Coupled Gaussian Process Regression for Pose-Invariant Facial Expression Recognition

Ognjen Rudovic[1], Ioannis Patras[2], and Maja Pantic[1,3]

[1] Comp. Dept, Imperial College, London, UK
[2] Elec. Eng. Dept, Queen Mary University, London, UK
[3] EEMCS, University of Twente, 7500 AE Enschede, The Netherlands
{o.rudovic,m.pantic}@imperial.ac.uk, i.patras@elec.qmul.ac.uk

Abstract. We present a novel framework for the recognition of facial expressions at arbitrary poses that is based on 2D geometric features. We address the problem by first mapping the 2D locations of landmark points of facial expressions in non-frontal poses to the corresponding locations in the frontal pose. Then, recognition of the expressions is performed by using any state-of-the-art facial expression recognition method (in our case, multi-class SVM). To learn the mappings that achieve pose normalization, we use a novel Gaussian Process Regression (GPR) model which we name Coupled Gaussian Process Regression (CGPR) model. Instead of learning single GPR model for all target pairs of poses at once, or learning one GPR model per target pair of poses independently of other pairs of poses, we propose CGPR model, which also models the couplings between the GPR models learned independently per target pairs of poses. To the best of our knowledge, the proposed method is the first one satisfying all: (i) being face-shape-model-free, (ii) handling expressive faces in the range from -45° to $+45^\circ$ pan rotation and from -30° to $+30^\circ$ tilt rotation, and (iii) performing accurately for continuous head pose despite the fact that the training was conducted only on a set of discrete poses.

1 Introduction

Facial expression recognition has attracted significant attention because of its usefulness in many applications such as human-computer interaction, face animation and analysis of social interaction [1,2]. Most existing methods deal with images (or image sequences) in which depicted persons are relatively still and exhibit posed expressions in nearly frontal view [1]. However, most of real-world applications relate to spontaneous human-to-human interactions (e.g., meeting summarization, political debates analysis, etc.), in which the assumption of having immovable subjects is unrealistic. This calls for a joint analysis of head pose and facial expressions. Nonetheless, this remains a significant research challenge mainly due to the large variation in the appearance of facial expressions in different views and the difficulty in decoupling these different sources of variation.

Most of the existing approaches that perform pose-invariant facial expression recognition are based on 3D face models. For example, Chang et al. [3] built

K. Daniilidis, P. Maragos, N. Paragios (Eds.): ECCV 2010, Part II, LNCS 6312, pp. 350–363, 2010.

a probabilistic model on the generalized expression manifold obtained from 3D facial expression range data to recognize the prototypic facial expressions. To the same aim and to analyze the dynamics of facial expressions, Sun and Yin [4] applied 3D dynamic facial surface descriptors. Furthermore, several works proposed to apply 3D Active Appearance Models (AAM) for pose-invariant facial expression analysis (e.g. Sung and Kim [5], Cheon and Kim [6]). Zhu and Ji [7] used 3D Point Distribution Model (3D-PDM) and normalized SVD to recover the facial expression and pose. Wang and Lien [8] used similar 3D-PDM to separate the rigid head rotation from non-rigid facial expressions. Kumano et al. [9] applied a rigid face shape model to build person-dependent descriptors that were later used to decompose facial pose and expression simultaneously. Despite the fact that 3D face models have advantage over 2D approaches in that the effect of head pose on the facial expression analysis can be removed (although this usually comes at the expense of the recovered facial expression accuracy), the main disadvantage is the use of generative models and fitting techniques that can fail to converge. Also, most of these methods are computationally expensive and in need of time-consuming initialization process (e.g. due to manual annotation of more than 60 facial landmark points). Moreover, some of the aforementioned methods such as AAM need to be trained for each person/ facial expression/ head pose separately which makes those methods difficult to apply in real-world applications where unknown subjects/ expressions can be expected.

In contrast to increasing interest in pose-invariant facial expression analysis based on 3D and 2D face-shape models, pose-invariant facial expression analysis based on 2D shape-free methods has been scarcely investigated. This is mostly due to the fact that rigid head motions and non-rigid facial expressions are non-linearly coupled in 2D and difficult to decouple using existing algorithms [7]. For this reason, most of the proposed 2D pose-invariant methods address the problem of (expressionless) face recognition but not the problem of facial expression recognition (e.g. [10]). To the best of our knowledge, the only work that analyzed the problem of pose-invariant facial expression recognition using a 2D shape-free approach is the work by Hu et al. [11]. They proposed a set of pose-wise facial expression classifiers that are used to discriminate simultaneously facial expressions and horizontal head orientations at five pan angles (0°, 30°, 45°, and 90°). However, the performance of this method has not been analyzed for unknown head poses, i.e. poses that were not used to train the classifiers. Moreover, because the classifiers were trained pose-wise, it is not possible to perform recognition of facial expressions that were not included in the training dataset for the given pose (in other words, this facial expression recognition method cannot generalize across poses).

In this paper we propose a 2D face-shape-free method for pose-invariant facial expression recognition. We address the problem by mapping 2D facial points (e.g., mouth corners) from non-frontal poses to the frontal pose where the recognition of facial expressions can be performed by using any state-of-the art facial expression recognition method. The proposed three-step approach is illustrated in Fig. 1. In the first step, we perform head pose estimation by projecting the

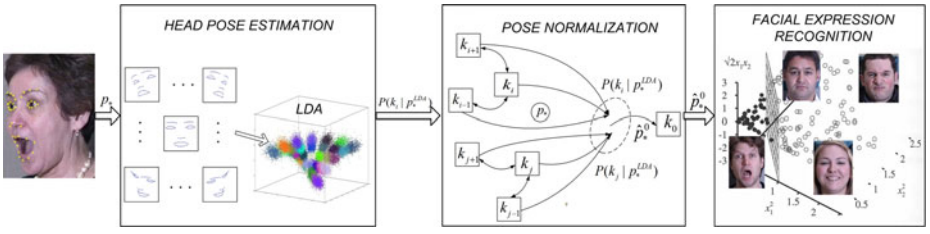

Fig. 1. The overview of the proposed three-step approach. Legend: p_* are the 2D locations of facial landmarks from the input facial image, $P(k_i|p_*^{LDA})$ is the probability of p_* belonging to the pose k_i, where k_0 is the frontal pose. The bidirectional lines in the pose normalization step represent the coupled head poses, and the directed lines represent the CGPR models learned per pair of poses (k_i, k_0). $\hat{p}_*^0$ is the prediction of p_* in the frontal pose obtained as a linearly weighted combination of the aforementioned CGPR models where the weights are proportional to $P(k_i|p_*^{LDA})$.

input datum (i.e. 2D facial points locations) to a low-dimensional manifold (attained by the means of multi-class LDA) and by estimating the probability of it belonging to each of the discrete poses for which training data are available. In the second step, we use the novel Coupled Gaussian Process Regression (CGPR) model to perform pose normalization, that is, to learn mappings between the 2D locations of landmark points of the facial expressions in non-frontal poses and their locations in the frontal pose. Instead of using single Gaussian Process Regression (GPR) model for all target pairs of poses at once, or using only one GPR model per target pair of poses, we propose CGPR models, which also model the couplings between the GPR models learned independently per target pairs of poses. To enable accurate performance for continuous head pose (i.e. for unknown poses), the predictions of the facial landmark locations in the frontal pose obtained by CGPR models from different poses are linearly combined (where the weights are based on head-pose probabilities obtained by the pose estimator in the first step of the proposed approach). The last step in our approach is facial expression classification in the frontal pose attained using the multi-class Support Vector Machine classifier.

The contributions of the proposed methodology are summarized as follows.

1. We propose a 2D face-shape-model-free approach to pose-invariant facial expression recognition that can handle expressive faces in the range from $-45°$ to $+45°$ pan rotation and from $-30°$ to $+30°$ tilt rotation. The proposed approach performs accurately for continuous head pose despite the fact that the training was conducted only on a set of discrete poses. It can also handle successfully the problem of having an unbalanced training dataset (i.e., when examples of certain facial expression categories are not included in the training dataset for a given discrete pose).
2. We propose a novel head pose normalization approach based on the linearly weighted combination of the newly proposed Coupled Gaussian Process Regression (CGPR) models, which model the couplings between the GPR

models learned per target pairs of poses. We employ GPR model since it provides not only the predictions of the facial landmark points in the frontal pose but also the uncertainty in these predictions (obtained through its covariance function) [12]. Moreover, the couplings between the GPR models can be embedded in their covariance structure in a very natural and straightforward manner. Although CGPR is a multiple-output GPR model, it does not model the dependences between its outputs (as done by the dependent-output GPR model such as the one proposed in [13]). Instead, CGPR models the dependences between the predictions obtained by different GPR models (i.e., GPR models learned for different poses). For these newly proposed CGPR models, we show experimentally that the proposed scheme outperforms a linearly weighted combination of GPR models learned per target pairs of poses which, in turn, outperforms baseline methods for pose normalization as 2D- and 3D-PDM.

The rest of the paper is organized as follows. In Section 2 we present our approach to pose-invariant facial expression recognition. In Section 3 we describe the newly proposed CGPR model. Experimental studies are discussed in Section 4, while Section 5 concludes the paper.

2 Pose-Invariant Facial Expression Recognition

In this section we describe a novel 2D face-shape-model-free approach to pose-invariant facial expression recognition given the 2D locations $p \in \mathbf{R}^d$ of $L = d/2$ facial landmarks of a face at an arbitrary pose. The proposed approach consists of three main steps: (i) head pose estimation by using a pose classifier on p, (ii) pose normalization by mapping the positions p of the facial landmarks from a non-frontal pose to the corresponding 2D positions p^0 in the frontal pose, and (iii) facial expression classification in the frontal pose. These steps are described in detail in the following sections and are summarized in Alg. 1. The theory behind the second step, that is the proposed CGPR model, is described in detail in Section 3.

In what follows, we assume that we have training data for each of P discrete poses and the correspondences between the points for each target pair of poses (non-frontal and frontal pose). In our case, we discretized the head pose space which resulted in $P = 35$ poses evenly distributed across the range from $-45°$ to $+45°$ pan rotation and from $-30°$ to $+30°$ tilt rotation. We denote by $D^k = \{p_1^k, ..., p_{N_k}^k\}$ the data set from pose k, and by $D = \left\{D^0, ..., D^k, ..., D^{P-1}\right\}$ the whole training data set, where N_k represents the number of training data in the pose k.

2.1 Head Pose Estimation

Various head pose estimation methods based on appearance and/or geometric features are proposed in the literature [14]. We propose to estimate the probability of each head pose belonging to a discretized head-pose space represented by

a low-dimensional manifold attained by means of multi-class LDA. Firstly, we normalize all examples from D (2D locations of facial landmarks in P poses), to remove the scale and translation components, as described in [15]. Secondly, to learn the manifold from such normalized data, we employ multi-class LDA since it is a simple linear transform that, given a training set with known pose labels, finds a low dimensional manifold which best represents pose variations while ignoring variations due to facial expressions. The estimated probability of input 2D facial points locations p being in pose k is given by $P(p^{lda}|k) = G(p^{lda}; \mu^k, \Sigma^k)$, where G is a normal density centered at μ^k and with covariance Σ^k. p^{lda} is the projection of p onto the low dimensional manifold. By applying Bayes' rule, we obtain the probability of being in pose k as $P(k|p^{lda}) \propto P(p^{lda}|k)P(k)$, where we assume a uniform prior $P(k) = 1/P$.

2.2 Head Pose Normalization

Given input data p_* containing the 2D locations of the facial points in an unknown head pose, our goal is to predict the location of these points in the frontal pose $\hat{p}_*^0$. To this end, we learn the functions $f_C^{(k)}(p_*)$ $(1 \leq k \leq P)$ which are later used to make predictions for input data p_*. These functions are modeled by the proposed CGPR models described in detail in Section 3. Thus, given p_*, $P(k|p_*^{lda})$ and $f_C^{(k)}(p_*)$, we obtain the locations of the frontal facial landmarks $\hat{p}_*^0$ as a linearly weighted combination of $f_C^{(k)}(p_*)$ for all k which satisfy $P(k|p_*^{lda}) > P_{min}$, where the weights are proportional to the head pose probabilities $P(k|p_*^{lda})$. The mathematical formulation of this is given in Step 2 in Alg. 1. Let us mention here that before $f_C^{(k)}$ is applied to p_*, it is registered to a reference face in pose k using a simple affine transform. The latter is calculated using five referential points: the nasal spine point and the inner and outer corners of the eyes (because they are stable facial points and the contractions of the facial muscles do not affect them).

2.3 Facial Expression Classification in Frontal Pose

We address the problem of pose-invariant facial expression recognition by performing pose normalization first, and subsequently applying any 2D-geometric-feature-based facial expression recognition method to the normalized input data (see [1]). In this paper, we use the multi-class SVM with decision function is given by

$$l = \arg\max_z \Big(\sum_{i:p_i^0 \in T_z} \alpha_i K(p_i^0, \hat{p}_*^0) + b_z \Big), \quad z = 1...Z, \tag{1}$$

where α_i and b_z are the weight and bias parameters, and $K(p_i^0, \hat{p}_*^0)$ is a vector of inner products between the training data $p_i^0 \in D^0$, containing Z facial expressions, and an estimate of p_* in the frontal pose, $\hat{p}_*^0$. The set T_z contains data points that depict facial expression z.

Algorithm 1. Pose-Invariant Facial Expression Recognition

Input: Positions of facial landmarks in an unknown pose (p_*)
Output: Facial expression label (l)
1. Apply the pose estimation (Sec. 2.1) to obtain $P(k|p_*^{lda})$, $k = 0 \,..\, P-1$
2. Register p_* to poses $k \in \mathcal{K}$ which satisfy $P(k|p_*^{lda}) > P_{min}$ (Sec. 2.2), and predict the locations of the facial landmarks points in frontal pose

$$\hat{p}_*^0 = \frac{1}{\sum_{k \in \mathcal{K}} P(k|p_*^{lda})} \sum_{k \in \mathcal{K}} P(k|p_*^{lda}) f_C^{(k)}(p_*^k)$$

3. Facial expression classification in frontal pose (Sec. 2.3)

$$l \leftarrow \arg\max_z \Big(\sum_{i:p_i^0 \in T_z} \alpha_i K(p_i^0, \hat{p}_*^0) + b_z \Big)$$

3 Coupled Gaussian Process Regression (CGPR)

In this section we describe a novel methodology for learning functions that map the 2D locations of facial points p in non-frontal poses to the corresponding 2D locations in the frontal pose. We learn a set of such functions, denoted by $f_C^{(k)}$, each one of which is associated with a certain pose k, where k is one of the discrete poses P for which training examples are available (i.e. $0 \leq k \leq P-1$). Roughly speaking, $f_C^{(k)}(p_*)$ is expected to provide good mappings for p_* obtained at an arbitrary pose that is relatively close to the pose k.

In order to learn $f_C^{(k)}$, we learn a set of $P-1$ mapping functions $\{f^{(1)}, .., f^{(P-1)}\}$ first, where the function $f^{(k)}$ maps the positions of the landmark points p^k in pose k to the corresponding points p^0 in the frontal pose. $f^{(k)}$ is learned using a GPR model for the target pair of poses $(k, 0)$ based on the datasets D^k and D^0, i.e., the sets that contain landmark points p in pose k and in the frontal pose denoted by 0.

3.1 Gaussian Process Regression (GPR)

In this section we describe the base GPR model for learning the mapping functions f^k. Formally, given a set of N_k examples of facial images containing the landmark locations in pose k, and the corresponding landmark locations in the frontal pose 0 (i.e. $\{D^k, D^0\}$), we learn the function $f^{(k)}$: $\mathbb{R}^d \rightarrow \mathbb{R}^d$ that maps $p_i^k \in D^k$ to $p_i^0 \in D^0$, where $i = 1..N_k$. Assuming Gaussian noise ε_i with zero mean and covariance matrix $\sigma_n^2 I$, this is expressed by $p_i^0 = f^{(k)}(p_i^k) + \varepsilon_i$. In GPR model, a zero mean Gaussian process prior is placed over the function $f^{(k)}$, that is $f^{(k)} \sim GP(0, K + \sigma_n^2 I)$, where $K(D^k, D^k)$ denotes $N_k \times N_k$ matrix of the covariances evaluated at pairs (p_i^k, p_j^k) by applying the kernel

$$k(p_i^k, p_j^k) = \sigma_s^2 \exp(-\frac{1}{2}(p_i^k - p_j^k)^T W (p_i^k - p_j^k)) \; + \; \sigma_l p_i^k p_j^k \; + \; \sigma_b, \qquad (2)$$

where $i, j = 1..N_k$. σ_s and $W = diag(w_1, ..., w_d)$ are the parameters of the radial basis function with different length scales for each input dimension (each coordinate of each landmark point), σ_l is the process variance which controls

the scale of the output function $f^{(k)}$, and σ_b is the model bias. This kernel has been widely used due to its ability to handle both linear and non-linear data structures [16]. During inference, we obtain the predictive mean $f^{(k)}(p_*^k)$ and the corresponding variance $V^{(k)}(p_*^k)$ for a new input p_*^k as

$$f^{(k)}(p_*^k) = k_*^T(K + \sigma_n^2 I)^{-1} D^0 \tag{3}$$

$$V^{(k)}(p_*^k) = k(p_*^k, p_*^k) - k_*^T(K + \sigma_n^2 I)^{-1} k_* \tag{4}$$

where $k_* = k(D^k, p_*^k)$, and $k(\cdot,\cdot)$ is given by Eq.(2). The kernel parameters $\theta = \{\sigma_s, W, \sigma_l, \sigma_b, \sigma_n\}$ are found by maximizing the log marginal likelihood of the training outputs using the conjugate gradient algorithm [12]. We assume here that the output dimensions (each coordinate of each landmark point in p_i^0) are *a priori* identically distributed [12]. This allows us to easily handle multiple outputs by applying the same covariance matrix to each output.

3.2 Learning Couplings

The mapping functions $\{f^{(1)}, ..., f^{(k)}, ..., f^{(P-1)}\}$ are learned independently for each target pair of poses; however, they need not be independent. Moreover, if the outputs obtained by different mapping functions are correlated, inferring the couplings between them may help obtain better predictions [17]. We model the coupling between two functions, $f^{(k_1)}$ and $f^{(k_2)}$, for pose k_1, using Gaussian distribution on the differences of their predictions obtained by evaluating these functions on the training data D^{k_1}. It is expressed by

$$P(f^{(k_1)}, f^{(k_2)}|k_1) \propto \exp(-\frac{1}{2} d^T \Sigma^{-1} d), \tag{5}$$

where $d = f^{(k_1)}(p_*^{k_1}) - f^{(k_2)}(p_*^{k_1})$, and $\Sigma = \sigma^2_{(k_1,k_2)} I$. The variance $\sigma^2_{(k_1,k_2)}$ measures the extent to which $f^{(k_2)}$ is coupled (i.e., similar) to $f^{(k_1)}$. Alternatively, this can be seen as an independent noise component in the predictions made by $f^{(k2)}$ because it is evaluated on data from different pose, i.e., pose k_1. Since we assume that this noise is Gaussian and independent of the noise process modeled by $f^{(k_2)}$, their covariances can simply be added [12]. Accordingly, we update the mean and variance given by Eq.(3) and Eq.(4), respectively, to obtain the mean and variance of CGPR model, that is

$$f^{(k_1,k_2)}(p_*^{k_1}) = k_*^T(K_2 + (\sigma_n^2 + \sigma^2_{(k_1,k_2)})I)^{-1} D^0 \tag{6}$$

$$V^{(k_1,k_2)}(p_*^{k_1}) = k(p_*^{k_1}, p_*^{k_1}) - k_*^T(K_2 + (\sigma_n^2 + \sigma^2_{(k_1,k_2)})I)^{-1} k_*, \tag{7}$$

where $k_* = k(D^{k_2}, p_*^{k_1})$. It is clear that the smaller the coupling between the functions $f^{(k_1)}$ and $f^{(k_2)}$, the higher the uncertainty in the predictions obtained by $f^{(k_1,k_2)}$. In the case of perfect coupling (when $\sigma^2_{(k_1,k_2)} \to 0$), we do not increase the uncertainty in the predictions obtained by $f^{(k_1,k_2)}$ (which converges to $f^{(k_2)}$).

On the other hand, when there is no coupling ($\sigma^2_{(k_1,k_2)} \to \infty$), we obtain the prior mean and covariance of $f^{(k_2)}$ ($f^{(k_1,k_2)}(p_*^{k_1}) \to 0$ and $V^{(k_1,k_2)}(p_*^{k_1}) \to k(p_*^{k_1}, p_*^{k_1})$). Because the variance of such prediction is the highest possible (learned by the model), this prediction will be suppressed by the covariance intersection rule described in Sec 3.4. Finally, the covariance matrix of CGPR model is guaranteed to be positive definite (the covariance matrix of the base GPR models learned in Sec. 3.1 is positive definite) since we only add a positive term to the diagonal of the covariance matrix in Eq.(6)&(7) [12].

3.3 Pruning CGPR Models

The couplings between all functions pairs $(f^{(k_1)}, f^{(k_2)})$ can be easily learned. Nevertheless, inference utilizing all them would be slow. Also, not all of the coupled functions $f^{(k_1,k_2)}$ contribute significantly in reducing the uncertainty in the predictions. As a pruning criterion, we propose using a measure based on the number of effective degrees of freedom of a GP [18]. In the framework of CGPR model, it has the following form

$$C_{eff}^{(k_1,k_2)} = \sum_{i=1}^{N_{k_2}} \frac{\lambda_i}{\lambda_i + \sigma_n^2 + \sigma^2_{(k_1,k_2)}} \tag{8}$$

where λ_i are the eigenvalues of the matrix K_2. If $\sigma^2_{(k_1,k_2)}$ is large, $C_{eff}^{(k_1,k_2)} \to 0$ and the predictions made by $f^{(k_1,k_2)}$ can be neglected. We compare the ratio $C_{eff}^{(k_1,k_2)}/C_{eff}^{(k_1)}$ to a threshold C_{min} to decide which coupled functions are relevant.

3.4 Covariance Intersection (CI)

In this section we describe how to fuse the predictions obtained by different mapping functions $f^{(k_1)}$ and $f^{(k_1,k_2)}$ in order to obtain a single prediction $f_C^{(k_1)}$ associated with pose k_1. A straightforward solution would be to select weighting functions inversely proportional to the variance of the predictions obtained by the individual functions. However, this fusion rule is optimal only if the predictions (i.e. their errors) are uncorrelated [17]. Since for a query point p_* we do not *a priori* know whether the predictions are correlated or not, the above fusion rule may not be optimal. Recently, a fusion rule, called Covariance Intersection (CI), for combining predictions in the presence of unknown cross covariance, has been proposed in [19]. To illustrate this, consider two GPR models, $f^{(k_1)}$ and $f^{(k_1,k_2)}$, with the mean and covariance pairs, $\{f^{(k_1)}(p_*), V^{(k_1)}(p_*)\}$ and $\{f^{(k_1,k_2)}(p_*), V^{(k_1,k_2)}(p_*)\}$. The CI yields the mean and covariance pair $\{f_C^{(k_1)}(p_*), V_C^{(k_1)}(p_*)\}$ obtained as

$$V_C^{(k_1)^{-1}}(p_*) = \omega(V^{(k_1)}(p_*))^{-1} + (1-\omega)(V^{(k_1,k_2)}(p_*))^{-1} \tag{9}$$

$$f_C^{(k_1)}(p_*) = V_C^{k_1}(p_*)(\omega(V^{(k_1)}(p_*))^{-1} f^{(k_1)}(p_*) + (1-\omega)(V^{(k_1,k_2)}(p_*))^{-1} f^{(k_1,k_2)}(p_*)) \tag{10}$$

where $\omega \in [0, 1]$ is a scalar that minimizes some criterion of uncertainty. In all our experiments we minimize the trace of $V_C^{k_1}(p_*)$ that we use as the uncertainty criterion, as proposed in [19].

4 Experiments

The experimental evaluation of the proposed methodology has been carried out using two datasets: the BU-3D Facial Expression (BU3DFE) database [20] containing 3D range data and the CMU Pose, Illumination and Expression Database (MultiPie) [21] containing multi-view facial expression data. BU3DFE contains 3D scans of 7 facial expressions, Angry, Disgust, Fear, Happy, Sad, Surprise and Neutral, performed by 100 subjects (60% female of various ethnic origin). All facial expressions except Neutral were sampled in four different levels of intensity. We generate 2D multi-view images of facial expressions from the available 3D data by rotating 39 facial landmark points provided by the database creators (see Fig. 3), which were used further as the features in our study. The data in our experiments include images of 50 subjects (54% female) at $\pm 15°$,$\pm 30°$ and $\pm 45°$ pan angles, and $\pm 15°$ and $\pm 30°$ tilt angles (see Fig. 1), with 5° increment, resulting in 1250 images for each of 247 poses. The training data are subsampled from this dataset to include images of expressive faces in 35 poses (15° increment in pan and tilt angles). These data (referred to as BU-TR dataset in the text below) as well as the rest of the data (referred to as BU-TST dataset and used to test the performance of the proposed methods) were partitioned into five folds in a person-independent manner for use in a 5-fold cross validation procedure. To evaluate the performance of the method in case of real data (as opposed to synthetic BU-TR/TST data), we used a subset of MultiPie containing images of 50 subjects (22% female) displaying 4 expressions (neutral, disgust, surprise, and happy) captured at 4 pan angles ($0°$,$-15°$,$-30°$ and $-45°$), resulting in 200 images per pose. All images were hand labeled in terms of 39 landmark points and the dataset was partitioned in a person-independent manner for use in a 5-fold cross validation procedure.

The rest of this section is organized as follows. First we present the experiments aimed at evaluation of the accuracy of the proposed head pose normalization method. To measure the accuracy of the method, we used the root-mean squared (RMSE) distance between the predicted image positions of the facial landmarks in the frontal pose and the ground truth (the manually annotated facial landmarks in frontal pose). As suggested by the results attained when testing on BU-TST dataset (see Fig. 2), the proposed CGPR-based method outperforms both GPR-based method and the 'baseline' methods for pose normalization, namely, 2D-PDM [22] and 3D-PDM [7]. The superior performance of the proposed CGPR-based method is also shown in the case of noisy data (see Table 1). Secondly, we evaluate the performance of the proposed pose-invariant facial expression recognition method. Testing was performed on faces from BU-TST images in (i) frontal pose (FP), (ii) non-frontal training poses (tp), and (iii) unknown poses (ntp), where the pose normalization was achieved using

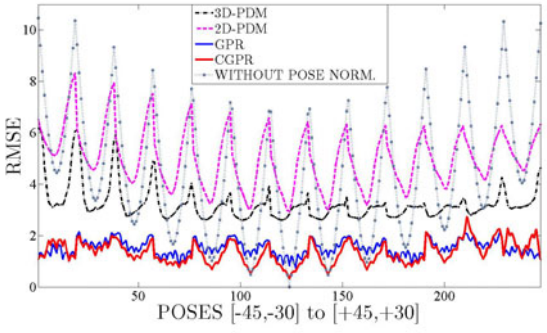

Fig. 2. Comparison of head pose normalization methods CGPR, GPR, 3D-PDM and 2D-PDM, trained on BU-TR (35 head poses) and tested on BU-TST (247 head poses) in a person-independent manner in terms of RMSE

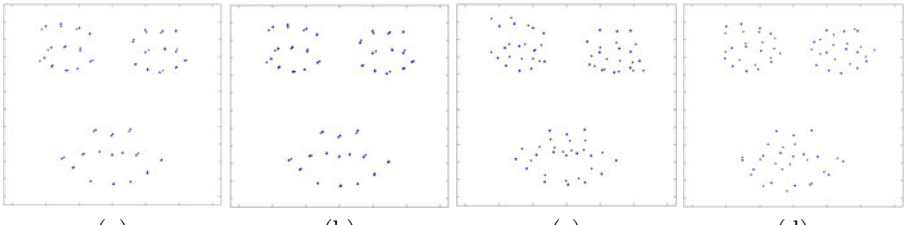

Fig. 3. Prediction of the facial landmarks in the frontal pose for an BU3DFE image of Happy facial expression in pose $(-45°, -30°)$ obtained by using (a) CGPR (b) GPR (c) 3D-PDM and (d) 2D-PDM. The blue $\bigtriangledown$ represent the ground truth and the black $\square$ are the predicted points. As can be seen, the alignment of the predicted and the corresponding ground truth facial landmarks is far from perfect in case of 3D/2D-PDM.

the CGPR-based method (Table 2). Finally, to evaluate the performance of the method in case of real data and in case of unbalanced data (i.e. when the method is trained on data where some of facial expression categories are missing in certain poses), we carry out experiments on MultiPie dataset (Table 3). For all experiments carried out on BU-TR/TST datasets, we did the following: the head pose estimator was trained on BU-TR dataset and when tested on BU-TST data, it predicted the correct (closest) head pose in 95% of cases. The base GPR models in Alg. 2 were trained on BU-TR dataset for each of the 34 target pairs of poses. Furthermore, we set P_{min} in Alg. 1 and C_{min} in Alg. 2 to experimentally found optimal values that are 0.1 and 0.75, respectively. The 2D-PDM and 3D-PDM were trained using the frontal data from BU-TR dataset (for 3D-PDM, the corresponding 3D data were used), retaining 15 modes of non-rigid shape variation.

Table 1. Comparison of head pose normalization methods CGPR, GPR, 3D-PDM and 2D-PDM, trained on BU-TR (35 head poses) and tested on BU-TST (247 head poses) corrupted with different levels of Gaussian noise with standard deviation σ, in a person-independent manner, in terms of RMSE

	$\sigma = 0$		$\sigma = 0.5$		$\sigma = 1$		$\sigma = 2$		$\sigma = 3$	
	tp	ntp	tp	ntp	tp	ntp	tp	ntp	tp	ntp
3D-PDM	3.1±0.9	3.2±0.6	3.1±0.8	2.9±0.6	3.2±0.7	2.9±0.6	3.7±0.9	3.3±0.6	4.0±0.5	3.8±0.5
2D-PDM	5.2±1.2	4.9±1.1	5.4±1.4	5.3±1.3	5.7±1.4	5.4±1.3	6.0±1.3	5.8±1.3	6.3±1.2	6.1±1.1
GPR	1.1±0.2	1.6±0.3	1.3±0.3	1.5±0.3	1.6±0.2	1.8±0.3	2.4±0.2	2.5±0.2	3.3±0.1	3.4±0.1
CGPR	1.1±0.3	1.4±0.4	1.2±0.3	1.4±0.2	1.5±0.2	1.6±0.2	2.3±0.2	2.4±0.2	3.2±0.2	3.3±0.1

Table 2. Facial expression recognition results using 7-class-SVM trained on frontal-pose expressive images from BU-TR and tested on BU-TST images in (i) frontal pose (FP), (ii) non-frontal training poses (tp), and (iii) unknown poses (ntp), where the pose normalization was achieved using the CGPR-based method. The best results reported by Hu *et al.* [11] for BU3DFE are reported for comparison purposes. All results are given in terms of correct recognition rate percentages.

	Disgust	Angry	Fear	Happy	Sad	Surprise	Neutral
FP+SVM	74.5±2.1	69.9±1.8	58.3±1.2	80.4±2.1	76.3±2.0	91.1±1.4	73±2.5
CGPR+SVM (tp)	71.0±3.1	72.8±1.6	58.0±1.7	81.9±2.9	73.8±2.7	89.9±1.9	73±3.0
CGPR+SVM (ntp)	70.1±3.4	71.1±2.2	56.2±2.2	80.2±1.8	72.1±2.9	88.1±2.0	72±2.4
Hu *et al.* [11] (tp)	69.3	71.3	52.5	78.3	71.5	86.0	-

Evaluation of the accuracy of the proposed head pose normalization method – Fig. 2 shows the comparative results in terms of RMSE of the tested head pose normalization methods along with the results obtained when no pose normalization is performed and only the translation component has been removed. As can be seen, both GPR- and CGPR-based methods significantly outperform the 2D/3D 'baseline' methods for pose normalization. Judging from Fig 3, this is probably due to the fact that the tested 2D/3D deformable face-shape-based models were not able to accurately model the non-rigid facial movements present in facial expression data. The performance of the aforementioned models in the presence of noise in test data was evaluated on BU-TST data corrupted by adding four different levels of noise. As can be seen from Table 1, even in the presence of high levels of noise the performance of GPR/CGPR-based methods is comparable to that of 2D/3D-PDM achieved for noise-free data. The performance of GPR- and CGPR-based methods is highly comparable in the aforementioned experiments where the utilized data were balanced (i.e. when the method is trained on data containing examples of all facial expression categories in all target poses). However, the results shown in Table 1 (i.e. when no noise is present in the data) suggest that the proposed CGPR-based method slightly outperforms the GPR-based method when tested on unknown poses (ntp).

Table 3. Facial expression recognition results using 4-class-SVM trained and tested on unbalanced data from MultiPie, where the pose normalization was achieved using GPR- or CGPR-based method. The unbalanced dataset was prepared by removing all examples of one facial expression category from one of the non-frontal poses. The testing was performed on the removed examples in a cross-validation person-independent manner for each expression and each pose. The performance of the classifier trained/tested on frontal-pose expressive images from MultiPie is also reported for the purposes of baseline comparison. All results are given in terms of correct recognition rate percentages.

		Disgust	Happy	Surprise	Neutral
FP+SVM	RR[%]	94.2±2.3	95.6±1.3	97.4±0.9	93.7±1.9
GPR+SVM	RR[%]	68.9±6.2	74.2±4.5	69.4±5.2	73.8±3.9
	RMSE	3.10±0.7	3.21±0.9	3.62±0.9	2.80±0.7
CGPR+SVM	RR[%]	85.2±4.3	90.2±3.1	89.8±3.2	88.2±3.1
	RMSE	1.95±0.3	1.80±0.4	2.40±0.3	1.90±0.4

Algorithm 2. Learning and inference with CGPR

OFFLINE: Learning base GPR models and coupling parameters
1. Learn $P-1$ base GPR models $\{f^{(1)}, ..., f^{(P-1)}\}$ for target pairs of poses (Sec. 3.1)
2. Perform coupling of base GPR models learned in Step 1
  for k_1=1 to P-1 **do**
    for k_2=1 to P-1 & $k_1 \neq k_2$ **do**
      estimate $\sigma_{(k_1,k_2)}$ (Sec. 3.3)
      if $C_{eff}^{(k_1,k_2)} > C_{min}$ **then** $\sigma_C^{k_1} = [\sigma_C^{k_1}, \sigma_{(k_1,k_2)}]$ **end if**
    end for
    store $\sigma_C^{k_1}$
  end for

ONLINE: Inference of the facial landmarks $p_*^{k_1}$ in pose k_1
S_{k_1}: number of the functions coupled to $f^{(k_1)}$
1. Evaluate base GPR model for pose k_1 (Sec. 3.1): $Pr(0) = \{f^{(k_1)}(p_*^{k_1}), V^{(k_1)}(p_*^{k_1})\}$
2. Evaluate CGPR models for pose k_1 (Sec. 3.3)
  for i=1 to S_{k_1} **do** $\sigma_{(k_1,i)} = \sigma_C^{k_1}(i)$, $Pr(i) = \{f^{(k_1,i)}(p_*^{k_1}), V^{(k_1,i)}(p_*^{k_1})\}$ **end for**
3. Combine estimates using CI (Sec. 3.4): $\{f_C^{(k_1)}(p_*^{k_1}), V_C^{(k_1)}(p_*^{k_1})\} = CI(Pr)$

Evaluation of the proposed pose-invariant facial expression recognition method – The results presented in Table 2 clearly suggest that the proposed pose-invariant facial expression recognition method performs accurately for continuous head pose (i.e. for unknown poses; ntp-case in the Table 2) despite the fact that the training was conducted only on a set of discrete poses (i.e. on BU-TR). As can be seen further from Table 2, even in case of unknown poses, the proposed method outperforms the method reported by Hu *et al.* [11], where pose-wise SVM classifiers were trained and tested only on known poses. While the aforementioned experiments suggest that the performance of GPR- and CGPR-based methods is highly comparable when the utilized data are balanced, the same is not the case when the utilized data are unbalanced.

Specifically, the results presented in Table 3 clearly suggest that the proposed CGPR-based pose-invariant facial expression recognition method significantly outperforms the GPR-based method in case of unbalanced data, i.e., when one facial expression category is missing in a certain pose. Judging from Table 3, RMSE rows, the reason for this is that the CGPR-based head pose normalization is significantly better than that obtained by the GPR-based method. In turn, this can be explained by the non-parametric nature of the GPR-based method due to which it cannot generalize well beyond the training data. On the contrary, the CGPR-based method overcomes this by employing the knowledge (training data) provided by the underlying CGPR models.

5 Conclusion

We presented a novel 2D-shape-free method for the recognition of facial expressions at arbitrary poses that is based on pose normalization of 2D geometric features. For pose normalization, we proposed Coupled Gaussian Process Regression (CGPR) model that learns direct mappings between the facial positions at an arbitrary pose and the positions in the frontal pose. Experimental results demonstrate the advantages of the proposed pose normalization in comparison to generative methods and its robustness to incomplete training data (i.e. expressions and poses that do not belong to the training dataset). For the problem of expression recognition, the proposed method is shown to demonstrate classification performance comparable to the ones obtained by pose-specific classification schemes for the significantly more difficult problem of expression recognition at an unknown pose.

Acknowledgments. This work is funded in part by the European Community's 7th Framework Programme [FP7/2007-2013] under grant agreement no. 211486 (SEMAINE), and in part by the European Research Council under the ERC Starting Grant agreement no. ERC-2007-StG-203143 (MAHNOB). The work of Ioannis Patras is partially supported by EPSRC project EP/G033935/1.

References

1. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Trans. Pattern Analysis and Machine Intelligence 31, 39–58 (2009)
2. Vinciarelli, A., Pantic, M., Bourlard, H.: Social signal processing: Survey of an emerging domain. Image and Vision Computing 27, 1743–1759 (2009)
3. Chang, Y., Vieira, M., Turk, M., Velho, L.: Automatic 3d facial expression analysis in videos. In: Proc. Int'l Workshop Analysis and Modelling of Faces and Gestures, pp. 293–307 (2005)
4. Sun, Y., Yin, L.: Facial expression recognition based on 3d dynamic range model sequences. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 58–71. Springer, Heidelberg (2008)

5. Sung, J., Kim, D.: Real-time facial expression recognition using staam and layered gda classifier. Image and Vision Computing 27, 1313–1325 (2009)
6. Cheon, Y., Kim, D.: Natural facial expression recognition using differential-aam and manifold learning. Pattern Recognition 42, 1340–1350 (2009)
7. Zhu, Z., Ji, Q.: Robust real-time face pose and facial expression recovery. In: Proc. Int'l Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 681–688 (2006)
8. Wang, T.H., Lien, J.J.J.: Facial expression recognition system based on rigid and non-rigid motion separation and 3d pose estimation. Pattern Recognition 42, 962–977 (2009)
9. Kumano, S., Otsuka, K., Yamato, J., Maeda, E., Sato, Y.: Pose-invariant facial expression recognition using variable-intensity templates. Int'l J. Computer Vision 83, 178–194 (2009)
10. Chai, X., Shan, S., Chen, X., Gao, W.: Locally linear regression for pose-invariant face recognition. IEEE Trans. Image Processing 16, 1716–1725 (2007)
11. Hu, Y., Zeng, Z., Yin, L., Wei, X., Tu, J., Huang, T.: A study of non-frontal-view facial expressions recognition. In: Proc. Int'l Conf. Pattern Recognition, pp. 1–4 (2008)
12. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). The MIT Press, Cambridge (2005)
13. Boyle, P., Frean, M.: Dependent gaussian processes. In: Advances in Neural Information Processing Systems, vol. 17, pp. 217–224. MIT Press, Cambridge (2005)
14. Murphy-Chutorian, E., Trivedi, M.M.: Head pose estimation in computer vision: A survey. IEEE Trans. Pattern Analysis and Machine Intelligence 31, 607–626 (2009)
15. Rudovic, O., Patras, I., Pantic, M.: Facial expression invariant head pose normalization using gaussian process regression. In: Proceedings of IEEE Int'l Conf. Computer Vision and Pattern Recognition, vol. 3 (in Press, 2010)
16. Chen, T., Morris, J., Martin, E.: Gaussian process regression for multivariate spectroscopic calibration. Chemometrics and Intelligent Laboratory Systems 87, 59–71 (2007)
17. Tresp, V., Taniguchi, M.: Combining estimators using non-constant weighting functions. In: Advances in Neural Information Processing Systems, pp. 419–426 (1995)
18. Tresp, V.: A bayesian committee machine. Neural Computing 12, 2719–2741 (2000)
19. Julier, S.J., Uhlmann, J.K.: A non-divergent estimation algorithm in the presence of unknown correlations. In: Proc. American Control Conf., pp. 2369–2373 (1997)
20. Wang, J., Yin, L., Wei, X., Sun, Y.: 3d facial expression recognition based on primitive surface feature distribution. In: Proc. Int'l Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 1399–1406 (2006)
21. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. Image and Vision Computing 28, 807–813 (2010)
22. Cootes, T., Taylor, C.: Active shape models - smart snakes. In: Proc. British Machine Vision Conf., pp. 266–275 (1992)