pages 1919–1926, 2011. [1](#)
- [24] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.*, 22(10):1345–1359, 2010. [2](#), [5](#)
- [25] D. Pizarro, J. Peyras, and A. Bartoli. Light-invariant fitting of active appearance models. In *CVPR*, 2008. [1](#)
- [26] M. B. Stegmann, B. K. Ersbøll, and R. Larsen. FAME – a flexible appearance modelling environment. *IEEE Trans. on Medical Imaging*, 22(10):1319–1331, 2003. [5](#), [6](#)
- [27] J. Sung and D. Kim. Adaptive active appearance model with incremental learning. *PRL*, 30(4):359–367, 2009. [1](#), [2](#)
- [28] B.-J. Theobald, I. A. Matthews, J. F. Cohn, and S. M. Baker. Real-time expression cloning using appearance models. In *ICMI*, pages 134–139, 2007. [2](#), [3](#), [4](#), [6](#)
- [29] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999. [4](#)
- [30] G. Tzimiropoulos, J. A. i Medina, S. Zafeiriou, and M. Pantic. Generic active appearance models revisited. In *ACCV*, pages 650–663, 2012. [1](#), [5](#)
- [31] C. Widmer, M. Kloft, X. Lou, and G. Rätsch. Regularization-based multitask learning. *KI*, 2013. [2](#)
- [32] Q. Xu and Q. Yang. A survey of transfer and multitask learning in bioinformatics. *JCSB*, 5(3):257–268, 2011. [2](#)
- [33] Y. Zhang. Smart PCA. In *IJCAI*, pages 1351–1356, 2009. [4](#), [6](#)
- [34] J. Zhu, S. C. H. Hoi, and M. R. Lyu. Real-time non-rigid shape recovery via active appearance models for augmented reality. In *ECCV*, pages 186–197, 2006. [1](#)
- [35] L. Zhuang, A. Y. Yang, Z. Zhou, S. S. Sastry, and Y. Ma. Single-sample face recognition with image corruption and misalignment via sparse illumination transfer. In *CVPR*, pages 3546–3553, 2013. [3](#)