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

W. Papiez B., J. Matuszewski B., Shark L. and Quan W..
FACIAL EXPRESSION RECOGNITION USING LOG-EUCLIDEAN STATISTICAL SHAPE MODELS.
DOI: 10.5220/0003867503510359
In *Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods* (SADM-2012), pages 351-359
ISBN: 978-989-8425-98-0

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 (Vercauteren et al., 2009), or producing the principal logarithm directly as an output of image registration e. g. inverse consistent image registration (Ashburner, 2007; Vercauteren 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 expressions 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

$$\mathbf{I^{ne}} = \{I_k^{ne} : \Omega \subset \mathbb{R}^n \to \mathbb{R}, k = 1, \ldots, K\} \qquad (1)$$

where K is the number of subjects included in training data, the objective is to estimate a set of displacement fields $\hat{\mathbf{u}}^{\mathbf{ne}}$ to map the image taken from $\mathbf{I^{ne}}$ to the *mean* face model I^{mean}.

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

$$\hat{\mathbf{u}}^{\mathbf{ne}} = \arg\min_{\mathbf{u^{ne}}} \varepsilon(\mathbf{u^{ne}}; \mathbf{I^{ne}}) \qquad (2)$$

where $\varepsilon(\mathbf{u^{ne}})$ is defined as

$$\varepsilon(\mathbf{u^{ne}}) = \sum_k \sum_l \int_\Omega Sim(I_k^{ne}(\vec{x} + \vec{u}_k(\vec{x}), I_l^{ne}(\vec{x} + \vec{u}_l))dx + \alpha \sum_k \int_\Omega Reg(\vec{u}_k(\vec{x}))dx \qquad (3)$$

where $\vec{x} = [x^1, \ldots, x^n] \in \Omega$ denotes given voxel position, *Sim* denotes a similarity measure between each pair of the images, I_k^{ne} and I_l^{ne} $(l \neq k)$ from $\mathbf{I^{ne}}$, *Reg* denotes a regularisation term, and α is a weight of the regularisation term. In this work, the deformation fields are parameterised by recently proposed stationary velocity fields $\vec{v}(\vec{x})$ via exponential mapping (Arsigny et al., 2006):

$$\varphi(\vec{x}) = \vec{x} + \vec{u}(\vec{x}) = \vec{x} + exp(\vec{v}(\vec{x})). \qquad (4)$$

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

$$\vec{du}_{kl}^{\,i} = \frac{(I_k^{\varphi_k^i} - I_l^{\varphi_l^i})(\nabla I_k^{\varphi_k^i} + \nabla I_l^{\varphi_l^i})}{\|\nabla I_k^{\varphi_k^i} + \nabla I_l^{\varphi_l^i}\|^2 + (I_k^{\varphi_k^i} - I_l^{\varphi_l^i})^2} \qquad (5)$$

where $I_k^{\varphi_k^i} = I_k^{ne}(\varphi_k^i(\vec{x}))$, $I_l^{\varphi_l^i} = I_l^{ne}(\varphi_l^i(\vec{x}))$ are warped images and $\nabla I_k^{\varphi_k^i}$, $\nabla I_l^{\varphi_l^i}$ are gradients of those images, and i is an iteration index. The average update of the velocity field is calculated using the Log-Euclidean mean for vector fields $\vec{du}_{kl}^{\,i}$ given by (Arsigny et al., 2006):

$$\vec{dv}_k^{\,i} = \frac{1}{K} \sum_l log(\vec{du}_{kl}^{\,i}) \qquad (6)$$

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

$$\vec{v}_k^{i+1}(\vec{x}) = \vec{v}_k^{i}(\vec{x}) + \vec{d}v_k^{i}(\vec{x}) \quad (7)$$

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 Diffemorphic 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:

$$I^{mean} = \frac{1}{K}\sum_{k}^{K} I_k^{ne}(\varphi_k(\vec{x})) \quad (8)$$

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

```
repeat
   for k=1:K
      for l=1:K and l!=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>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.

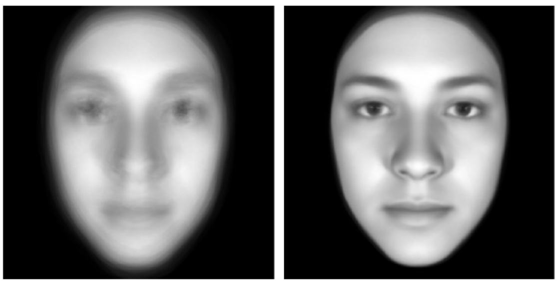

Figure 1: Grey-level average of mean face before registration (left), and after registration (right), obtained for 40 images from training data set.

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.

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

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:

$$\mathbf{I_k^{ex}} = \{I_{km}^{ex} : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}, m = 1, \ldots, M\} \quad (9)$$

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):

$$\mathbf{I_k^{cex}} = \{I_{km}^{ex}(\varphi_k(\vec{x}))\} \quad (10)$$

By applying the Log-Domain image registration approach based on the consistent symmetric Demon algorithm (Vercauteren et al., 2008), each image in set $\mathbf{I_k^{cex}}$ is registered to image of neutral expression in common face space $I_k^{ne}(\varphi_k(\vec{x}))$, the set of the velocity fields $\mathbf{v_k^{ex}}$ is estimated, and the set of the corresponding deformation fields $\mathbf{u_k^{ex}}$ 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.

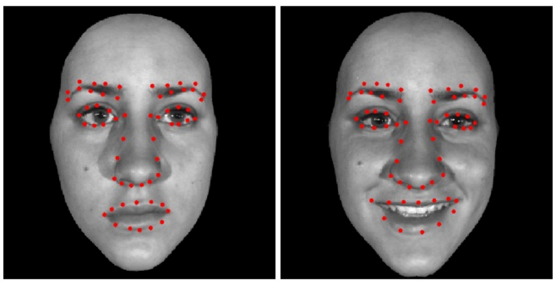

Figure 3: Manually selected landmarks for neutral expression(left), and happiness expression(right).

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 λ is eigenvalue of the covariance matrix calculated from the training data set.

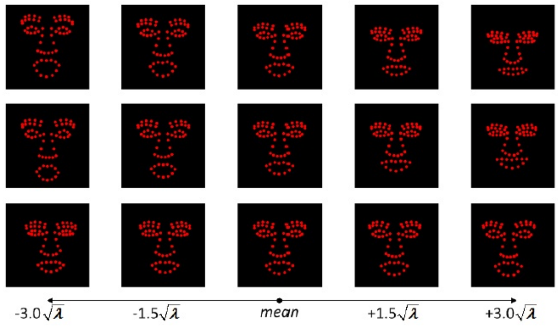

Figure 4: Variations of the first (top row), the second (middle row), and the third (bottom row) major mode of the Point Distribution Model for automatically selected landmarks.

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.

Figure 5: Four sample subjects showing seven expressions (neutral, angry, disgust, fear, happiness, sadness, and surprise).

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-

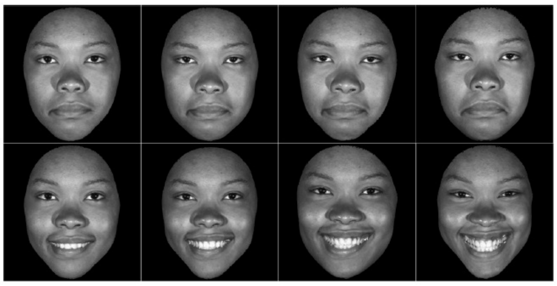

Figure 6: Samples of expressions: sadness (top) and happiness (bottom) for of different expression intensity ranges from low (left), middle, high to highest (right).

ers. 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:

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

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

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

where: $\vec{x}_k^i$ is d-dimensional feature, n_i is the number of samples in ith 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 ith class defined as:

$$\vec{m}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} \vec{x}_k^i \tag{13}$$

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

$$\vec{m} = \sum_{i=1}^{c} P_i \vec{m}_i \tag{14}$$

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:

$$J_2(\vec{x}) = \ln \frac{det(S_B + S_W)}{det(S_W)} \tag{15}$$

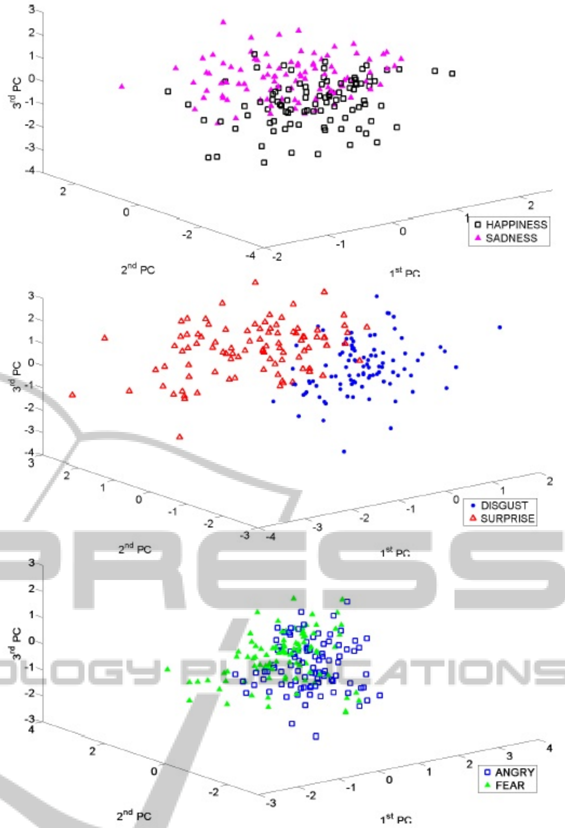

Figure 7: Separability Analysis for automatically selected landmarks using first three principal components.

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

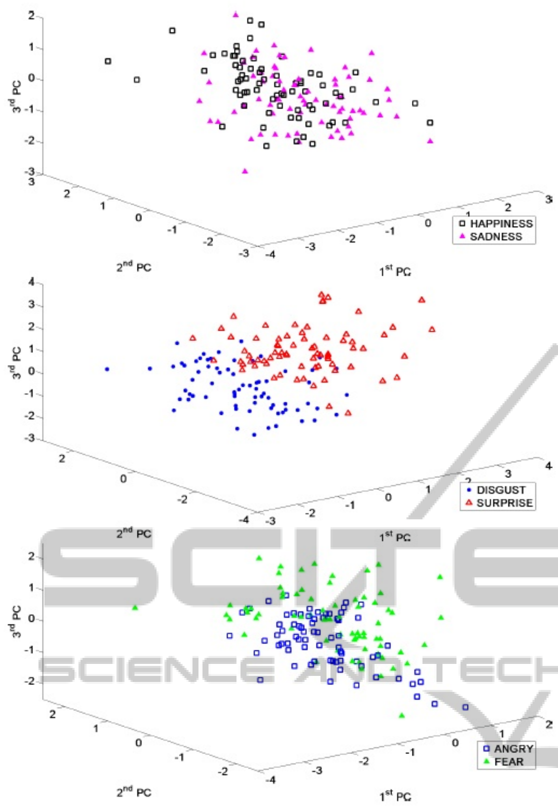

Figure 8: Separability Analysis for full deformation field using first three principal components.

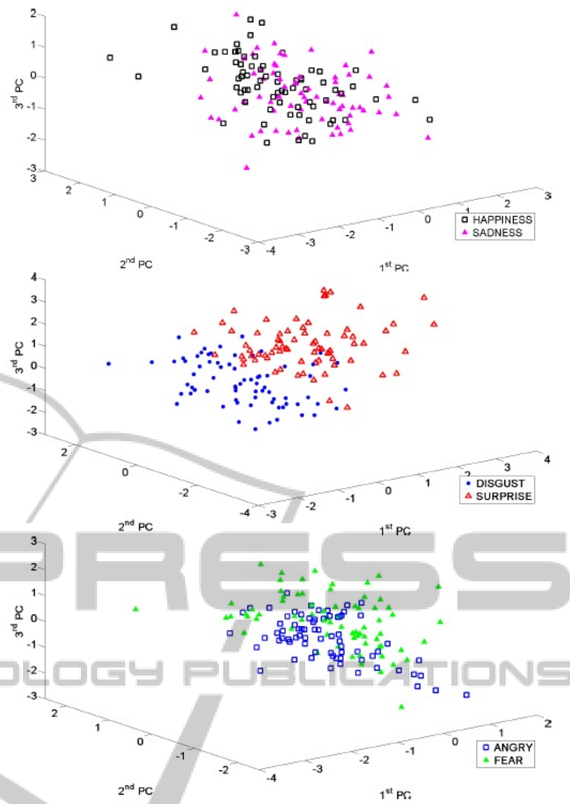

Figure 9: Separability Analysis for full velocity field using first three principal components.

$S_W^{ex_i,ex_j}$ for two classes case (c=2) is defined as follows:

$$S_W^{ex_i,ex_j} = \frac{1}{n}\Big(\sum_{k=1}^{n_{ex_i}}(\vec{x}_k^{ex_i}-\vec{m}_{ex_i})(\vec{x}_k^{ex_i}-\vec{m}_{ex_i})^T + \sum_{l=1}^{n_{ex_j}}(\vec{x}_l^{ex_j}-\vec{m}_{ex_j})(\vec{x}_l^{ex_j}-\vec{m}_{ex_j})^T\Big) \quad (16)$$

and between-class scatter matrix $S_B^{ex_i,ex_j}$ is defined:

$$S_B^{ex_i,ex_j} = \frac{n_{ex_i}n_{ex_j}}{n^2}(\vec{m}_{ex_i}-\vec{m}_{ex_j})(\vec{m}_{ex_i}-\vec{m}_{ex_j})^T \quad (17)$$

where ex_i and ex_j are analysed expressions, n_{ex_i}, n_{ex_j} are the numbers of samples in ith and jth class, $n = n_{ex_i}+n_{ex_j}$. Then for each pair of selected expressions $J_2^{ex_i,ex_j}(\vec{x})$ 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 $\exp^{J_2(\vec{x})}$ (the minimum 2.93), while the pair angry and fear lower (the maximum 2.62).

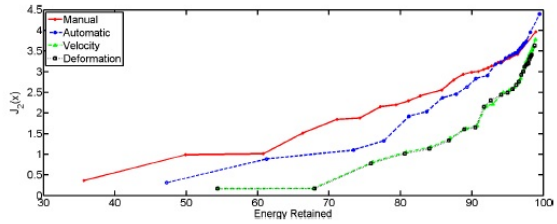

Figure 10: Separability of expression for different features in term of separability criterion $J_2(\vec{x})$.

Table 1: Confusion matrix of the expression separability criterion of $J_2^{ex_i,ex_j}(\vec{x})$ for the manually selected landmarks.

	Ang	Dis	Fea	Hap	Sad	Sur
Ang	-	2.09	2.26	3.61	1.61	4.19
Dis	-	-	2.37	3.65	2.80	3.56
Fea	-	-	-	1.98	2.02	2.66
Hap	-	-	-	-	3.93	4.44
Sad	-	-	-	-	-	3.94
Sur	-	-	-	-	-	-

Table 2: Confusion matrix of the expression separability criterion of $J_2^{ex_i,ex_j}(\vec{x})$ for the automatically selected landmarks.

	Ang	Dis	Fea	Hap	Sad	Sur
Ang	-	1.91	2.01	3.08	1.44	3.33
Dis	-	-	2.03	3.08	2.33	2.93
Fea	-	-	-	1.90	1.84	2.33
Hap	-	-	-	-	3.27	3.68
Sad	-	-	-	-	-	3.20
Sur	-	-	-	-	-	-

Table 3: Confusion matrix of the expression separability criterion of $J_2^{ex_i,ex_j}(\vec{x})$ for the full deformation fields.

	Ang	Dis	Fea	Hap	Sad	Sur
Ang	-	2.15	2.53	2.68	1.74	4.21
Dis	-	-	2.23	3.39	2.68	3.68
Fea	-	-	-	1.72	2.05	2.67
Hap	-	-	-	-	3.25	3.82
Sad	-	-	-	-	-	4.08
Sur	-	-	-	-	-	-

Table 4: Confusion matrix of the expression separability criterion of $J_2(\vec{x})$ for the full velocity fields.

	Ang	Dis	Fea	Hap	Sad	Sur
Ang	-	2.23	2.62	2.83	1.78	4.10
Dis	-	-	2.27	2.44	2.83	3.63
Fea	-	-	-	1.88	2.09	2.73
Hap	-	-	-	-	3.53	3.96
Sad	-	-	-	-	-	4.27
Sur	-	-	-	-	-	-

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.

Table 5: Confusion matrix of the LDA for the manually selected landmarks.

Input/ Output	Ang (%)	Dis (%)	Fea (%)	Hap (%)	Sad (%)	Sur (%)
Ang	74.5	4.7	3.1	3.1	14.6	0.0
Dis	8.3	81.8	4.7	0.5	3.6	1.0
Fea	7.8	1.6	59.9	11.5	16.1	3.1
Hap	4.2	2.1	8.3	85.4	0.0	0.0
Sad	16.7	1.6	4.2	0.0	77.6	0.0
Sur	1.0	2.1	4.2	0.5	2.6	89.6

Table 6: Confusion matrix of the LDA for the automatic selected landmarks.

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

Table 7: Confusion matrix of the LDA for the full deformation fields.

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

Table 8: Confusion matrix of the LDA for the full velocity fields.

Input/ Output	Ang (%)	Dis (%)	Fea (%)	Hap (%)	Sad (%)	Sur (%)
Ang	77.6	7.8	0.5	2.1	11.5	0.5
Dis	8.9	77.1	5.2	5.2	2.6	1.0
Fea	4.7	3.6	61.5	9.9	13.0	7.3
Hap	3.1	6.3	14.1	76.0	0.0	0.5
Sad	15.1	0.0	6.8	1.6	76.6	0.0
Sur	1.6	1.6	3.6	1.0	1.6	90.6

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-

Table 9: Summary of diagonal for confusion matrix of the LDA for different features.

Fea./ Exp.	Ang (%)	Dis (%)	Fea (%)	Hap (%)	Sad (%)	Sur (%)
Man.	74.5	81.8	59.9	85.4	77.6	89.6
Aut.	68.8	76.6	55.2	82.3	72.4	87.0
Def.	74.5	75.5	56.8	74.0	78.1	91.1
Vel.	77.6	77.1	61.5	76.0	76.6	90.6

Table 10: Recognition rate for different classifier's methods.

Feature/ classifier	LDA (%± SD)	QDA (%± SD)	NNC (%± SD)
Manually	78.1±4.2	74.0±4.8	61.5±1.1
Automatic	73.4±6.0	**69.1±6.0**	**61.9±4.9**
Deformation	75.0±5.2	58.9±2.9	56.2±5.0
Velocity	**76.6±4.3**	59.3±4.1	57.7±5.1

pression data are presented in Table 10. It can be seen that LDA classifier achieves the highest recognition rate for every facial expression representation. As shown in Table 10 all facial expression representation achieve a similar recognition rate for the same classifier with the highest rate for the manually selected landmarks. The manually selected landmarks are included only for a reference for other automatic methods. The recognition rates obtained by the automatic methods are lower (maximum 4% less for deformation field based representation) than that obtained by manual landmark selection.

The confusion matrices for LDA for different data are given in Tables 5 - 8. From the classification performance, it can be concluded that the surprise, disgust, happiness and sadness expressions can be classified in most cases with above 75% accuracy, anger with about 70% accuracy, whereas fear is only classified correctly in 58% . The best recognition rates (about 90%) are found for surprise, similarly as for work in (Quan et al., 2009) for data sets taken the same database.

The results of misclassification support the conclusion of the separability analysis conducted in the previous section. The pair of expressions with low value of separability criterion $J_2^{ex_i,ex_j}(\vec{x})$ are more prompt to be misclassified (e g fear and sadness). The expression of fear achieves low values of separability criterion $J_2^{ex_i,ex_j}(\vec{x})$ for each facial expression representation and as it is expected the misclassification error is the highest. The expressions with high value of separability criterion $J_2^{ex_i,ex_j}(\vec{x})$ achieve recognition rates (e g happiness, or surprise).

Table 9 summaries the diagonal of Tables 5 - 8. Taking into account the "subjective" nature of the ground truth data (Quan et al., 2009), the results can be considered as reasonable.

4 CONCLUSIONS

A statistical analysis of different facial expression representations based on the Log-Euclidean statistics has been presented in this paper. The proposed method generates first the *mean* face by simultaneous registration of faces with neutral expression included in the training data set, thereby enabling all faces to be mapped to the common face space based on the estimated transformations. The obtained results show that the Space Shape Vectors built based on the velocity fields can be consider as an effective facial expression representation for the Statistical Shape Model. 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

Arsigny, V., Commowick, O., Pennec, X., and Ayache, N. (2006). A log-euclidean framework for statistics on diffeomorphisms. *Medical Image Computing and Computer Assisted Intervention*, 9(Pt 1):924–931.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95–113.

Bartlett, M. S., Littlewort, G., Fasel, I., and Movellan, J. R. (2003). Real time face detection and facial expression recognition: Development and applications to human computer interaction. In *In CVPR Workshop on CVPR for HCI*.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Bossa, M., Hernandez, M., and Olmos, S. (2007). Contributions to 3d diffeomorphic atlas estimation: application to brain images. *Med Image Comput Comput Assist Interv*, 10(Pt 1):667–674.

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models - their training and application. *Computer Vision. Image Understanding*, 61:38–59.

Fasel, B. and Luettin, J. (2003). Automatic facial expression analysis: A survey. *PATTERN RECOGNITION*, 36(1):259–275.

Geng, X., Christensen, G. E., Gu, H., Ross, T. J., and Yang, Y. (2009). Implicit reference-based group-wise image registration and its application to structural and functional mri. *Neuroimage*, 47(4):1341–1351.

Han, X., Hibbard, L. S., and Willcut, V. (2010). An efficient inverse-consistent diffeomorphic image registration method for prostate adaptive radiotherapy. In *Proceedings of the 2010 international conference on Prostate cancer imaging: computer-aided diagnosis, prognosis, and intervention*, MICCAI'10, pages 34–41, Berlin, Heidelberg. Springer-Verlag.

Hsieh, C.-K., Lai, S.-H., and Chen, Y.-C. (2010). An optical flow-based approach to robust face recognition under expression variations. *IEEE Transactions on Image Processing*, 19(1):233–240.

Kobayashi, H. and Hara, F. (1997). Facial interaction between animated 3d face robot and human beings. In *Proc. IEEE Int Systems, Man, and Cybernetics Computational Cybernetics and Simulation. Conf*, volume 4, pages 3732–3737.

Matuszewski, B. J., Quan, W., and Shark, L.-K. (2011). *Biometrics - Unique and Diverse Applications in Nature, Science, and Technology*, chapter Facial Expression Recognition. InTech.

Pantic, M., Member, S., and Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1424–1445.

Papiez, B. W. and Matuszewski, B. J. (2011). Direct inverse deformation field approach to pelvic-area symmetric image registration. In *Proceedings Medical Image Understanding and Analysis (MIUA'2011)*.

Quan, W., Matuszewski, B. J., and Shark, L.-K. (2010). Improved 3-d facial representation through statistical shape model. In *17th IEEE International Conference on Image Processing (ICIP 2010)*, pages 2433–2436.

Quan, W., Matuszewski, B. J., Shark, L.-K., and Ait-Boudaoud, D. (2007a). 3-d facial expression representation using b-spline statistical shape model. In *Proceedings of the Vision, Video and Graphics Workshop*.

Quan, W., Matuszewski, B. J., Shark, L.-K., and Ait-Boudaoud, D. (2007b). Low dimensional surface parameterisation with applications in biometrics. In *Proceedings of the International Conference on Medical Information Visualisation - BioMedical Visualisation*, pages 15–22, Washington, DC, USA. IEEE Computer Society.

Quan, W., Matuszewski, B. J., Shark, L.-K., and Ait-Boudaoud, D. (2009). Facial expression biometrics using statistical shape models. *EURASIP Journal on Advances in Signal Processing*, 2009:15:4–15:4.

Shan, C., Gong, S., and McOwan, P. W. (2005). Robust facial expression recognition using local binary patterns. In *ICIP (2)*, pages 370–373.

Tian, Y.-L., Kanade, T., and Cohn, J. F. (2011). *Handbook of Face Recognition*, chapter Facial Expression Analysis. Springer.

Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2008). Symmetric log-domain diffeomorphic registration: a demons-based approach. *Medical Image Computing and Computer Assisted Intervention*, 11(Pt 1):754–761.

Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1, Supp.1):S61–S72.

Wang, J. and Yin, L. (2007). Static topographic modeling for facial expression recognition and analysis. *Computer Vision and Image Understanding*, 108(1-2):19–34.

Yin, L., Wei, X., Sun, Y., Wang, J., and Rosato, M. J. (2006). A 3d facial expression database for facial behavior research. In *Proc. 7th Int. Conf. Automatic Face and Gesture Recognition FGR 2006*, pages 211–216.