CONSTRAINED LOCAL MODEL WITH INDEPENDENT COMPONENT ANALYSIS AND KERNEL DENSITY ESTIMATION: APPLICATION TO DOWN SYNDROME DETECTION

Qian Zhao[1], *Kazunori Okada*[2], *Kenneth Rosenbaum*[3], *Marshall Summar*[3], *Marius George Linguraru*[1,4]

[1] Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Health System, Washington DC
[2]Computer Science Department, San Francisco State University, San Francisco CA
[3]Division of Genetics and Metabolism, Children's National Health System, Washington DC
[4]School of Medicine and Health Sciences, George Washington University, Washington DC

ABSTRACT

Statistical shape models generally characterize shape variations linearly by principal component analysis (PCA), which assumes that the non-rigid shape parameters are drawn from a Gaussian distribution. This practical assumption is often not valid. Instead, we propose a constrained local model based on independent component analysis (ICA) and use kernel density estimation (KDE) for non-parametrically modeling the distribution of the shape parameters. The model fitting is achieved by maximum a posteriori via the expectation-maximization algorithm and results in a mean shift-like update optimizer. The proposed approach is capable of modeling non-Gaussian shape priors and significantly outperformed the PCA-based model ($p=0.03$) and ICA-based model with Gaussian shape prior ($p=0.01$) in experiments to detect facial landmarks. Moreover, we applied the model to Down syndrome detection from frontal facial photographs and obtained higher accuracy than the best results reported in literature.

Index Terms— Constrained Local Model, Non-Parametric Shape Prior, Kernel Density Estimation, Independent Component Analysis, Down syndrome

1. INTRODUCTION

Statistical shape models (SSM) are popular techniques for interpreting shape variations through analysis of training samples. Principal component analysis (PCA) is most commonly used to describe shape variations in the form of point distribution model (PDM) [1]. However, PCA-based models assume a Gaussian distribution of shape variations, which is often not valid and may lead to inaccurate statistical descriptions of shapes and generation of implausible shape instances. Moreover, the principal components (PCs) of PCA tend to represent global shape variations [2]. Independent component analysis (ICA) is another popular method to model non-Gaussian and localized statistical shape variations, addressing the above issues. To the best of our knowledge, the studies of PDM with ICA are scarce. Uzümcü et al. compared SSM with ICA vs. PCA, where the independent components (ICs) are ordered based on the locality of the shape variations [2]; but their SSM fitting simply used the ICA's linear projection and they constrained plausible shapes within an empirical shape variation range. Zhao et al. investigated constrained local model (CLM) with ICA by optimizing the appearance likelihood from local patches and shape reconstruction error introduced by ICA [3]; they assumed the shape prior to be uniformly distributed. ICA-based SSM were also investigated in 3D and 4D cardiac MR images [4, 5] using the method in [2].

Our key claim is that modeling the shape prior in ICA-based PDM by using a generalized non-Gaussian distribution improves the accuracy of shape modeling. This claim is theoretically justified by the definition of ICA to learn a linear basis that maximizes the *non-Gaussianity* of model parameters [6].

As our novel methodological contribution, we estimate the distribution of shape parameters using non-parametric kernel density estimation (KDE) in the ICA-based PDM and embed it to a CLM to locate facial landmarks. We propose a new updated formula for CLM fitting by optimizing the maximum *a posteriori* (MAP) objective function with the expectation-maximization (EM) algorithm. The resulting formula takes a form similar to the mean shift optimizer when solving the M-step by Gauss-Newton method. The previous CLM methods in literature [7-9] employed a Gaussian shape prior in their PCA-based PDM, which fall short of accurately describing the multi-modal shape variations. Other researchers introduced Gaussian mixtures to model the multi-modal distribution of shape parameters [10, 11], where the number of Gaussians were fixed to a value selected empirically. However, choosing an appropriate number of Gaussian components is not trivial and is data and domain-specific.

The proposed KDE-based ICA-CLM (ICA-KDE) method is validated by applying it to the automatic identification of facial landmarks and detection of Down syndrome (DS) from facial photography. DS is the most common chromosomal condition; one out of 691 infants is born with DS every year in the US [12]. DS causes lifelong mental retardation, heart defects and respiratory problems and its early detection is fundamental for managing DS and providing patients with lifelong medical care. ICA-KDE significantly improves the automatic identification of facials

landmarks (p=0.03/0.01, respectively) when compared with PCA/ICA-based models with Gaussian priors, and detects DS with 97.8% accuracy, which is an improvement of the best accuracy reported in [3].

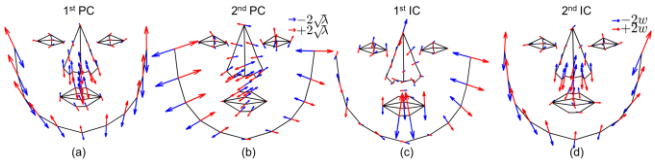

Fig. 1. The first two principal modes obtained by PCA (a, b) and independent modes obtained by ICA (c, d) of facial variation, where $\boldsymbol{w}$ is the empirical standard variation of training samples and λ the eigenvalues of PCA.

2. MATERIALS AND METHODS

2.1. Data

We employed the dataset used in [3] for facilitating simple comparison. The dataset contains 100 frontal facial images (one image per subject) including 50 DS patients and 50 healthy individuals, who were acquired with various illuminations, resolutions, expressions and poses. The subjects are from multiple ethnicities and both genders. The age of patients varies from 0 to 36 month. Forty-four anatomical landmarks were manually placed on each image on the eyes (10), nose (14), mouth (9), and along the contour of the face (11). The manual landmarks were used as ground truth in the evaluation of landmark detection.

2.2. ICA-based CLM with KDE

Building ICA-based CLM: A CLM consists of a PDM and a local texture model. The PDM describes the shape variation of the training samples and the local texture model characterizes the region appearance around each landmark. We denote a shape with n landmarks in two dimensions as $\boldsymbol{x} = [x_1, y_1, x_2, y_2 \cdots x_n, y_n]^T$. All shapes X are first aligned using Procrustes analysis [13] to remove translation, rotation and scaling. Then the non-rigid shape variation is modeled using ICA: $X = \bar{X} + A \cdot S$, where A is the mixing matrix and S the independent components (ICs). The ICs, i.e. the shape parameters, are computed as $S = W\left(X - \bar{X}\right)$, where $W = A^{-1}$ is the de-mixing matrix. ICA is defined as the method that finds a linear transformation A maximizing the non-Gaussianity of S. In this study, we use a robust algorithm to estimate A, W and S by iterative maximization of the kurtosis contrast with algebraic optimal step size [6], and select ICs following the method proposed in [3] by using entropy and the interquartile range to measure the sample variation and locality. The complete PDM consists of a global rigid transformation and non-rigid variation: $\boldsymbol{x} = sR\left(\bar{\boldsymbol{x}} + A\boldsymbol{q}\right) + \boldsymbol{t}$, where s, R and $\boldsymbol{t}$ denote the scaling, rotation matrix and translation, respectively. We denote the shape parameters as $\boldsymbol{p} = \{s, R, \boldsymbol{t}, \boldsymbol{q}\}$, including both the rigid shape parameters $\boldsymbol{p}_{rigid} = \{s, \theta, t_x, t_y\}$ and non-rigid shape parameters $\boldsymbol{q}$. The first two principal (PC) and independent modes (IC) of facial variation built from DS and healthy cases combined are compared in Fig. 1.

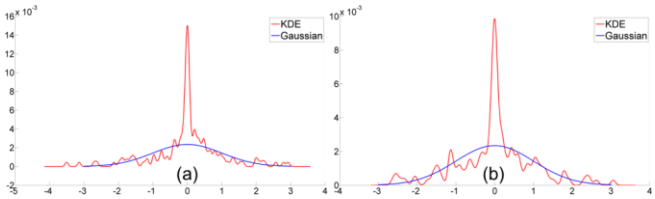

Fig. 2. The distribution of shape parameters estimated by Gaussian and KDE for (a) the first IC and (b) the second IC obtained by ICA. See Fig. 1 for modes of variation.

The local texture model is built with linear support vector machines (SVM) as described in [3, 8]. The direct use of gray scale intensities as descriptors may include noise and we improve the model by using the histogram of oriented gradients (HoG) [14] as patch descriptor to remove noise and enhance edges. For HoG descriptors, the parameters including the number of orientation bins, cell size and block size were set to 9, 8 and 2, respectively. Thus for m extracted patches in an image, we obtain m training vectors $[\boldsymbol{g}_1, \boldsymbol{g}_2, \cdots, \boldsymbol{g}_m]$ with assigned output SVM values $o_i = \{-1, 1\}, i = 1, 2, \ldots, m$ .The decision function of the linear SVM can be represented by the linear combination of the input vectors $f(\boldsymbol{g}_i) = \varphi^T \cdot \boldsymbol{g}_i$, where φ is the weight for each input vector and used as the local texture model.

Fitting ICA-based CLM with KDE: The CLM fitting objective can be interpreted probabilistically as the MAP of the shape parameters. The specific form of the objective function implicitly assumes conditional independence among the detections of each landmark

$$p(\mathbf{p}|\{l_i = 1\}_{i=1}^n, \mathcal{I}) \propto p(\mathbf{p}) \prod_{i=1}^{n} p(l_i = 1|\mathbf{x}_i, \mathcal{I}), \qquad (1)$$

where $l_i = \{-1, 1\}$ is a discrete variable denoting whether the i^{th} landmark $\mathbf{x}_i = (x_i, y_i)$ is misaligned or aligned, $p(l_i = -1|\mathbf{x}_i, \mathcal{I}) + p(l_i = 1|\mathbf{x}_i, \mathcal{I}) = 1$, and $\mathcal{I}$ is the image. If the shape parameters are assumed to be sampled from an underlying distribution, the objective function leads to MAP estimation, otherwise it leads to a maximum likelihood estimation by assuming a uniform shape prior.

In this study, we propose to use KDE to approximate the probability density function (*pdf*) of the non-rigid shape parameters. Fig. 2 shows the *pdf* of the first and second IC in Fig. 1 estimated by Gaussian and KDE. A uniform prior is commonly placed on the rigid transformation, which assumes all rigid shape parameters are equally likely.

Following this convention, the shape prior can be written as a KDE given a set of training shape parameters $\mathbf{q}_j,\ j = 1, 2, \ldots, N$:

$$p(\mathbf{p}) \propto p(\mathbf{q}) = \frac{1}{N}\sum_{j=1}^{N} \mathcal{K}(\mathbf{q}-\mathbf{q}_j, \Sigma),\ \mathcal{K}(\mathbf{x}, \Sigma) = |2\pi\Sigma|^{-1/2} e^{\left(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right)}, \quad (2)$$

where d is the dimension of the data ($d = |\mathbf{q}|$ here), N the number of training samples, Σ the bandwidth (smoothing parameter) and $\mathcal{K}(\cdot)$ the multivariate Gaussian kernel. Estimating the bandwidth of the kernel function is not a trivial problem. Here we estimate Σ as the homoscedastic covariance $h^2\mathbf{I}$ by using the nearest neighbor distance: $h^2 = \frac{1}{N}\sum_{i=1}^{N} min_{j\neq i}|\mathbf{q}_i - \mathbf{q}_j|^2$ [15].

Following the method in [8], the data likelihood term in (1) is also modeled using Gaussian kernel density estimation $p(l_i = 1|\mathbf{x}_i, \mathcal{I}) = \sum_{\mathbf{y}_i\in\mathbf{\Psi}_i} \pi_{\mathbf{y}_i}\mathcal{K}(\mathbf{x}_i - \mathbf{y}_i, \rho\mathbf{I})$, where $\mathbf{x}_i, \mathbf{y}_i$ denote the i^{th} landmark and its ground truth, respectively, and $\pi_{\mathbf{y}_i} = p(l_i = 1|\mathbf{y}_i, \mathcal{I})$. The true location of the landmark $\mathbf{y}_i$ is treated as a hidden variable in a set of candidate locations $\{\Psi_i\}_{i=1}^{n}$ and $\rho\mathbf{I}$ is a homoscedastic isotropic Gaussian covariance. Substituting (2) into (1), the objective function becomes

$$p(\mathbf{p}|\{l_i = 1\}_{i=1}^{n}, \mathcal{I}) \propto \frac{1}{N}\sum_{j=1}^{N} \mathcal{K}(\mathbf{q}-\mathbf{q}_j, \Sigma) \cdot \prod_{i=1}^{n}\sum_{\mathbf{y}_i\in\mathbf{\Psi}_i} \pi_{\mathbf{y}_i}\mathcal{K}(\mathbf{x}_i - \mathbf{y}_i, \rho\mathbf{I}). \quad (3)$$

We then derive a maximizer of the above objective function by using the EM algorithm. In the E-step, the *posterior* over the candidates is evaluated as

$$\omega_{\mathbf{y}_i} = p(\mathbf{y}_i|l_i = 1, \mathbf{x}_i, \mathcal{I}) = \frac{\pi_{\mathbf{y}_i}\mathcal{K}(\mathbf{x}_i - \mathbf{y}_i, \rho\mathbf{I})}{\sum_{\mathbf{z}_i\in\mathbf{\Psi}_i}\pi_{\mathbf{z}_i}\mathcal{K}(\mathbf{x}_i - \mathbf{z}_i, \rho\mathbf{I})}. \quad (4)$$

Then the M-step minimizes the following auxiliary function

$$\begin{aligned} Q(\mathbf{p}) &= \mathbb{E}\left[-\log\left(p(\mathbf{p})\prod_{i=1}^{n} p(l_i = 1, \mathbf{y}_i|\mathbf{x}_i, \mathcal{I})\right)\middle|\mathcal{Y}\right] \\ &\propto -\log\rho\left(\sum_{j=1}^{N}\mathcal{K}(\mathbf{q}-\mathbf{q}_j, \Sigma)\right) + \sum_{i=1}^{n}\sum_{\mathbf{y}_i\in\mathbf{\Psi}_i}\omega_{\mathbf{y}_i}(\mathbf{x}_i - \mathbf{y}_i)^T(\mathbf{x}_i - \mathbf{y}_i). \quad (5) \end{aligned}$$

The function in (5) is minimized using the Gauss-Newton optimization that yields the following form of the shape parameter update

$$\Delta\mathbf{p} = -\mathbf{H}(\mathbf{p})^{-1}\mathbf{g}(\mathbf{p}) = -\mathbf{H}(\mathbf{p})^{-1}\left(-\frac{\rho}{h^2}\mathbf{m}(\mathbf{q}) - \mathbf{J}^T\mathbf{v}\right), \quad (6)$$

$$\mathbf{H} = \rho\Sigma^{-1}\frac{\sum_{j=1}^{N}(\mathcal{K}_j(1+\log\mathcal{K}_j))}{\sum_{j=1}^{N}\mathcal{K}_j} - \rho\left(\frac{\mathbf{m}_j}{h^2}\right)^2 + \mathbf{J}^T\mathbf{J}, \quad (7)$$

$$\mathbf{v}_i = \sum_{\mathbf{y}_i\in\mathbf{\Psi}_i}\omega_{\mathbf{y}_i}\mathbf{y}_i - \mathbf{x}_i^c,\ \mathbf{m}_j(\mathbf{q}) = \frac{\sum_{j=1}^{N}\mathbf{q}_j\mathcal{K}_j}{\sum_{j=1}^{N}\mathcal{K}_j} - \mathbf{q}, \quad (8)$$

where $\mathbf{H}$ is the Gauss-Newton Hessian, $\mathbf{v}_i, \mathbf{m}_j$are the mean shift vectors for shape prior and likelihood term, respectively, and **J** the Jacobian of the PDM.

For comparison, we also built CLM with Gaussian prior based on both PCA and ICA following the method in [8].

2.3. Feature Extraction, Selection and Classification

Following [3], we extract 27 geometric features including normalized Euclidean distance and angles and 132 texture features based on local binary patterns [16]. These features are combined then undergone feature selection. Finally, the DS patients are separated from the healthy groups using support vector machines (SVM) [14] with RBF kernel, linear SVM, *k* nearest neighbor (*k*-NN) [15], random forests (RF) [16] and linear discriminant analysis (LDA) [17]. The parameters for classifiers are found optimally by grid search. Leave-one-subject-out validation was conducted throughout the framework, including SVM texture model training, CLM building, feature selection and classification, performed for each fold separately.

3. EXPERIMENTS AND RESULTS

A CLM-based Down syndrome detection system [3] is used as our baseline system. The performance of the proposed ICA-KDE method to detect facial landmarks was compared with results obtained with PCA-based CLM (PCA) and ICA-based CLM with Gaussian shape prior (ICA-Gaussian).The validation metric was the point-to-point distance between the landmarks obtained by the automated method and the ground truth normalized by the distance between the pupils. The normalized errors were 0.038±0.028, 0.040±0.036 and 0.036±0.025 for PCA, ICA-Gaussian and ICA-KDE, respectively. The comparison of different models is shown in Fig. 3. Significance was computed using the Wilcoxon rank-sum test [17] based on results obtained on the detection of all facial landmarks. A significant improvement was recorded by using ICA-KDE vs. PCA (p=0.026) and by using ICA-KDE vs. ICA-Gaussian (p=0.010). ICA-KDE outperformed the other methods on average and on all anatomical areas, except for the nose region. The error on the nose was larger by using ICA-based methods than PCA due to the lack of non-prominent texture features at the top of the nose. The performance of ICA-Gaussian was worse than PCA which may be caused by the inappropriate shape prior imposed on ICA-based PDM. The average time for analyzing one new case with ICA-KDE (30 iterations) was 39.14s using MATLAB on a Windows 8 core workstation with 12GB RAM, shown in Fig. 3 (b).

Next, we evaluated the proposed ICA-KDE method in terms of accuracy, precision and recall for DS detection with different model fitting methods. The experimental results are shown in Table 1. For DS detection, ICA-KDE outperformed PCA and ICA-Gaussian models. The best performance of 97.8% accuracy with 97.7% precision and 97.7% recall was achieved by using linear SVM and 72 selected combined features. The area under the receiver operating characteristic (AUROC) was 0.997. The accuracy, precision and recall were improved by 2.2%, 2.4% and 2.4% from those reported by Zhao et al. [3] using the same data, respectively. This improvement is the equivalent of 50% reduction of misclassification rate over the state-of-the-art in the tight space left for performance improvement.

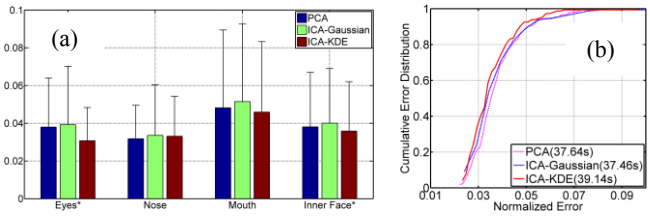

Fig. 3. Comparison of PCA, ICA- Gaussian and ICA- KDE: (a) the normalized landmark detection errors for eyes, nose, mouth and inner facial landmarks; * indicates the significantly improved cases on eyes and inner face; (b) the cumulative error distribution curves.

4. CONCLUSION

We proposed an ICA-based CLM with the shape prior estimated by KDE instead of a Gaussian. The model fitting was achieved by optimizing a maximum a posteriori via EM algorithm, where the M-step solved by Gauss-Newton yielded a mean shift-like update iterator. The proposed method significantly outperformed PCA-based CLM ($p=0.03$) and ICA-based CLM with Gaussian shape prior ($p=0.01$). When applied to DS detection, the ICA-KDE achieved 97.8% accuracy that was 2.2% higher than the previous state-of-the-art. One of the weaknesses of the method is the increased but acceptable computational complexity and time (39.1 s per new case based on 30 iterations). Another question is the automatic estimation of KDE's bandwidth, which strongly influences the optimization results. We will investigate other more advanced methods to estimate the bandwidth of KDE in the future. The promising results of our general methodology encourage us to further investigate the detection of other types of facial dysmorphology and use in other medical and non-medical applications.

Acknowledgement. This work was supported by a gift from the Government of Abu Dhabi to Children's National Health System.

5. REFERENCES

[1] T. F. Cootes, *et al.*, "Active Shape Models-Their Training and Application," *Computer Vision and Image Understanding,* vol. 61, pp. 38-59, 1995.

[2] M. Uzümcü, *et al.*, "ICA vs. PCA Active Appearance Models: Application to Cardiac MR Segmentation," in *Proc. MICCAI 2003, LNCS 2878*, 2003.

[3] Q. Zhao, *et al.*, "Hierarchical Constrained Local Model Using ICA and Its Application to Down Syndrome Detection," in *Proc. MICCAI 2013*. vol. 8150, 2013, pp. 222-229.

[4] S. Faghih Roohi and R. A. Zoroofi, "4D statistical shape modeling of the left ventricle in cardiac MR images," *International Journal of Computer Assisted Radiology and Surgery,* pp. 1-17, 2012.

[5] J. Lötjönen, *et al.*, "Statistical shape model of atria, ventricles and epicardium from short- and long-axis MR images," *Medical image analysis,* vol. 8, pp. 371-386, 2004.

[6] V. Zarzoso and P. Comon, "Robust Independent Component Analysis by Iterative Maximization of the Kurtosis Contrast With Algebraic Optimal Step Size," *IEEE Transactions on Neural Networks,* vol. 21, pp. 248-261, 2010.

[7] D. Cristinacce and T. Cootes, "Automatic feature localisation with constrained local models," *Pattern Recognition,* vol. 41, pp. 3054-3067, 2008.

[8] J. M. Saragih, *et al.*, "Face alignment through subspace constrained mean-shifts," in *2009 IEEE 12th International Conference on Computer Vision,* pp. 1034-1041, 2009.

[9] W. Yang, *et al.*, "Enforcing convexity for improved alignment with constrained local models," in *CVPR 2008.*, pp. 1-8, 2008.

[10] T. F. Cootes and C. J. Taylor, "A mixture model for representing shape variation," *Image and Vision Computing,* vol. 17, pp. 567-573, 1999.

[11] K. Jang, *et al.*, "Lip Localization Based on Active Shape Model and Gaussian Mixture Model," in *Advances in Image and Video Technology*. vol. 4319, pp. 1049-1058, 2006.

[12] S. E. Parker, *et al.*, "Updated national birth prevalence estimates for selected birth defects in the United States, 2004–2006," *Birth Defects Research Part A: Clinical and Molecular Teratology,* vol. 88, pp. 1008-1016, 2010.

[13] J. C. Gower, "Generalized procrustes analysis," *Psychometrika,* vol. 40, pp. 33-51, 1975.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *CVPR 2005*, pp. 886-893 vol. 1, 2005.

[15] J. M. Leiva and A. Artés, "Algorithms for maximum-likelihood bandwidth selection in kernel density estimators," *Pattern Recogn. Lett.,* vol. 33, pp. 1717-1724, 2012.

[16] T. Ojala, *et al.*, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 971-987, 2002.

[17] H. Toutenburg, "Hollander, M., D. A. Wolfe: Nonparametric statistical methods. John Wiley & Sons, New York-Sydney-Tokyo-Mexico City 1973. 503 S., $9.50," *Biometrische Zeitschrift,* vol. 17, pp. 526-526, 1975.

Table 1. Comparison of different classifiers for DS detection

	PCA	ICA-Gaussian	ICA-KDE	
SVM-RBF	0.934	0.934	0.967	**Accuracy**
	0.930	0.911	0.976	**Precision**
	0.930	0.953	0.953	**Recall**
Linear SVM	0.956	0.901	**0.978**	**Accuracy**
	0.953	0.886	**0.977**	**Precision**
	0.953	0.907	**0.977**	**Recall**
***k*-NN**	0.912	0.945	0.956	**Accuracy**
	0.927	0.952	0.976	**Precision**
	0.884	0.930	0.930	**Recall**
RF	0.912	0.901	0.923	**Accuracy**
	0.927	0.886	0.929	**Precision**
	0.884	0.907	0.907	**Recall**
LDA	0.956	0.967	0.945	**Accuracy**
	0.953	0.976	0.952	**Precision**
	0.953	0.953	0.930	**Recall**