Cross-pose Facial Expression Recognition

Abstract— In real world facial expression recognition (FER) applications, it is not practical for a user to enroll his/her facial expressions under different pose angles. Therefore, a desirable property of a FER system would be to allow the user to enroll his/her facial expressions under a single pose, for example frontal, and be able to recognize them under different pose angles. In this paper, we address this problem and present a method to recognize six prototypic facial expressions of an individual across different pose angles. We use Partial Least Squares to map the expressions from different poses into a common subspace, in which covariance between them is maximized. We show that PLS can be effectively used for facial expression recognition across poses by training on coupled expressions of the same identity from two different poses. This way of training lets the learned bases model the differences between expressions of different poses by excluding the effect of the identity. We have evaluated the proposed approach on the BU3DFE database [1]. We experiment with intensity values and Gabor filters for local face representation. We demonstrate that two representations perform similarly in case frontal is the input pose, but Gabor outperforms intensity for other pose pairs. We also perform a detailed analysis of the parameters used in the experiments. We have shown that it is possible to successfully recognize expressions of an individual from arbitrary viewpoints by only having his/her expressions from a single pose, for example frontal pose as the most practical case. Especially, if the difference in view angle is relatively small, that is less than 30 degrees, then the accuracy is over 90%. The correct recognition rate is often around 99% if there is only 15 degrees difference between view angles of the matched faces. Overall, we achieved an average recognition rate of 87.6% when using frontal images as gallery and 86.6% when considering all pose pairs.

I. INTRODUCTION

Facial expressions constitute an essential part of non-verbal communication between human beings. Automated analysis of facial expressions has been an active research topic in computer vision over the last years. Most of the facial expression recognition studies attempt to recognize a set of prototypic emotional expressions, that is happiness, surprise, anger, sadness, fear, and disgust.

Automatic recognition of expressions can be a very challenging task, since there are many underlying factors that affect the appearance of facial expressions. One factor is the presence of subject differences such as texture of the skin, hair style, age, gender, and ethnicity. In addition to the differences in appearance, there might be differences in expressiveness; that is, individuals perform expressions differently from each other [2]. Another factor that makes automatic facial expression recognition a hard problem is the presence of pose variations. The change in pose causes a non-linear transformation of the 2D face image. Moreover, some areas of the face become self-occluded and some areas might

have a very different appearance from different viewpoints. Variations caused by different poses impose extra burden on the task of recognizing expressions, which is already a difficult problem due to the large differences across subjects.

Most of the existing studies focus on recognizing expressions from frontal or nearly frontal view facial images. However, expression recognition systems have to deal with arbitrary viewpoints to be able to work under the uncontrolled real world situations. Most of the existing multi-view studies on expression recognition discretize the viewpoints into a set of intervals and use a separate model for each viewpoint. Each model functions as a recognizer for a particular pose angle and needs representative data from that pose angle. In [3], Hu *et al.* train different classifiers for each pose to compare the performance of non-frontal view classifiers with the frontal view ones. They obtain higher results for non-frontal views. In a different study, [4], they utilize two classification schemes. The first one is a 2-step cascade classification, they first train a pose classifier, then separate expression classifiers are trained for each pose. The second one is a composite classification which treats each pose-emotion combination as a class. They compare different feature descriptors and dimensionality reduction techniques for these schemes. Similar to their first scheme, Moore and Bowden utilize a two-step classification approach in [5] with different forms of Local Binary Patterns (LBP). In [6], Rudovic *et al.* learn a mapping of facial landmarks, like mouth corner, from non-frontal to frontal views, so that a frontal classifier can also be used for non-frontal views. They compare a number of regression models for this mapping.

A universal multi-view facial expression recognition system may not generalize well and perform robustly while classifying facial expressions of an unseen person. Therefore, learning person specific facial expressions would be desirable. However, it is not convenient for a user to enroll his/her facial expressions under different pose angles. In this respect, a FER system should allow the user to enroll his/her facial expressions under a single pose, for example frontal, and be able to recognize them under different pose angles. To achieve this goal, we model the relations between expressions of an individual from different viewpoints by learning a mapping from one pose to another. This way, having only frontal view expressions is sufficient to recognize expressions from any other arbitrary viewpoints. A FER system capable of relating expressions from different viewpoints has many interesting applications. Autism studies as a prevalent application area of FER, could highly benefit from such an improvement. Autistic children may have problems with directly looking at a camera and even if they

look, they may move their head very quickly. Due to these, for example in an emotion capturing game, it may be very problematic to make the child perform expressions from frontal view. In such cases, it would be very beneficial to take an expression from one pose and then, to be able to generalize this to other poses.

Recently, cross-pose face matching has become a popular research topic, especially for identification purposes. Partial Least Squares (PLS) has been found as a promising approach to model the relations between pose pairs [7], [8], [9]. In [7], Sharma *et al.* use PLS to find a latent space for each pair of poses. Recognition is then performed in this latent space using nearest neighbor matching. Similarly, Li *et al.* use PLS to produce pose-independent feature vectors in [8]. The authors report improved results by applying PLS on local blocks of Gabor features compared to the holistic representation. In [9], Fischer *et al.* perform a detailed analysis of PLS for face recognition by using both holistic and local representation methods.

The previous works indicate that the PLS approach achieves very good results for pose-invariant face recognition. Inspired by this outcome, in this paper, we explore the use of PLS for cross-pose facial expression recognition. We present a method to recognize six prototypic facial expressions of an individual across different pose angles. We use PLS to learn a relationship between faces of two different poses belonging to the same emotion and the same subject. The reason for using faces of the same subject is to exclude variations caused by identity and to reduce the problem to modeling of the variance in expressions caused by pose changes. We achieved an average recognition rate of 87.6% when using frontal images as gallery and 86.6% when considering all pose pairs on the highest intensity level of the BU3DFE database [1], which are significantly superior to the performance obtained by view-specific facial expression classifiers. Overall, in this study we adopt PLS and propose a framework to utilize it efficiently for cross-pose facial expression classification. We have shown that it is possible to successfully recognize expressions of an individual from arbitrary viewpoints by only having his/her expressions from a single pose, for example frontal pose as the most practical case, especially if the difference in view angle is relatively small, that is less than 30 degrees, then the accuracy is over 90%. The correct recognition rate is often around 99% if there is only 15 degrees difference between view angles of the matched faces.

The rest of the paper is organized as follows: the details of PLS are described in Section 2, and the proposed method is introduced in Section 3. Experimental results and discussions are given in Section 4; followed by conclusion in Section 5.

II. PARTIAL LEAST SQUARES (PLS)

PLS models the relations between blocks of observed variables by means of latent variables. Input and output vectors are mapped into a common vector space in such a way that covariance between projected input and output vectors is maximized [10]. Both input X and output Y are

considered as indicators of p latent variables, or scores, t and u, respectively. PLS maximizes the squares of covariance between the score vectors t and u by finding weight (basis) vectors w and c such that:

$$
\begin{aligned}
[cov(t, u)]^2 &= [cov(Xw, Yc)]^2 \qquad (1) \\
&= max_{|r|=|s|=1}[cov(Xr, Ys)]^2
\end{aligned}
$$

where $cov(t, u) = t^T u / n$ denotes the sample covariance between score vectors t and u.

We refer to representing X and Y by their corresponding score vectors t and u as projecting them to the latent space, where their samples are highly correlated. Weight vectors w and c are computed by the NIPALS algorithm [10] and saved into the projection matrices W and C, respectively. Then, input and output data can be projected into the latent space by using these projections: $\hat{x} = W^T x$ and $\hat{y} = C^T y$. After projection, facial expression recognition methods can be applied on these pose-independent latent vectors $\hat{x}$ and $\hat{y}$.

III. PROPOSED APPROACH

In this section, we explain the steps of the PLS approach which are alignment, feature extraction, PLS, and expression recognition.

A. Alignment

We utilize an alignment method which works for all pose angles, and gives a consistent scale and rotation of the face similar to [9]. Parameters of the alignment are illustrated in Fig. 1.
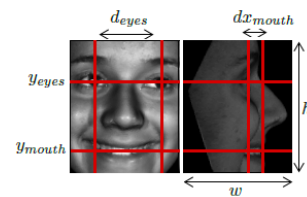

Fig. 1.   Illustration of the alignment parameters.

We first calculate the positions of the visible eyes, $s_{l.eye}$ and/or $s_{r.eye}$, and the mouth center s_{mouth} in the input image by using the annotated landmarks. Then, we compute a similarity transform, T by specifying two point correspondences between the input image and the aligned image.

The first point correspondence is always the mouth center. Let $x_{center} = \frac{(w-1)}{2}$ be the horizontal center of the aligned image and ϕ be the pose angle. The positions in the input and aligned image are computed as follows:

$$
\begin{aligned}
s_1 &= s_{mouth} & t_1 &= \begin{pmatrix} x_{center} + dx_{mouth}\ sin(\phi) \\ y_{mouth} \end{pmatrix} \\
s_2 &= \frac{s_{l.eye} + s_{r.eye}}{2} & t_2 &= \begin{pmatrix} x_{center} \\ y_{eyes} \end{pmatrix} \qquad (2)
\end{aligned}
$$

In the case that only one eye is visible, the visible eye $s_{v.eye}$ is used for the second correspondence. The homogeneous transformation matrix, T is computed by solving
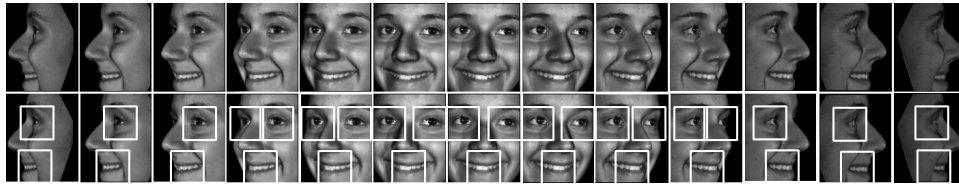
Fig. 2. Examples of aligned face images in different poses (first row) and visualization of the extracted local blocks on the aligned face (second row).

the system of linear equations given by the two point correspondences.

B. Feature Extraction

In facial image analysis, local face representations have shown better performance than holistic representations. In this study, we use a local face representation by either using intensity values or Gabor wavelets as features.

1) Gabor Features: Gabor wavelets have been extensively used for facial image analysis due to their powerful representation capabilities [11]. The conventional Gabor wavelet is defined as follows:

$$\psi(\vec{x}; \nu, \mu) = \frac{k_{\nu,\mu}^2}{\sigma^2} e^{(-\frac{k_{\nu,\mu}^2 ||\vec{x}||^2}{2\sigma^2})} [e^{(ik_{\nu,\mu}\vec{x})} - e^{(-\frac{\sigma^2}{2})}] \quad (3)$$

where μ and ν define the orientation and scale of the Gabor kernels. We perform full convolution of the Gabor wavelet with the aligned face image to obtain the Gabor Magnitude Images (GMI) in five different scales $\nu \in \{0, 1, 2, 3, 4\}$ and eight orientations $\mu \in \{0, 1, .., 7\}$.

2) Extraction of Local Blocks: We extract local blocks around facial landmarks similar to [8] and [9] from the representation of each face, which is gray-scale image in case of intensity features and GMIs in case of Gabor features. More specifically, we use local blocks around left eye, right eye and mouth since these regions provide the most discriminative information for facial expression recognition.

We extract blocks of size $w_b \times h_b$ centered on the eye centers and the mouth center. For frontal pose, local blocks can be extracted directly. On the other hand, in non-frontal poses, some of the directly extracted local blocks may include more and more of the background. In order to avoid background, we apply the following modifications as in [9]: First, we discard the eye block unless it is clearly seen in the non-frontal aligned face image. Secondly, we compute a horizontal offset Δ_{mouth} as in [9] for the mouth block. Then, we shift the mouth block horizontally and move it further into the face to decrease the number of background pixels in the block. The horizontal offset is computed as follows:

$$\Delta_{mouth} = -f_{mouth} \ w_{mouth} \ sin(\phi) \quad (4)$$

where f_{mouth} is the mouth coefficient, and ϕ is the pose angle of the face. Fig. 2 shows some examples of the blocks extracted on aligned face images.

C. PLS for Cross-pose Expression Recognition

For each pose pair, we use PLS to compute a latent space for the existing blocks in both poses. We use a custom GPU implementation of the NIPALS algorithm [10] to compute the PLS bases. The input and output vectors are centered by subtracting mean and scaled by dividing by standard deviation before training. Test data is transformed by using the values learned during training.

In training, for a pose pair (p_i, p_j), we construct input $\boldsymbol{X}$ and output $\boldsymbol{Y}$ matrices, where samples in $\boldsymbol{X}$ are from pose p_i and samples in $\boldsymbol{Y}$ are from pose p_j. Corresponding samples are coupled by both identity and expression. We then perform PLS to compute projections $\boldsymbol{W}$ and $\boldsymbol{C}$ that maximize the covariance of score vectors. Since training faces from two different poses are coupled by expression and identity, covariance between different poses of an expression is maximized. In testing, learned projections are used to estimate the score vectors for new samples and then classification methods for expression recognition can be applied on these pose-independent latent vectors.

In faces with large pose angles, one of the eyes is not visible. This causes a problem for the pose pairs with opposite signed and large angles. In one pose, only the right eye is visible and in the other only the left eye, therefore eyes cannot be used at all. To solve this problem, we assume that left and right eye are sufficiently symmetric and exploit this symmetry property of eyes as in [9]. We train a PLS latent space for the opposite eye blocks.

D. Facial Expression Recognition

We compute the distance between the query and target face images in different poses by first extracting the blocks for both images as explained in the previous sections. Then, for those pairs of blocks that have trained PLS for the pose pair, we project the blocks into the latent space and compute the distance between the latent vectors. The computed differences for each block are averaged to yield the global difference.

After projection, the dimension of the input and output score vectors is the same and equal to the number of extracted PLS bases. Therefore, latent representations of input and output vectors lie in the same vector space. Moreover, PLS bases are learned based on a criterion that maximizes the covariance between the score vectors. Therefore, we can compare projected input and output samples by using the Nearest Neighbor (NN) algorithm. To compute the distance between query-target pairs, we use L2-norm and normalized cross correlation (NCC) in our experiments. NCC is defined as follows:

$$d_{ncc}(\boldsymbol{x}, \boldsymbol{y}) = 1 - \frac{(\boldsymbol{x} - \mu_x) \ (\boldsymbol{y} - \mu_y)}{(N-1) \ \sigma_x \ \sigma_y} \quad (5)$$

IV. EXPERIMENTS AND RESULTS

A. Data

We evaluated our approach on a commonly used database for multi-view facial expression recognition, the Binghamton University 3D Facial Expression Database (BU-3DFE) [1]. BU-3DFE contains 3D models of 100 subjects with texture and 83 annotated landmark points per model. Each subject shows 7 expressions, which are neutral, anger, disgust, fear, happiness, sadness, and surprise. All subjects display all expressions except neutral at four different levels of intensity from low to high. For this study, 3D models are rendered together with the texture by rotating at yaw angle from -90 to $+90$ degrees in steps of 15 degrees.

B. Experimental Setup

Data taken from the BU-3DFE database is divided into three sets of similar size, each containing different sets of subjects. One set is used for learning PLS bases, one for the optimization of parameters, and the last one for testing. In our experiments, we use a single image for each expression, which belongs to the highest intensity level (level 4). Later, we also experiment with other intensity levels to show the effect of intensity level on the results.

We repeat alignment, feature extraction, and projection steps for each face image. Then, we match an expression of a subject from the input pose to the all expressions of that subject from the output pose, and assign the label by using the NN algorithm.

C. Effects of Parameters

There are a number of factors that affect the recognition results in our experiments. These factors can be organized as alignment parameters, feature extraction parameters, number of PLS bases in the NIPALS algorithm, and the distance type used in classification. We performed a series of experiments with changing parameter settings to show the effects of these parameters. We report results as the average of all pose pairs, where input pose is always the frontal pose.

Alignment is a common, critical step for facial image analysis problems. In our experiments, we used the optimal parameters used for face recognition in [9] for the alignment. These are, $w = 104$, $h = 128$, $y_{eyes} = 42$, $y_{mouth} = 106$, $d_{eyes} = 62$, and $dx_{mouth} = 20$. Example alignment results for all pose angles can be seen in Fig. 2.

For alignment, we calculate mouth and eye locations from annotated landmarks. In an automatic system, these three points have to be automatically extracted. In order to measure the robustness of our proposed approach on localization errors, we trained pose specific Active Appearance Models (AAM) [12] for each pose and then fit these models on face images to automatically extract the location of eyes and mouth. We performed experiments to compare the results obtained from annotated landmarks and estimated points by AAM. In these experiments, pose pairs are constituted between a pose angle and frontal view, where input pose is always the frontal. As can be seen from the Fig. 3, there is a slight decrease in the results, therefore effect of facial feature

localization errors on alignment and feature extraction is negligible.
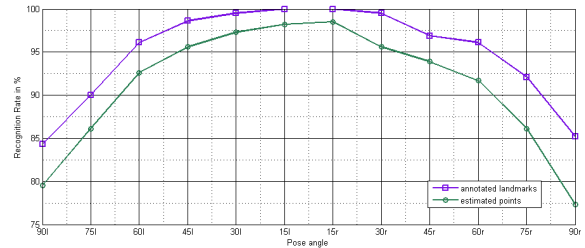

Fig. 3. Comparison of the results obtained from annotated landmarks and estimated points by AAM.

1) Effects of Feature Extraction Parameters: Parameters of the feature extraction are block size, mouth offset, and additional Gabor specific parameters in case of Gabor features.
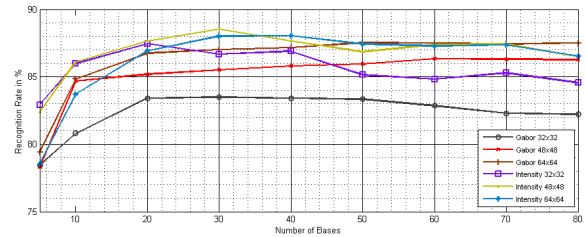

Fig. 4. Effects of different block sizes for intensity values and Gabor features with changing number of PLS bases on the average recognition rate.

We experiment with three different block sizes: 32×32, 48×48, and 64×64. Fig. 4 illustrates the effects of each block size and feature type on the recognition rates with changing number of PLS bases. Local blocks of intensity values and Gabor features produce similar results for two largest block sizes. As can be seen from the Fig. 4, block size of 32×32 gives the lowest results for both feature types. For Gabor features, the highest recognition rates are obtained by using the largest block size, 64×64. On the other hand, block sizes of 64×64 and 48×48 result in similar recognition rates in case of intensity features, and the highest result is achieved by using block size of 48×48 at 30 bases. The extracted blocks also cover some of the background together with the additional areas on the face in case of large block sizes. Since we have a uniform background in BU3DFE, Gabor features are not affected from that and produce the highest results with the largest block size by benefiting from the extra areas covered. However, direct intensity values of the background are zero, and this causes the largest block size to produce similar results with block size of 48×48.

To show the effects of the mouth offset parameter on the recognition rates, we experiment with a set of parameters. As can be seen from Fig. 5, location of the mouth has an important effect on recognition rates, especially for large pose angles. The highest recognition rates for many of the
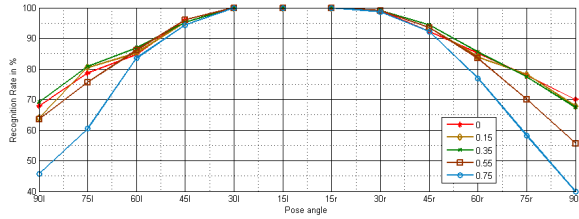
Fig. 5. Effects of the mouth offset parameter for all pose angles.

pose angles are achieved when the mouth offset parameter equals to 0.35. This value minimizes background pixels inside the mouth block and still contains the outline of the mouth.

Gabor representation produces similar recognition rates with intensity values. So far, we performed experiments by using the default values in the literature ($k_{max} = \pi/2$, $\sigma = 3\pi/2$), but these default values may not be the optimal ones for facial expression recognition. Therefore, we experimented with different parameters of Gabor wavelets to improve the representation. k_{max} is responsible for the scaling part, and consequently highly related to the size of the structures in the image. Gaussian window width σ finds a compensation between the representation of coarse and fine structures in the image. We perform a grid search for these parameters to find their optimal values. As shown in Fig. 6, optimal values, $k_{max} = \frac{\pi}{0.5}$ and $\sigma = 1.0\pi$, are different from the default values. This shows the importance of representing small and fine details of the face for expression recognition.
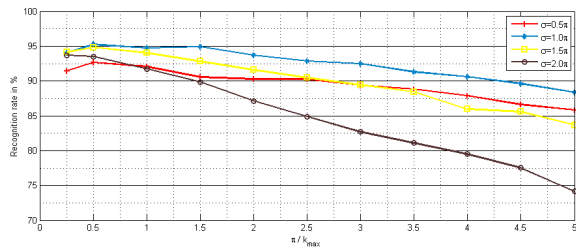

Fig. 6. Effects of different Gabor parameters, k_{max} and σ.

2) Effects of Number of PLS bases: To show the effect of the number of PLS bases, we use a set of PLS bases from 10 to 80 in our experiments. Fig. 4 and Fig. 7 show the influence of the number of bases in different experiments. From these experiments, we realize that the optimal recognition performance is usually achieved when 30 PLS bases are used.

3) Effects of the Distance Type: The results of two different distance types for both local blocks of Gabor features and intensity values can be seen from Fig. 7 with changing number of PLS bases. For both feature types, high recognition rates are achieved starting from 30 bases. There is no significant advantage of using one distance type over another.
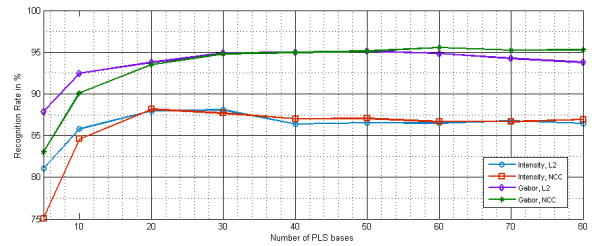

Fig. 7. Results of two different distance types for both local blocks of Gabor features and intensity values with changing number of PLS bases.

D. Cross-pose Recognition Results

We relate an expression image of a subject from one pose to another pose by using the PLS method. This relation shows how well an expression from a viewpoint can be recognized by matching expressions from another viewpoint, if we exclude the subject differences.

In this section, we evaluate our method by obtaining recognition rates for each pose pair. Results for each pose pair by using the intensity values of local blocks as features can be seen from Table I and Gabor features can be seen from Table II. These results show that expressions of a subject from different poses are projected into a space in which they remain closer to each other than other expressions of the subject despite the differences caused by the pose change. We also see that overall performance of Gabor features only outperform intensity features significantly when using non-frontal poses as gallery.

It is clear from the results in Tables I and II that pose pairs whose angles are close to each other are likely to produce higher recognition rates. Therefore, table elements are higher as they get close to the diagonal. If the difference in view angle is relatively small, that is less than 30 degrees, then the accuracy is over 90%. The correct recognition rate is often around 99% if there is only 15 degrees difference between view angles of the matched faces. Although tables are not exactly symmetric due to the stopping criterion in the NIPALS algorithm, recognition rates of the symmetric pose pairs are close to each other as expected as a consequence of the symmetric modeling of the input and output matrices in the algorithm.

E. Results for All Intensity Levels

There are four different intensity levels for each expression in BU3DFE. In this section, we repeat our experiments for each intensity level by averaging the results of using the frontal pose as the input pose. As shown in Table III, recognition rates are higher for higher intensity levels.

F. Comparison with the Previous Works

Most of the multi-view expression recognition studies in the literature train pose-specific classifiers and results are reported according to this scheme, therefore an exact comparison of our approach with the previous studies is not possible. As an example, we use the same experimental setup

TABLE I
RESULTS FOR ALL INPUT AND OUTPUT POSE PAIRS BY USING INTENSITY FEATURES.

g/p	90l	75l	60l	45l	30l	15l	0	15r	30r	45r	60r	75r	90r	Avg.
90l	-	97.7	91.8	75.4	65.6	58.0	50.8	52.1	51.7	50.8	54.6	53.8	51.4	62.8
75l	99.3	-	99.8	93.1	85.3	76.7	67.2	65.1	61.4	63.8	65.2	66.5	59.1	75.2
60l	94.9	99.1	-	98.1	93.8	88.0	80.7	74.3	71.0	70.1	72.9	69.7	62.5	81.2
45l	81.9	94.6	98.3	-	98.8	96.7	87.9	80.1	72.1	68.9	71.6	67.4	60.2	81.5
30l	75.1	89.0	95.6	99.7	-	98.9	95.8	88.3	77.7	72.2	74.5	66.9	57.1	82.5
15l	72.1	84.6	91.3	98.9	100	-	99.8	97.6	91.1	81.9	77.0	69.2	56.4	84.9
0	64.0	76.9	85.2	92.7	98.0	100.0	-	99.9	98.0	91.4	85.3	75.7	63.0	85.8
15r	58.0	70.8	77.0	80.9	89.4	97.4	99.8	-	100	98.3	91.8	82.8	67.1	84.4
30r	56.7	68.3	72.0	72.0	78.1	87.9	95.2	99.0	-	99.5	94.8	87.4	73.0	81.9
45r	61.0	68.4	73.5	72.2	73.3	80.2	87.7	97.0	99.5	-	97.9	94.1	81.1	82.1
60r	60.7	68.1	73.4	69.0	70.1	73.3	79.9	87.7	94.3	98	-	99.9	95.7	80.8
75r	56.7	64.5	64.5	61.3	59.0	61.9	67.4	74.2	81.6	91.3	99.7	-	99.6	73.4
90r	52.1	53.8	54.4	52.1	50.2	51.6	55.2	58.2	61.4	71.3	91.3	98.2	-	62.4
Avg.	69.3	77.9	81.4	80.4	80.1	80.8	80.6	81.1	79.9	79.7	81.3	77.6	68.8	78.4

TABLE II
RESULTS FOR ALL INPUT AND OUTPUT POSE PAIRS BY USING GABOR FEATURES.

g/p	90l	75l	60l	45l	30l	15l	0	15r	30r	45r	60r	75r	90r	Avg.
90l	-	96.8	89.0	81.4	75.4	70.4	65.1	66.6	70.7	72.3	70.9	75.8	79.0	76.1
75l	97.8	-	98.7	95.0	90.1	83.0	73.2	75.9	78.3	81.6	84.3	85.4	76.5	84.9
60l	92.2	99.6	-	99.6	97.4	93.5	85.0	86.3	86.6	89.5	90.8	84.5	74.0	89.9
45l	84.2	95.6	99.5	-	99.9	97.8	91.5	90.4	91.9	92.7	86.5	80.8	72.5	90.2
30l	78.3	90.7	97.3	100	-	99.9	98.0	96.1	94.9	92.0	84.6	78.8	70.9	90.1
15l	73.0	84.5	93.7	99.1	99.9	-	99.8	98.9	97.1	92.1	84.6	78.2	69.8	89.2
0	69.7	78.7	87.2	94.2	99.1	99.9	-	99.9	98.4	94.6	85.6	76.5	68.1	87.6
15r	69.1	77.7	83.6	91.1	96.6	99.4	99.9	-	99.9	98.1	92.5	83.7	74.0	88.8
30r	69.2	77.7	84.6	90.7	94.0	95.1	97.8	100	-	99.9	97.9	89.7	79.8	89.7
45r	73.3	79.4	86.5	91.8	91.5	90.6	92.8	96.9	99.6	-	99.9	96.0	84.9	90.2
60r	73.8	84.4	91.5	89.2	85.6	85.0	84.8	91.2	97.3	99.8	-	99.7	93.6	89.6
75r	74.7	85.9	85.2	80.6	77.4	73.5	73.5	80.8	87.4	94.0	99.4	-	98.6	84.2
90r	79.6	76.3	73.3	71.0	69.2	65.8	64.5	66.8	74.6	79.3	90.5	98.4	-	75.7
Avg.	77.9	85.6	89.1	90.3	89.6	87.8	85.4	87.4	89.7	90.4	88.9	85.6	78.4	86.6

TABLE III
RESULTS FOR ALL INTENSITY LEVELS BY USING GABOR FEATURES.

intensity	90l	75l	60l	45l	30l	15l	15r	30r	45r	60r	75r	90r	Avg.
1	63.4	69.5	83.9	89.1	96.5	100	99.5	97.3	89.5	80.8	73.0	61.7	83.6
2	73.3	74.6	89.5	94.7	99.5	100	100	98.6	92.5	88.2	80.3	72.0	88.6
3	75.3	83.1	92.2	97.4	98.7	100	100	99.1	95.6	91.7	84.8	80.9	91.5
4	84.3	90.0	96.1	98.6	99.5	100	100	99.5	96.9	96.1	92.1	85.2	94.8
Avg.	74.0	79.3	90.4	94.9	98.5	100	99.8	98.6	93.6	89.2	82.5	74.9	89.6

with [13] whose pose-specific results are lower compared to the last row of Table III for each pose. Average recognition rate of all poses by using all intensity levels is reported as 74.1% in [13]. Here, we have an average recognition rate of 89.6% in Table III. In [3], Hu *et al.* obtain the highest recognition rate as 66.5% for pose-specific classifiers trained using normalized annotated landmarks. They report improved results of using SIFT features with LPP feature selection as 73.1% in [4]. Moore and Bowden obtain a recognition rate of 71.1% in [5]. These show that higher results can be achieved by matching the expressions of a subject from other poses to the frontal pose instead of matching expressions from the same pose. In [**?**], authors use a different dataset.

V. CONCLUSION

In this paper, we presented an approach to match an individual's expressions across different pose angles. We used an alignment method which works for all pose angles and experimented with local blocks of intensity values and Gabor wavelets to represent faces. We showed that size and location of the extracted blocks affect the performance and parameters might differ for two representations. We found that using local blocks of intensity values performs almost as well as local blocks of Gabor features for the cases in which input pose is the frontal pose. However, Gabor features outperform intensity features significantly for other pose pairs. We achieved an average recognition rate of 86.6% when all pose pairs are considered, and 87.6% when only the frontal gallery pose is considered. We showed that PLS method is very good at generalizing a subject's expressions over different pose angles and that cross-pose recognition might be a good alternative for pose-specific systems in multi-view facial expression recognition.

REFERENCES

[1] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D Facial Expression Database For Facial Behavior Research," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pp. 211–216, IEEE, 2006.

[2] F. De la Torre and J. F. Cohn, *Guide to Visual Analysis of Humans: Looking at People*, ch. Facial Expression Analysis. Springer, 2011.

[3] Y. Hu, Z. Zeng, L. Yin, X. Wei, J. Tu, and T. Huang, "A study of non-frontal-view facial expressions recognition," in *19th International Conference on Pattern Recognition (ICPR 2008)*, pp. 1 –4, 2008.

[4] Y. Hu, Z. Zeng, L. Yin, X. Wei, and X. Zhou, "Multi-view facial expression recognition," in *Face and Gesture Recognition*, pp. 1–6, IEEE, 2008.

[5] S. Moore and R. Bowden, "Local Binary Patterns for Multi-view Facial Expression Recognition," *Computer Vision and Image Understanding*, vol. 115, no. 4, pp. 541–558, 2011.

[6] O. Rudovic, I. Patras, and M. Pantic, "Regression-based Multi-view Facial Expression Recognition," in *20th International Conference on Pattern Recognition (ICPR 2010)*, pp. 4121–4124, IEEE, 2010.

[7] A. Sharma and D. W. Jacobs, "Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '11, (Washington, DC, USA), pp. 593–600, IEEE, 2011.

[8] A. Li, S. Shan, and W. Gao, "Coupled bias-variance tradeoff for cross-pose face recognition," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 305 –315, 2012.

[9] M. Fischer, H. K. Ekenel, and R. Stiefelhagen, "Analysis of partial least squares for pose-invariant face recognition," *IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems*, 2012.

[10] R. Roman and K. Nicole, "Overview and recent advances in partial least squares," in *Lecture Notes in Computer Science in Subspace, Latent Structure and Feature Selection Techniques*, pp. 34–51, Springer, 2006.

[11] L. Chengjun and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 467 –476, 2002.

[12] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 681–685, 2001.

[13] N. Hesse, T. Gehrig, H. Gao, and H. K. Ekenel, "Multi-view facial expression recognition using local appearance features," in *21st International Conference on Pattern Recognition (ICPR 2012)*, IEEE, 2012.