Hierarchical Constrained Local Model Using ICA and Its Application to Down Syndrome Detection

Qian Zhao[1], Kazunori Okada[2], Kenneth Rosenbaum[3], Dina J. Zand[3], Raymond Sze[1,4], Marshall Summar[3], and Marius George Linguraru[1]

[1] Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Medical Center, Washington, DC
[2] Computer Science Department, San Francisco State University, San Francisco, CA
[3] Division of Genetics and Metabolism, Children's National Medical Center, Washington, DC
[4] Department of Radiology, Children's National Medical Center, Washington, DC

Abstract. Conventional statistical shape models use Principal Component Analysis (PCA) to describe shape variations. However, such a PCA-based model assumes a Gaussian distribution of data. A model with Independent Component Analysis (ICA) does not require the Gaussian assumption and can additionally describe the local shape variation. In this paper, we propose a Hierarchical Constrained Local Model (HCLM) using ICA. The first or coarse level of HCLM locates the full landmark set, while the second level refines a relevant landmark subset. We then apply the HCLM to Down syndrome detection from photographs of young pediatric patients. Down syndrome is the most common chromosomal condition and its early detection is crucial. After locating facial anatomical landmarks using HCLM, geometric and local texture features are extracted and selected. A variety of classifiers are evaluated to identify Down syndrome from a healthy population. The best performance achieved 95.6% accuracy using support vector machine with radial basis function kernel. The results show that the ICA-based HCLM outperformed both PCA-based CLM and ICA-based CLM.

Keywords: hierarchical constrained local model, independent component analysis, Down syndrome detection, classification.

1 Introduction

Conventional statistical models such as Active Shape Model (ASM) [1], Active Appearance Model (AAM) [2] and other variants have been widely applied in face alignment, expression recognition and medical image interpretation. The Constrained Local Model (CLM) first proposed by Cristinacce and Cootes [3] also uses a template appearance model with a more robust constrained search technique. CLM has demonstrated good performance in non-rigid object alignment. Most of these techniques describe the principal modes of shape variation in the training dataset using Principal Component Analysis (PCA). However, PCA-based shape models assume a Gaussian distribution of the input data which is often not valid and may lead to the inaccurate

K. Mori et al. (Eds.): MICCAI 2013, Part II, LNCS 8150, pp. 222–229, 2013.

statistical description of shapes and generation of implausible shapes. Furthermore, the principal components (PCs) of PCA tend to represent global shape variations: changing parameter value of a PC may deform the entire extent of shape, failing to capture localized deformations that are important for model discrimination in certain medical applications. To address the above problem, Independent Component Analysis (ICA) is considered in this study as an alternative method to build a statistical shape model. To the best of our knowledge, previous work related to constructing statistical models using ICA is scarce. In [4], the authors compared the ICA and PCA in AAM for cardiac MR segmentation. The ICA-based AAM outperformed PCA-based model in terms of border localization accuracy. But they did not present how to select a relevant subset of ICs.

In this study, we propose a Hierarchical Constrained Local Model (HCLM) based on ICA and apply it to detect Down syndrome from facial photographs. Down syndrome (DS) is the most common chromosomal condition; one out of 691 infants born with DS every year in the US [5]. DS causes lifelong mental retardation, heart defects and respiratory problems. The early detection and intervention of DS is fundamental for managing the disease and providing patients lifelong medical care.

DS may be diagnosed before or after birth. During pregnancy, screening and diagnostic tests can be performed. The accuracy of screening tests is estimated to be 75% [6]. After birth, the initial diagnosis of DS is often based on a number of minor physical variations and malformations. Some common features include a flattened facial profile, upward slanting eyes and protruding tongue. These differences may be subtle and are influenced by a variety of factors, frequently making a rapid, accurate diagnosis difficult. The accuracy of a clinical diagnosis of DS for pediatricians prior to cytogenetic results approximates 50%-60% and is likely to be lower in many instances [7]. The development of automated remote computer-aided diagnosis of DS has the potential for dramatically improving the diagnostic rate and providing early guidance to families and involved professionals. Recently, facial image analysis methods were investigated for DS detection [8, 9]. However, they all require manual pre-processing and do not incorporate any disease specific information.

In this study, we present a simple, automated and non-invasive assessment method for DS based on anatomical landmark analysis and machine learning techniques. After locating landmarks using HCLM with ICA, syndrome-specific geometric features and local texture features based on local binary patterns (LBP) are extracted and selected. Then various classifiers including support vector machine (SVM), k-nearest neighbor (kNN), random forest (RF) and linear discriminant analysis (LDA) are compared for the identification of DS.

Our main contributions lie in: (a) the proposal of a hierarchical constrained local model to locate facial landmarks; (b) investigation of constrained local model using ICA to describe local shape variations (c) the proposal of a data-driven ordering method for sorting the independent components; (d) construction of separate models for Down syndrome and a healthy group to refine the locations of the relevant facial landmarks (e) combination of the syndrome specific geometric features and local texture features to characterize pathology.

2 Methods

2.1 Hierarchical Constrained Local Model (HCLM)

A hierarchical constrained local model is proposed to locate facial landmarks. The HCLM consists of two levels. The first level CLM is trained by using the full landmark set (inner and outer face) and facial images including both healthy and DS populations. It roughly locates all the facial landmarks. For the second level, two separate CLMs are trained using inner facial landmarks for the healthy and DS groups, respectively. It refines the locations of inner face landmarks which are clinically relevant for diagnosis, therefore important for feature extraction. The overall framework for HCLM is shown in Fig.1.

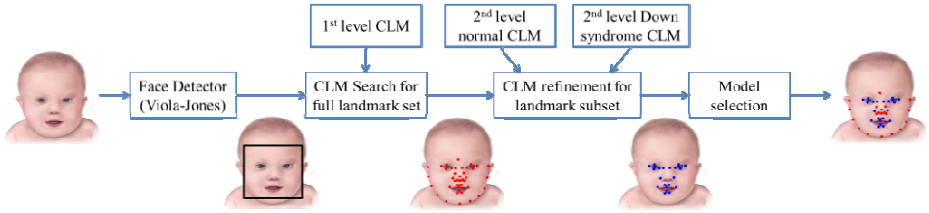

Fig. 1. The framework of hierarchical constrained local model (HCLM)

Building HCLM with ICA

The CLM consists of a shape model and a patch model. The shape model defines the plausible shape domain and describes how face shape can vary. The patch model describes how the local region around each facial feature point should look like. With these two models, both face morphology and textures are described. We denote an n-point shape in two dimensions as $\boldsymbol{x} = [x_1, y_1, x_2, y_2 \cdots x_n, y_n]^T$. To study the shape variation in the training data, all shapes are aligned to each other using Procrustes analysis. Then the mean shape is subtracted from each aligned shape of the training set that is represented by the vector $\boldsymbol{x}$, where $\boldsymbol{x}$ now contains the new zero-mean coordinates resulting from alignment. After this pre-processing, the statistical shape model is built using ICA. The shape matrix X containing all the training shapes can be written as $X = A \cdot S$, where A is the mixing matrix containing the mixing parameters and S the source shapes. It can also be written in a vector format $\boldsymbol{x} = \sum \boldsymbol{a}_i s_i$, where $\boldsymbol{a}_i$ is the columns of A and s_i is the ith independent component (IC). After estimating the matrix A, the de-mixing matrix W and ICs can be computed by $S = W \cdot X$. The de-mixing matrix can be computed by maximizing some measure of independency. In this study, we use the Joint Approximated Diagonalization of Eigenmatrices (JADE) method [10] as suggested by [4].

For PCA, the eigenvectors are sorted according to their corresponding variances naturally, while for ICA, the variances and the order of ICs are not determined naturally. We propose to sort the columns of the mixing matrix A based on non-parametric estimate of variance and locality. First the shapes X are projected onto

each $\boldsymbol{a}_i$. A histogram and its corresponding normalized cumulative histogram (CH) are computed from these projections. Then the histogram width ω_i is determined by the range spanning the total 96% histogram area (from $\mathrm{CH}^{-1}(0.02)$ to $\mathrm{CH}^{-1}(0.98)$). The width can be regarded as a robust non-parametric estimate of sample variance along ICs. The shape variation along $\boldsymbol{a}_i$ to the limit $\omega_i/2$ is given by $\boldsymbol{v}_i = |\boldsymbol{a}_i \cdot \omega_i|$. The criterion C to order the ICs is then determined by

$$C = \omega_i \cdot v_{\max} / H(\boldsymbol{v}_i), \tag{1}$$

where $v_{\max}$ is the maximum value of $\boldsymbol{v}_i$ and $H(\boldsymbol{v}_i)$ is the entropy of $\boldsymbol{v}_i$. For modes that describe relevant independent directions in the training data, the variations are localized and have large peaks, therefore have large C values. While for noisy modes, the variations are relatively small and not localized, thus have small C values. After sorting the ICs with this criterion, noisy ICs with very small C values are removed. Fig. 2 compares the first three principal modes for PCA and ICA. Note that PCA modes depict only global variations, while ICA modes highlight local variations.

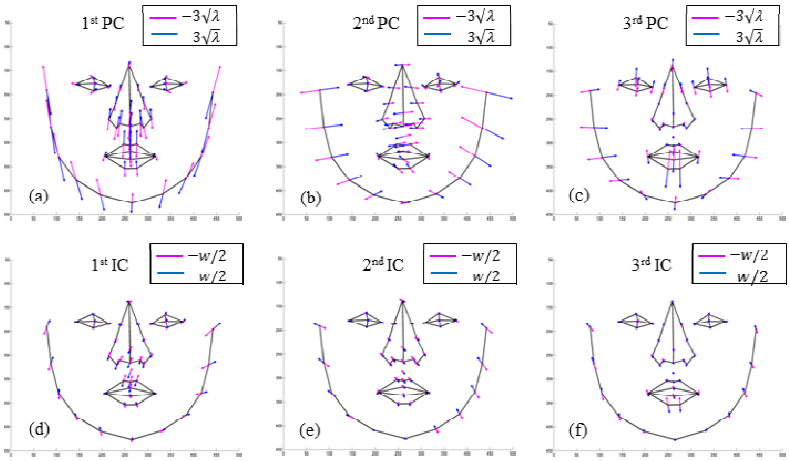

Fig. 2. The first three principal modes for PCA (a-c) and ICA (d-f)

The patch model is built by using SVM with a linear kernel. For each landmark, we extract 25 (40×40) square patch samples from each image as training data for SVM, containing both negative and positive examples. A SVM is trained for each landmark. For CLM, the output of SVM with linear kernel can be written as a linear combination of the input vector $y^{(i)} = \boldsymbol{\alpha}^T \cdot p^{(i)} + \theta$, where $\boldsymbol{\alpha}^T = [\alpha_1 \alpha_2 \ldots \alpha_N]$ represents the weight for each input pixel $p^{(i)}$, and θ is a bias. The weight matrix is used as the patch model. We note that the linear SVM can be implemented efficiently by convolution. Therefore, it reduces the computational complexity and time dramatically by avoiding sliding window process.

Searching Landmarks with HCLM

The CLM model is built to search landmarks around their local region. First, we detect the face, eyes and tip of the nose in the image by using Viola-Jones face detector [11] to initialize the first level HCLM. Then each landmark is searched in the local region of its current position using the patch model. We denote the SVM response image with $R(x,y)$ which is fitted with a quadratic function $r(x,y)=zHz^T-2Fz^T+ax_0^2+by_0+c$, where $z=[x\ y]$, $H=diag(a,b)$, $F=[ax_0\ by_0]$, a,b are the quadratic function parameters and (x_0,y_0) is the center point. The parameters are solved by a least square optimization.

Finally, the optimal landmark positions are found by optimizing quadratic functions and shape constraints. The joint objective function is given by

$$\begin{aligned} \boldsymbol{x}^* &= \arg\max_{\boldsymbol{x}} \boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x}-2\boldsymbol{F}\boldsymbol{x}-\beta\left(\boldsymbol{x}-AW\boldsymbol{x}\right)^T\left(\boldsymbol{x}-AW\boldsymbol{x}\right), \\ &\text{subject to } -\boldsymbol{\omega}/2<W\cdot\boldsymbol{x}<\boldsymbol{\omega}/2 \end{aligned} \tag{2}$$

where $\boldsymbol{H}=diag(H_1,\ldots H_n)$, $\boldsymbol{F}=[F_1\ \cdots\ F_n]$, and $\boldsymbol{\omega}$ is the histogram width vector. In a PCA-based shape model, the shape parameters are usually limited by three times square root of eigenvalues. The range of ICA shape parameters $[-\omega_i/2,\omega_i/2]$ covers the same amount of sample variance as that by 3-sigma range with PCA given a normal distribution by recalculating $\boldsymbol{\omega}$ to cover 99.7% sample variance along ICs.

The clinical differences between DS and a healthy population mainly lie in the inner face features (around the eyes, nose and mouth) shown in Fig.3 (a). So it is desirable to build separate models for the DS and healthy groups.

The results of the inner facial landmarks from the first level HCLM serves as the initialization of the second level search. The above searching process is repeated for both the second level DS model and normal model. The best fitted model is selected as the one whose result is closer to its own mean shape and holds smaller changes to the second level initialization.

2.2 Feature Extraction, Selection and Classification

DS presents both special morphology (e.g. upward slanting eyes, small nose and wide-opened mouth) and textures (e.g. smooth philtrum and prominent epicanthic folds) [9]. To describe the facial information, geometric and texture features are extracted on the patient image registered to a reference image using Procrustes analysis to remove the translation and in-plane rotation. Geometric features are defined via interrelationships among 22 anatomical landmarks to incorporate clinical criteria used for DS diagnosis. Geometric features include horizontal and vertical distances normalized by the face size, and corner angles between landmarks. There are a total of 27 geometric features.

The local texture features are extracted based on LBP [12]. First, an LBP histogram is extracted from a square patch around each inner facial landmark. Then six

statistical measures of the histogram are computed, which are the mean, variance, skewness, kurtosis, energy and entropy. Finally, the feature vectors in all patches are concatenated to form the 132-dimensional local texture features.

The geometric and local texture features are concatenated to 159 combined features. Feature selection is performed using the method in [13]. The optimal dimension for feature space is found based on maximizing the area under the receiver operating characteristic (AUROC) curves by empirical exhaustive search.

The classification performances of SVM [14] with RBF kernel, linear SVM, k-NN [15], RF [16] and LDA [17] are compared in this study. The parameters for classifiers are found optimally by gird search.

3 Experiments

The image dataset consists of 100 frontal facial photos (one photo per subject) with 50 DS patients and 50 healthy individuals acquired with a variety of cameras and under variable illumination. The subjects include 75 Caucasian, 16 African American and 9 Asian and both genders. The ages of patients vary from 0 to 3 years.

3.1 Landmark Detection Using HCLM

We compared three models: PCA-based CLM, ICA-based CLM, and ICA-based HCLM. The performance of landmark detection was evaluated by a normalized error

$$\varepsilon = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{N_{\mathrm{I}}}\sum_{x\in \mathrm{I}} d\left(\tilde{x}, x\right) \Big/ d_n, \tag{3}$$

where I is the set of inner face landmarks, $d\left(\tilde{x}, x\right)$ is the Euclidean distance between inner face points located by the automatic search and the corresponding ground-truth landmarks placed manually by experts, d_n is the distance between the two pupils as normalizing factor, and N is the number of images. The comparison of detection errors is shown in Fig.3 (b). The overall error is 0.057±0.036, 0.049±0.028, and 0.041±0.028 for PCA-based CLM, ICA-based CLM, and ICA-based HCLM, respectively. A significant improvement was recorded by using ICA vs. PCA with CLM, and by using HCLM vs. CLM with ICA ($p<0.01$ for both).

3.2 Down Syndrome Detection

The experimental results are shown in Table I. Leave-one-subject-out validation is performed throughout the dataset. The performance is evaluated using accuracy, precision, and recall. We noted that the combined features achieved the best performance using SVM with RBF kernel with 95.6% accuracy with high precision and recall. The selected dimensions of geometric, texture and combined features are 24, 6 and 32, respectively. The texture features had slightly better performance than geometric features, probably due to the fact that anatomical location is already imbedded in the computation of the texture features. All the metrics improved when combining the

geometric and texture features, especially the precision. However, all classifiers achieved competitive performances. The accuracy for geometric, texture and combined features on ground truth landmarks were 0.923, 0.956 and 0.967, respectively. Our automatic results are at least as good as of studies [10, 11] using manual methods.

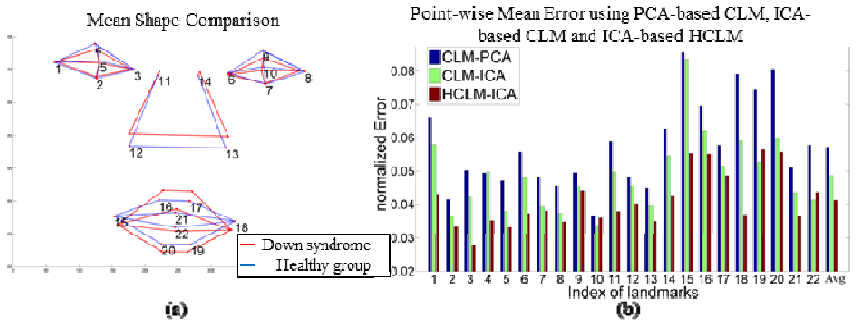

Fig. 3. (a) The mean shape comparison on inner face features between the DS and healthy groups; (b) The normalized landmark detection errors for PCA-based CLM (blue), ICA-based CLM (green) and ICA-based HCLM (refinement, red).

Table 1. performance comparison of Down syndrome detection using different features and classifiers

	Accuracy			**Precision**			**Recall**		
	Geometric	Texture	Combined	Geometric	Texture	Combined	Geometric	Texture	Combined
SVM-RBF	0.901	0.912	**0.956**	0.886	0.907	0.953	0.907	0.907	**0.953**
Linear SVM	**0.934**	0.934	0.945	**0.930**	0.930	**0.975**	0.930	0.930	0.907
k-NN	0.945	0.945	0.956	0.952	0.932	0.933	0.930	0.953	0.977
RF	0.923	0.901	0.923	0.929	0.905	0.929	0.907	0.884	0.907
LDA	**0.934**	**0.945**	0.945	0.911	**0.952**	0.932	**0.953**	0.930	**0.953**

4 Conclusion

In this study, a hierarchical constrained local model based on ICA was proposed to locate facial landmarks for DS detection. The ICA-based analysis gave a better representation of the local shape variations in the training data than PCA, an important factor in the analysis of medical image data. The two-level structure of HCLM significantly improved the accuracy of landmark detection over CLM. Based on the detected anatomical landmarks, geometric and local texture features were extracted and selected. Finally, several classifiers were employed to discriminate between the Down syndrome and healthy groups. The best performance was achieved by the combined geometric and texture features with SVM-RBF classifier with 95.6% diagnostic accuracy. The promising results also demonstrate the robustness of our method for analyzing highly variable photographic data. Data collection is on-going and future work will involve the investigation of other types of genetic syndromes.

Acknowledgements. This project was supported by a philanthropic gift from the Government of Abu Dhabi to Children's National Medical Center. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the donor.

References

1. Cootes, T.F., et al.: Active Shape Models-Their Training and Application. Computer Vision and Image Understanding 61, 38–59 (1995)
2. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 484–498. Springer, Heidelberg (1998)
3. Cristinacce, D., Cootes, T.: Automatic feature localisation with constrained local models. Pattern Recognition 41, 3054–3067 (2008)
4. Üzümcü, M., Frangi, A.F., Sonka, M., Reiber, J.H.C., Lelieveldt, B.: ICA vs. PCA Active Appearance Models: Application to Cardiac MR Segmentation. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2878, pp. 451–458. Springer, Heidelberg (2003)
5. Parker, S.E., 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 88, 1008–1016 (2010)
6. Benn, P.A.: Advances in prenatal screening for Down syndrome: I. general principles and second trimester testing. Clinica Chimica Acta; International Journal of Clinical Chemistry 323, 1–16 (2002)
7. Sivakumar, S., Larkins, S.: Accuracy of clinical diagnosis in Down's syndrome. Archives of Disease in Childhood 89, 691 (2004)
8. Burçin, K., Vasif, N.V.: Down syndrome recognition using local binary patterns and statistical evaluation of the system. Expert Systems with Applications 38, 8690–8695 (2011)
9. Saraydemir, Ş., et al.: Down Syndrome Diagnosis Based on Gabor Wavelet Transform. Journal of Medical Systems 36, 3205–3213 (2012)
10. Cardoso, J.F.: High-order contrasts for independent component analysis. Neural Comput. 11, 157–192 (1999)
11. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, vol. 1, pp. I-511--I-518 (2001)
12. Ojala, T., et al.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 971–987 (2002)
13. Cai, D., et al.: Unsupervised feature selection for multi-cluster data. Presented at the Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA (2010)
14. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning 20, 273–297 (1995)
15. Denoeux, T.: A k-nearest neighbor classification rule based on Dempster-Shafer theory. IEEE Transactions on Systems, Man and Cybernetics 25, 804–813 (1995)
16. Breiman, L.: Random Forests. Machine Learning 45, 5–32 (2001)
17. Mika, S., et al.: Fisher discriminant analysis with kernels. In: Proceedings of the 1999 IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing IX, 1999, pp. 41–48 (1999)