FACIAL EXPRESSION RECOGNITION USING LOG-EUCLIDEAN STATISTICAL SHAPE MODELS

Bartlomiej W. Papiez, Bogdan J. Matuszewski, Lik-Kwan Shark and Wei Quan

Applied Digital Signal and Image Processing Research Centre, University of Central Lancashire, PR1 2HE Preston, U.K.

Keywords: Facial expression representation, Facial expression recognition, Vectorial log-Euclidean statistics, Statistical shape modelling.

Abstract: This paper presents a new method for facial expression modelling and recognition based on diffeomorphic image registration parameterised via stationary velocity fields in Log-Euclidean framework. The validation and comparison are done using different statistical shape models (SSM) built using the Point Distribution Model (PDM), velocity fields, and deformation fields. The obtained results show that the facial expression representation based on stationary velocity field can be successfully utilised in facial expression recognition, and this parameterisation produces higher recognition rate than the facial expression representation based on deformation fields.

1 INTRODUCTION

Face is an important medium used by humans to communicate, but also reflecting a person’s emotional and awareness states, cognitive activity, personality or wellbeing. Over last ten years automatic facial expression representation and recognition have become area of significant research interest for the computer vision community, with applications in human-computer interaction (HCI) systems, medical/psychological sciences, and visual communications to name a few.

Although, significant efforts have been undertaken to improve the facial features extraction process and the recognition performance, automatic facial expression recognition is still a challenging task due to an inherent subjective nature of the facial expressions and their variation over different gender, age, and ethnicity groups. Detailed overview of existing methodologies, recent advances and challenges can be found in (Matuszewski et al., 2011; Tian et al., 2011; Fasel and Luettin, 2003; Pantic et al., 2000).

The facial expression representation can be seen as a process of extracting features, which could be generic as local binary patterns (Shan et al., 2005) or Gabor coefficients (Bartlett et al., 2003) or more specific such as landmarks of characteristic points located in areas of major facial changes due to articulation (Kobayashi and Hara, 1997), or a topographic context (TC) that treats the intensity levels of an image as a 3-D terrain surface (Wang and Yin, 2007). Recently, in (Quan et al., 2007b; Quan et al., 2009) authors postulated that the space shape vectors (SSV) of the statistical shape model (SSM) can constitute a significant feature space for the recognition of facial expressions. The SSM can be constructed in many different ways, and it was developed based on the point distribution model originally proposed by (Cootes et al., 1995). In (Quan et al., 2007a), the SSM is built based on the control points of the B-Spline surface of the training data set, and in (Quan et al., 2010) an improved version with multi-resolution correspondence search and multi-level model deformation was proposed. In this paper, the SSM is generated using the stationary velocity fields obtained from diffeomorphic face registration. The idea of using the motion fields as feature in computer vision and pattern recognition was used previously for face recognition where the optical flow was computed to robustly recognise face under different expressions based on a single sample per class in the training set (Hsieh et al., 2010).

In medical image analysis, the parameterisation of the diffeomorphic transformation based on the principal logarithm to non-linear geometrical deformations was introduced by (Arsigny et al., 2006). Using this framework, the Log-Euclidean vectorial statistics can be performed on the diffeomorphic vector fields via their logarithm, which always preserve the invertibility constraint contrary to the Euclidean statistics on...
the deformation fields. Recently, the stationary velocity field parametrisation has been utilised for deformable image registration in different way e. g. for exponential update of deformation field (Vercouteren et al., 2009), or producing the principal logarithm directly as an output of image registration e. g. inverse consistent image registration (Ashburner, 2007; Vercouteren et al., 2008) or symmetric inverse consistent image registration (Han et al., 2010). Those algorithms preserve the spatial topology of objects by maintaining diffeomorphism. As the facial shapes (mouth, eyes, eye brows) have constant intra- and inter-subject topology, it is interesting to check adequacy of the facial expressions represented using a stationary velocity fields as a result of performing diffeomorphic image registration and compare with the deformation field based facial expression representation in terms of separability in feature space and recognition performance.

The remainder of the paper is organised as follows. Section 2 introduces the concept of the SSM with detailed description of the group-wise registration algorithm (Section 2.1). Then, the velocity field based representation of facial expression is described in 2.2, and the Point Distribution Model is presented in Section 2.3. The experimental results of qualitative and quantitative evaluation are shown in Section 3 with concluding remarks in Section 4.

## 2 STATISTICAL SHAPE MODEL

The statistical shape model was developed based on the point distribution model originally proposed by (Cootes et al., 1995). The model represents the facial expression variations based on the statistics calculated for corresponding features during the learning process for the training data set. In order to build an SSM, the correspondence of facial features between different faces in the training data set must be established. This is done here first by generating a mean face model for the neutral facial expression data set to find the mappings from any face to the so called common face space. Then, by transferring subject specific facial expression data set into the common face space, the intra-subject facial expression correspondence is estimated. Finally, the principal component analysis (PCA) is applied to the training data set aligned in the common face space, to provide a low-dimensional feature space for facial expression representation.

### 2.1 Log-domain Group-wise Image Registration

Generation of the mean face model is an essential step during the training process because it allows a subject independent common face space to be established for further analysis.

For a given set of n-dimensional images representing neutral facial expressions denoted by $\mathbf{I}_n^\text{ne}$, the objective is to estimate a set of displacement fields $\hat{\mathbf{u}}^\text{ne}$ to map the image taken from $\mathbf{I}_n^\text{ne}$ to the mean face model $\mathbf{I}_n^\text{mean}$.

In general, this problem can be formulated as a minimisation problem:

$$\hat{\mathbf{u}}^\text{ne} = \arg\min_{\mathbf{u}^\text{ne}} \mathcal{E}(\mathbf{u}^\text{ne}; \mathbf{I}_n^\text{ne})$$

(2)

where $\mathcal{E}(\mathbf{u}^\text{ne})$ is defined as

$$\mathcal{E}(\mathbf{u}^\text{ne}) = \sum_{I} \int_{\Omega} \text{Sim}(\mathbf{I}_l^\text{ne}(\tilde{x} + \hat{\mathbf{u}}_l(x), \mathbf{I}_k^\text{ne}(\tilde{x} + \tilde{u}_k))) \, dx$$

$$+ \alpha \sum_{I} \int_{\Omega} \text{Reg}(\hat{\mathbf{u}}_k(x)) \, dx$$

(3)

where $\tilde{x} = [x_1, \ldots, x_n] \in \Omega$ denotes given voxel position, $\text{Sim}$ denotes a similarity measure between each pair of the images, $\mathbf{I}_l^\text{ne}$ and $\mathbf{I}_k^\text{ne}$ ($l \neq k$) from $\mathbf{I}_n^\text{ne}$, $\text{Reg}$ denotes a regularisation term, and $\alpha$ is a weight of the regularisation term. In this work, the deformation fields are parameterised by recently proposed stationary velocity fields $\tilde{v}(\tilde{x})$ via exponential mapping (Arsigny et al., 2006):

$$\varphi(\tilde{x}) = \tilde{x} + \hat{u}(\tilde{x}) = \tilde{x} + \exp(\tilde{v}(\tilde{x})).$$

(4)

To minimise Equation 2, Demon force (Vercouteren et al., 2009) was used in the symmetric manner (Papiez and Matuszewski, 2011) in the following way:

$$\tilde{u}^i_{kl} = \frac{(I^k_l - I^l_k) (\nabla I^k_l + \nabla I^l_k)}{||\nabla I^k_l + \nabla I^l_k||^2 + (I^k_l - I^l_k)^2}$$

(5)

where $I^k_l = I^k_l(\varphi_l(\tilde{x}))$, $I^l_k = I^l_k(\varphi_l(\tilde{x}))$ are warped images and $\nabla I^k_l$, $\nabla I^l_k$ are gradients of those images, and $i$ is an iteration index. The average update of the velocity field is calculated using the Log-Euclidian mean for vector fields $\tilde{u}^i_{kl}$ given by (Arsigny et al., 2006):

$$\tilde{u}^i_k = \frac{1}{K} \sum \log(\tilde{u}^i_{kl})$$

(6)
and the deformation field $\vec{u}_{i+1}^{k}(\vec{x})$ is calculated via exponential mapping for the updated velocity field:

$$\vec{v}_{i+1}^{k}(\vec{x}) = \vec{v}_{i}^{k}(\vec{x}) + \vec{dv}_{i}^{k}(\vec{x})$$

(7)

Although according to Equation 6 the Log-Euclidean mean requires calculating of the logarithm, which is reported to be a time-consuming process (Arsigny et al., 2006; Bossa et al., 2007), the symmetric Log-Domain Diffomorphic Demon approach (Vercauteren et al., 2008) is used which produces the principal logarithm of transformation as an output of image registration and therefore the logarithm is not calculated directly. Finally, the mean face model is generated by averaging the intensity of all images after registration:

$$I_{\text{mean}} = \frac{1}{K} \sum_{k=1}^{K} I_{\text{ne}}^{k}(\vec{x})$$

(8)

The procedure for estimation of the set of deformation fields for generation the common face space is summarised below:

repeat
  for $k=1$ to $K$
    for $l=1$ to $K$ and $l \neq k$
      Calculate update (Equation 5)
    end
    Calculate average of updates (Equation 6)
    Update velocity field (Equation 7)
    Smooth velocity field using Gaussian filter
  end
  $i = i+1$;
until (velocity fields do not change) or ($i > \text{max\_Iteration}$)

As an example, the mean face model estimated by applying the scheme to neutral expressions is illustrated in Figure 1, and the faces with neutral expression mapped into common face space are shown in Figure 2.

![Figure 1: Grey-level average of mean face before registration (left), and after registration (right), obtained for 40 images from training data set.](image)

![Figure 2: Examples of images representing neutral expression (top) and the same faces mapped into the common face space.](image)

The presented algorithm of generating the mean face model is similar to the work presented in (Geng et al., 2009). The main difference is in how the deformation fields are parameterised with the stationary velocity field used in the proposed method instead of the Fourier series in (Geng et al., 2009), and secondly in the method of solving Equation 2 with the Demon approach used instead of the linear elastic model. Using the Log-domain parametrisation for deformation fields is reported to produce smoother deformation fields and it allows vectorial statistics to be calculated directly on the velocity fields.

### 2.2 Velocity Field based Facial Expression Model

The next step is to warp all other training faces representing different facial expressions to the mean face (the reference face) via transformation $\varphi_{k}(\vec{x})$ estimated for neutral expressions. For a given set of facial expression images from subject $K$:

$$I_{\text{ke}}^{c} = \{ I_{\text{ke}}^{cm}(\varphi_{k}(\vec{x})), m = 1, \ldots, M \}$$

(9)

where $M$ denotes the number of images, the transformation $\varphi_{k}(\vec{x})$ is applied to get a set of facial expression images in the common face space (space of the reference image):

$$I_{\text{ke}}^{c} = \{ I_{\text{ke}}^{cm}(\varphi_{k}(\vec{x})), m = 1, \ldots, M \}$$

(10)

By applying the Log-Domain image registration approach based on the consistent symmetric Demon algorithm (Vercauteren et al., 2008), each image in set $I_{\text{ke}}^{c}$ is registered to image of neutral expression in common face space $I_{\text{ne}}^{c}(\varphi_{k}(\vec{x}))$, the set of the velocity fields $\vec{v}_{i}^{c}$ is estimated, and the set of the corresponding deformation fields $\vec{u}_{i}^{c}$ via exponential mapping is calculated as well. Utilising this particular method for image registration has two important advantages. Firstly, the consistency criterion is maintained during the registration process that helps to keep the smooth transformation especially for cases like matching between open-mouth and close-mouth
shapes. Secondly, the results of registration are the velocity fields so there is no necessity of calculating the principal logarithm of transformations.

2.3 Point Distribution Model

The point distribution model originally proposed by (Cootes et al., 1995) is one of most often used techniques for representing shapes. This model describes a shape as a set of positions (landmarks) in the reference image. The variations between different shapes require establishment of the correspondence between points detected in the reference image to images representing different deformation in the training set. Although this can be relatively reliably achieved during the model training phase by careful time consuming, often manual selection of corresponding points, such task is prone to occurrence of gross errors during the model evaluation where often near real time performance is required. The examples of the manually selected landmarks for neutral and happiness expression are shown in Figure 3. The automatically selected landmarks used later on in the experimental section are obtained with help of face image registration described in the previous section. In that case, the manually selected landmarks in the model face are automatically mapped into registered faces.

2.4 Principal Component Analysis

Using the standard principal component analysis (PCA), each face representation in the training data set can be approximately represented in a low-dimensional shape vector space instead of the original high-dimensional data vector space (Bishop, 2006). Figure 4 shows the effect of varying the first three largest principal component of the PDM for automatically selected landmarks, where \( \lambda \) is eigenvalue of the covariance matrix calculated from the training data set.

3 EXPERIMENTAL RESULTS

The data set used for validation (Yin et al., 2006) consists of 48 subjects with a wide variety of ethnicity, age and gender. Some example faces taken from that database are shown in Figure 5, and Figure 6 shows the range of expression intensity. The data used during the training procedure are mutually excluded with the data used for validation. The group-wise registration based on Demon minimises the Sum of Squared Difference (SSD) between images and hence due to different skin patterns an additional image intensities values adjustment was performed.

3.1 Separability Analysis

To assess whether the Shape Space Vectors based on the velocity fields can be used as a feature space for facial expression analysis and recognition, the separability of the SSV-based features has been analysed.

The first three element of the SSM are used to reveal clustering characteristics and separability pow-
The SSM for training was built using 24 subjects, each containing 25 faces, the SSV is based on the automatically selected points (with 60 landmarks per face), the velocity fields, and the deformation fields (with 512x512 pixels per image). The test data set was extracted from another 24 subjects. The training data set and the testing data set are mutually exclusive. Examples of some expressions given in Figures 7-9 exhibit good separability even in the low-dimensional space, especially for expressions such as "happiness vs. sadness" or "disgust vs. surprise". The expressions like "anger vs. fear" appear to overlap more each other, but the clusters can be identified.

In order to quantitatively assess the separability of the presented facial expression features, the appropriate criteria have to be calculated. A computable criterion for measurement of within-class and between-class distances was computed similarly as it was done by (Wang and Yin, 2007; Quan et al., 2009). The within-class scatter matrix $S_W$ is defined as follows:

$$S_W = c \sum_{i=1}^{c} \frac{1}{n_i} \sum_{k=1}^{n_i} (\vec{x}_k - \vec{m}_i)(\vec{x}_k - \vec{m}_i)^T$$  \hspace{1cm} (11)$$

and the between-class scatter matrix $S_B$ is defined as:

$$S_B = c \sum_{i=1}^{c} \frac{1}{n_i} (\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})^T$$ \hspace{1cm} (12)$$

where: $\vec{x}_k$ is $d$-dimensional feature, $n_i$ is the number of samples in $i$th class, $n$ is the number of samples in all classes, $c$ is the number of classes, $\vec{m}_i$ is the mean of samples in the $i$th class defined as:

$$\vec{m}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} \vec{x}_k$$  \hspace{1cm} (13)$$

$\vec{m}$ is the mean of all the samples:

$$\vec{m} = \sum_{i=1}^{c} P_i \vec{m}_i$$ \hspace{1cm} (14)$$

The separability criterion $J_2(\vec{x})$ is defined as a natural logarithm of the ratio within-class scatter matrix’s determinant and between-class scatter matrix’s determinant:

$$J_2(\vec{x}) = \ln \left( \frac{\det(S_B + S_W)}{\det(S_W)} \right)$$  \hspace{1cm} (15)$$

This separability criterion is efficient for comparison of different feature selection, lying in the completely different spaces (also with different dimensionalities), and it is intrinsically normalised and reflects the quantity of separability for features between different classes (Wang and Yin, 2007; Quan et al., 2009). The larger value of $J_2(\vec{x})$ means the better separability. The separability criterion was evaluated on the different facial expression representation and the results are shown in Figure 10. For the same ratio of retained energy in the training data, the value of $J_2(\vec{x})$ for the manually selected landmarks is the highest. The automatically selected landmarks in range above 80% is not significantly different than the manually selected landmarks. The velocity field and the deformation field based facial expression representation is the worst.

To quantify the between-expression separability, the two-class separability criterion is evaluated (Wang and Yin, 2007). The within-class scatter matrix
S_{W}^{e_{i}, e_{j}} for two classes case (c=2) is defined as follows:

\[ S_{W}^{e_{i}, e_{j}} = \frac{1}{n} \left( \sum_{k=1}^{n_{e_{i}}} (\vec{x}_{k}^{e_{i}} - \vec{m}_{e_{i}})(\vec{x}_{k}^{e_{j}} - \vec{m}_{e_{j}})^{T} \right) + \sum_{l=1}^{n_{e_{j}}} (\vec{x}_{l}^{e_{j}} - \vec{m}_{e_{j}})(\vec{x}_{l}^{e_{i}} - \vec{m}_{e_{i}})^{T} \] (16)

and between-class scatter matrix \( S_{B}^{e_{i}, e_{j}} \) is defined:

\[ S_{B}^{e_{i}, e_{j}} = \frac{n_{e_{i}}n_{e_{j}}}{n^{2}} (\vec{m}_{e_{i}} - \vec{m}_{e_{j}})(\vec{m}_{e_{i}} - \vec{m}_{e_{j}})^{T} \] (17)

where \( e_{i} \) and \( e_{j} \) are analysed expressions, \( n_{e_{i}}, n_{e_{j}} \) are the numbers of samples in \( i \)th and \( j \)th class, \( n = n_{e_{i}} + n_{e_{j}} \). Then for each pair of selected expressions \( J_{2}^{e_{i}, e_{j}}(\vec{x}) \) on the different facial expression representation is calculated.

Tables 1-4 shows the separability of all pairs of expression for different facial expression representation. Those results support the visual inspection of the qualitative analysis presented in Figures 7-9. The separability of the pair of expression such as happiness and sadness, or disgust and surprise gets higher values of separability criterion \( \text{exp}^{J_{2}}(\vec{x}) \) (the minimum 2.93), while the pair angry and fear lower (the maximum 2.62).

Table 1: Confusion matrix of the expression separability criterion \( J_{2}^{e_{i}, e_{j}}(\vec{x}) \) for the manually selected landmarks.

<table>
<thead>
<tr>
<th></th>
<th>Ang</th>
<th>Dis</th>
<th>Fea</th>
<th>Hap</th>
<th>Sad</th>
<th>Sur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ang</td>
<td></td>
<td></td>
<td>2.09</td>
<td>2.26</td>
<td>3.61</td>
<td>1.61</td>
</tr>
<tr>
<td>Dis</td>
<td>2.09</td>
<td></td>
<td>2.37</td>
<td>3.65</td>
<td>2.80</td>
<td>3.56</td>
</tr>
<tr>
<td>Fea</td>
<td>2.26</td>
<td>2.37</td>
<td></td>
<td>1.98</td>
<td>2.02</td>
<td>2.66</td>
</tr>
<tr>
<td>Hap</td>
<td>3.61</td>
<td>3.65</td>
<td>1.98</td>
<td></td>
<td>3.93</td>
<td>4.44</td>
</tr>
<tr>
<td>Sad</td>
<td>1.61</td>
<td>2.80</td>
<td>2.02</td>
<td>3.93</td>
<td></td>
<td>3.94</td>
</tr>
<tr>
<td>Sur</td>
<td>4.19</td>
<td>3.56</td>
<td>2.66</td>
<td>4.44</td>
<td>3.94</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Confusion matrix of the expression separability criterion of $J^2(x_i, x^j)$ for the automatically selected landmarks.

<table>
<thead>
<tr>
<th>Input/</th>
<th>Ang</th>
<th>Dis</th>
<th>Fea</th>
<th>Hap</th>
<th>Sad</th>
<th>Sur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
</tr>
<tr>
<td>Ang</td>
<td>74.5</td>
<td>4.7</td>
<td>3.1</td>
<td>3.1</td>
<td>14.6</td>
<td>0.0</td>
</tr>
<tr>
<td>Dis</td>
<td>8.3</td>
<td>5.2</td>
<td>2.6</td>
<td>1.6</td>
<td>0.5</td>
<td>3.6</td>
</tr>
<tr>
<td>Fea</td>
<td>7.8</td>
<td>1.6</td>
<td>59.9</td>
<td>11.5</td>
<td>16.1</td>
<td>3.1</td>
</tr>
<tr>
<td>Hap</td>
<td>4.2</td>
<td>2.1</td>
<td>8.3</td>
<td>85.4</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Sad</td>
<td>16.7</td>
<td>1.6</td>
<td>4.2</td>
<td>0.0</td>
<td>77.6</td>
<td>0.0</td>
</tr>
<tr>
<td>Sur</td>
<td>1.0</td>
<td>2.1</td>
<td>4.2</td>
<td>0.5</td>
<td>2.6</td>
<td>89.6</td>
</tr>
</tbody>
</table>

Table 3: Confusion matrix of the expression separability criterion of $J^2(x_i, x^j)$ for the full deformation fields.

<table>
<thead>
<tr>
<th>Input/</th>
<th>Ang</th>
<th>Dis</th>
<th>Fea</th>
<th>Hap</th>
<th>Sad</th>
<th>Sur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
</tr>
<tr>
<td>Ang</td>
<td>68.8</td>
<td>5.2</td>
<td>5.2</td>
<td>2.6</td>
<td>18.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Dis</td>
<td>12.5</td>
<td>76.6</td>
<td>5.7</td>
<td>0.5</td>
<td>3.6</td>
<td>0.0</td>
</tr>
<tr>
<td>Fea</td>
<td>7.8</td>
<td>2.6</td>
<td>55.2</td>
<td>14.1</td>
<td>19.3</td>
<td>0.0</td>
</tr>
<tr>
<td>Hap</td>
<td>4.1</td>
<td>1.6</td>
<td>11.5</td>
<td>82.3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Sad</td>
<td>19.8</td>
<td>3.1</td>
<td>4.7</td>
<td>0.0</td>
<td>72.4</td>
<td>0.0</td>
</tr>
<tr>
<td>Sur</td>
<td>1.0</td>
<td>3.1</td>
<td>7.8</td>
<td>0.5</td>
<td>2.6</td>
<td>87.0</td>
</tr>
</tbody>
</table>

Table 4: Confusion matrix of the expression separability criterion of $J^2(x)$ for the full velocity fields.

<table>
<thead>
<tr>
<th>Input/</th>
<th>Ang</th>
<th>Dis</th>
<th>Fea</th>
<th>Hap</th>
<th>Sad</th>
<th>Sur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
</tr>
<tr>
<td>Ang</td>
<td>74.5</td>
<td>9.9</td>
<td>1.0</td>
<td>2.6</td>
<td>10.9</td>
<td>1.0</td>
</tr>
<tr>
<td>Dis</td>
<td>9.4</td>
<td>75.5</td>
<td>6.3</td>
<td>5.7</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td>Fea</td>
<td>5.7</td>
<td>2.6</td>
<td>56.8</td>
<td>15.6</td>
<td>11.5</td>
<td>7.8</td>
</tr>
<tr>
<td>Hap</td>
<td>2.1</td>
<td>6.3</td>
<td>16.1</td>
<td>74.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Sad</td>
<td>12.0</td>
<td>0.5</td>
<td>7.3</td>
<td>2.1</td>
<td>78.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Sur</td>
<td>2.6</td>
<td>1.0</td>
<td>2.1</td>
<td>2.1</td>
<td>1.0</td>
<td>91.1</td>
</tr>
</tbody>
</table>

3.2 Experiments on Facial Expression Recognition

The separability analysis conducted in the previous section indicates that the SSV feature space based on the velocity can be used for classification of facial expressions. Data used for classification based validation again consists of 48 subjects, and contains neutral expression, and six basic facial expressions of anger, disgust, fear, happiness, sadness, and surprise with four different expression intensity ranges. These data were divided into six subsets containing 8 subjects with 25 faces per subject representing different expressions. During evaluation procedure one subset is chosen as the testing set, the remaining data are used for the training procedure. Four types of facial expression representation have been used for validation: the manually selected landmarks from the database (Yin et al., 2006), the automatically detected facial landmarks using Log-Domain Demon registration, the full velocity fields, and the full deformation fields.

Three commonly used classification methods were used for evaluation, namely linear discriminant analysis (LDA), quadratic classifier (QDC), and nearest neighbour classifier (NCC). The detailed description of these methods can be found in most of the textbooks on pattern recognition e.g. (Bishop, 2006).

The average recognition rates and standard deviations of all six experiments for different facial ex-
Taking into account the "subjective" nature of the rates (e.g., happiness, or surprise), of separability criterion $J$ error is the highest. The expressions with high value of separability criterion $J$ are more likely to be misclassified (e.g., fear and sadness). The performed tests show also that the parameterisation via stationary velocity fields in Log-Domain produces slightly higher recognition rate of facial expressions that produced by using deformation fields.

**ACKNOWLEDGEMENTS**

The work has been in part supported by the MEGURATH project (EPSRC project No. EP/D077540/1).

### REFERENCES


