A Bayesian Hierarchical Model for Classifying Craniofacial Malformations from CT Imaging

S. Ruiz-Correa, D. Gatica-Perez, H. J. Lin, L. G. Shapiro, and R.W. Sze

Abstract— Single-suture *craniosynostosis* is a condition of the sutures of the infant's skull that causes major craniofacial deformities and is associated with an increased risk of cognitive deficits and learning/language disabilities. In this paper we adapt to classification of *synostotic* head shapes a Bayesian methodology that overcomes the limitations of our previously published shape representation and classification techniques. We evaluate our approach in a series of large-scale experiments and show performance superior to those of standard approaches such as Fourier descriptors, cranial spectrum, and Euclidian-distance-based analyses.
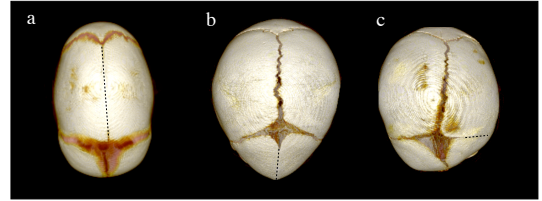
Fig. 1. Volumetric reformations (top view) of the skull of patient affected with isolated synostosis of the a) sagittal suture, b) metopic suture and c) unilateral suture. The suture that fused prematurely is shown as a dotted line.

I. INTRODUCTION

Craniosynostosis is the premature fusion of one or more calvarial sutures that normally separate the bony plates of the skull [4]. In normal developing infants, open sutures allow the skull to expand as the brain grows, producing normal head shape. If one or more sutures are prematurely fused, there is restricted growth perpendicular to the fused suture(s) and compensatory growth in the skull's open sutures, producing abnormal head shape (Figure 1).

Single-suture craniosynostosis (SSC) is the most common form of synostosis, with the prevalence of approximately 1 in 2,500 live births [23]. Among the isolated synostoses, fusions of the sagittal, coronal, and metopic sutures are most common. *Sagittal synostosis* is manifested at birth as a long, narrow head shape (scaphocephaly) with bifrontal and occipital bossing. *Unilateral coronal synostosis* is characterized by an asymmetrical skewed head (plagiocephaly) with retrusion of the forehead and brow ipsilateral to the fused suture and with compensatory contralateral frontal prominence on the side opposite of the fused suture. *Metopic synostosis* results in a triangular head shape called trigonocephaly, which features a midline forehead ridge, frontotemporal narrowing, and an increased biparietal diameter [13]. Imaging evaluation such as computed tomography (CT) scans are typically used to confirm the fused suture and to describe its effect on cranial morphology. In clinical practice, this evaluation is largely descriptive and qualitative, based on physical examination and radiography.

Single-suture craniosynostosis and its associated neurobehavioral features (such as a three- to five-fold increase in risk for cognitive deficits or learning/language disabilities [24]) represent an important area of research and clinical intervention for craniofacial providers and pediatric psychologists. Single suture synostoses are relatively common birth defects that present frequently in hospital-based craniofacial programs and neurodevelopmental centers. Unfortunately, the advancement of SSC research has been hindered due to the lack of quantitative methods to describe distinctive skull shape features and to reveal possible associations between skull deformity and other biological and psychological features.

The ability to classify craniosynostosis head shapes is a step in the development of techniques to characterize cranial anatomy [14]. A number of studies have presented approaches to classify SSC head shapes. Richtsmeier and collaborators [14] combined *Euclidian distance matrix analysis* (EDMA) and likelihood-based classification methods that lead to relatively large error rates, in the range of $18 - 32\%$. More recently, our own *severity indices* [19][21], *cranial spectrum* [20] and *symbolic shape descriptors* (SSDs) [15] have also been developed to quantify skull shapes with significantly improved error rates in the range of $6 - 10\%$. A main requirement of such approaches is the development of low-dimensional shape descriptors that enable accurate classification and good generalization ability.

In this paper, we adapt to representation and classification of synostotic head shapes recent work by Blei [2], which was originally designed to represent and learn document models in the Bayesian framework. We show that our proposed methodology provides an effective technique to classify synostotic head shapes that overcomes the shortcomings of our PLSA-based shape descriptors [15]. We also show that our approach produces SSDs of low-dimension that are able to represent and classify synostotic and normal control head shapes with low error rates.
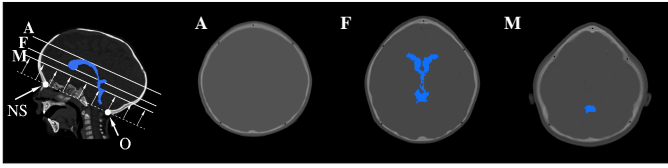
Fig. 2. In this work, all shape descriptors are computed from skull outlines that are derived from three CT image planes. Their location is defined by internal brain landmarks. The planes are parallel to the skull base plane, which is defined by using the frontal nasal suture (NS) anteriorly and opisthion (O) posteriorly. The A-plane is at the top of the lateral ventricle, the F-plane is at the Foramina of Munro, and the M-plane is at the level of the maximal dimension of the fourth ventricle. Ventricles are shown as colorized regions.

The paper is organized as follows. Section II.A describes our methods for acquiring imaging data. Section II.B and C summarize numeric shape descriptors that have been applied to classify synostotic head shapes, and techniques employed to reduce their dimensionality. Section II.D develops our approach to symbolic shape descriptors, which we derived from a hierarchical Bayesian model reported in [9]. This section also summarizes our methods for classifying synostotic skull shapes. Section III reports the results of our study. Section IV presents a discussion of our findings and section V concludes our paper.

II. METHODS

The task we approach in this work can be sumarized as follows. We are given a random sample of skull shapes labeled as sagittal, metopic, unicoronal, or normal control. Using skull shape information from CT imaging, we wish to construct a set of descriptors (numeric or symbolic) and a classification function, in order to predict the label of a new head shape.
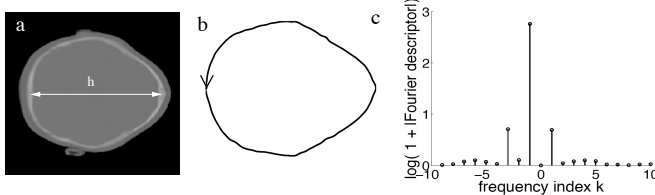

Fig. 3. An outline shape is computed from a CT image using standard segmentation techniques. a) CT image of a patient affected with metopic synostosis. The head length measured at the M-plane is denoted by h. b) Outline shape oriented in CCW direction. c) Twenty coefficients of the Fourier descriptor computed from the outline shape in b).

A. Source of CT images

All shape descriptors used in this study are computed from skull outline shapes extracted from CT images. The locations of the image planes are determined by internal brain landmarks (Fig. 2). The planes are parallel to the skull base plane, which is defined by using the frontal nasal suture anteriorly and opisthion posteriorly. The A-plane is at the top of the lateral ventricle, the F-plane is at the Foramina of Munro, and the M-plane is at the level of the maximal

dimension of the fourth ventricle. Outline shapes from the A, F and M-image planes are computed using standard image segmentation techniques and spline interpolation as shown in Fig 3. Each outline is represented as a contour with N points having vertex coordinates $\{x(n), y(n)\}$, $n = 0, 1, \cdots, N - 1\}$. In this study, outline shapes are oriented in the CCW direction.

B. Standard numeric shape descriptors

1) Fourier descriptors [18]: Given an outine with N points, we could form a complex sequence $z(n) = x(n) + jy(n)$. Traversing the outline, it is evident that $z(n)$ is a periodic function with period of N samples. The Fourier series of $z(n)$ can be derived as one period of its discrete Fourier transform $Z(k)$. The Fourier descriptor of $z(n)$ is a N-dimensional vector derived from $|Z(k)|$ by setting $Z(0)$ to 0, in order to make the descriptors independent of position, and dividing the remaining coefficients by $|Z(1)|$ in order to normalize for size. After these steps, the FD is independent of position, size, and starting point of the contour. We compute Fourier descriptors for the outlines of the A, F and M-image planes, and concatenate the $L = 3$ resulting vectors to form a single shape descriptor of size $O(LN)$.
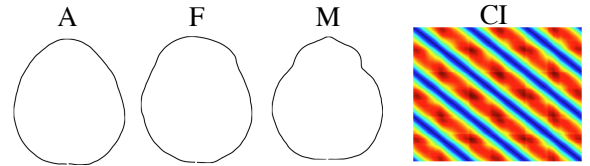

Fig. 4. Outline skull shapes computed at the A, F and M image planes of a patient affected with metopic synostosis. The corresponding cranial image descriptor (CI) is shown as a false color image.

2) Chord-length distributions: can be used to characterize 2-D closed outlines [18]. A chord-length measure l_{ij} is defined as the length of the line segment that links a pair of outline points i and j, normalized by a characteristic length such as the head length (h in Fig. 3). The complete set of chords for a given outline, consists of all possible chords drawn from every vertex to every other vertex (Fig. 4). The complete set of chords of an oriented outline can be represented as a distance matrix that we call *cranial image* (CI) in [19]. The vertices of an oriented outline can be numbered consecutively. Consequently, the i-th row of the CI stores the chord-lengths from the i-th vertex to every other vertex also in consecutive order. The i-th row of the CI is called the i-th *chord length distribution* (CLD) of the outline shape.

In order to avoid ambiguity in the numbering of the vertices, the first vertex is defined by the location of the metopic suture on the outline. Note that a CI is a coordinate-free, albeit redundant representation that has $N(N - 1)/2$ unique length parameters. The definition of CI can be extended to incorporate an arbitrary number L of outline shapes per skull. This is accomplished by computing inter- and intra-chord lengths for each of the L outlines. Therefore the size of a cranial image descriptor is $O(L^2 N^2)$.

C. Dimensionality reduction techniques

Principal component analysis (PCA) and random projection (RP) are standard techniques that are often applied to numeric shape descriptors to reduce their dimensionality. Although PCA analysis is widely used in a variety of applications, random projections have recently emerged as a powerful method for dimensionality reduction that offers many benefits over PCA for data sets that do not follow a multivariate Gaussian distribution [5]. In random projection, the original high-dimensional data is projected onto a lower-dimensional subspace using a random matrix whose columns have unit lengths. More specifically, a random projection from n dimensions to d dimensions is represented by a $d \times n$ matrix that can be generated using the following algorithm: a) set each entry of the matrix to an i.i.d. $N(0,1)$ value. b) Make the d rows of the matrix orthogonal by using the Gram-Schmidt algorithm, and then normalize them to unit length [16]. In practice, compact descriptors tend to improve the odds of statistical significance (generalization ability) in a classification task. In this work we used PCA and RP to reduce the dimensionality of the numeric shape descriptors presented in previous sections.
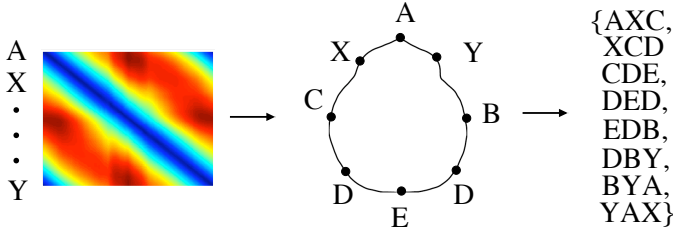

Fig. 5. Labels for the chord-length distributions of the CI shown on the left are computed by k-means clustering of all the chord-length distributions in the training set. The symbolic outline shape of a metopic skull is shown in the middle. The code-word generation process for the symbolic outline shown on the right side, consists of forming substrings of size W_s from adjacent vertices. In this toy example, $L = 1$ and $W_s = 3$.

D. Our approach

We begin with the algorithm to compute our symbolic shape descriptors (SSDs). To this end, we adapt to shape representation a variant of the generative model called latent Dirichlet allocation (LDA) [9] that we describe in section II-D.2.

The key points of our proposed algorithm are: a) the computation of a *symbolic outline representation* shape by k-means clustering of the chord-length distributions from a set of outline shapes; and b) the encoding of global geometric properties that differentiate our skull shape classes (sagittal, metopic, unicoronal and normal control) by LDA modeling of local shape features (*code words*), which are derived from the symbolic outline representation.

1) Computation of shape descriptors: The algorithm to construct SSDs is as follows. The input is a training set of skull shapes $\{S_1, \cdots, S_M\}$ and a testing skull shape S_{M+1}. We let $\mathcal{S}$ denote the set $\{S_1, \cdots, S_D\}$, where $D = M + 1$. The output is a set of SSD $\{\Theta_1, \cdots, \Theta_D\}$, where

$\Theta_i = (\Theta_{i1}, \cdots, \Theta_{iK})$, and K is a parameter to be defined below. It is assumed that each shape in $\mathcal{S}$ is represented by L oriented outlines, and each outline is discretized into N evenly spaced vertices. For the sake of simplicity and without loss of generality, we assume that $L = 1$.

1) Compute the cranial image representation for the outline shapes in the training set.
2) Assign symbolic labels to all the vertices of the outlines in $\mathcal{S}$. Labels are computed by applying k-means clustering to the CLDs associated with the CIs calculated in the previous step. The number of clusters k (i.e. the number of symbolic labels) is currently specified by the user at this time. The vertices are tagged according to the label assigned to their corresponding CLDs. An outline shape with labeled vertices is called a *symbolic outline shape* (Fig. 5).
3) Use the k-means cluster centers obtained in the previous step and a nearest neighbor rule to assign symbolic labels to the vertices of the testing outline S_{M+1}.
4) Compute code-words from the symbolic outline representation of the skulll-shapes in $\mathcal{S}$. More specifically, the labels associated with the vertices can be used to construct *code-words*. The word size is fixed at some integer $1 \leq W_s \leq N$. For instance, when $W_s = 3$, each word contains three labels. A toy example illustrates the code-word generation process. Let $XCDEDBYA$ be the string representing a symbolic outline in Fig. 5, where $N = 8$. We use capital letters as our symbolic labels and set $W_s = 3$. The code-words are $\{AXC, XCD, CDE, DED, EDB, DBY, BYA, YAX\}$.
5) Form a *code-book* by collecting all different code-words associated with all the skull shapes in $\mathcal{S}$. The number of code-words in the code-book is denoted by W.
6) Compute $D \times W$ co-occurrence matrix of counts n_{ji}, which stores the number of times the i-th code-word occurred in the symbolic outline representation associated with $S_j \in \mathcal{S}$.
7) Fore each $S_j \in \mathcal{S}$ compute the corresponding SSD

$$\Theta_j = (\Theta_{j1}, \cdots, \Theta_{jK})$$

by applying the LDA model to the co-ocurrence matrix calculated in the previous step.

2) **LDA model specification and inference:** In the context of our work, LDA models the codewords of each symbolic outline shape as a mixture over what we call *geometric topics*. A geometric topic is defined as a multinomial distribution over all the code-words in a code-book.

It is easier to understand the model by going through the generative process for creating the code-words of an outline of a particular shape class. We assume that there are K latent geometric topics, each being a multinomial distribution over a vocabulary of size W. For shape $S_j \in \mathcal{S}$, we first draw a mixing proportion $\theta_j = \{\theta_{jk}\}$ over K geometric topics from a symmetric Dirichlet distribution with concentration
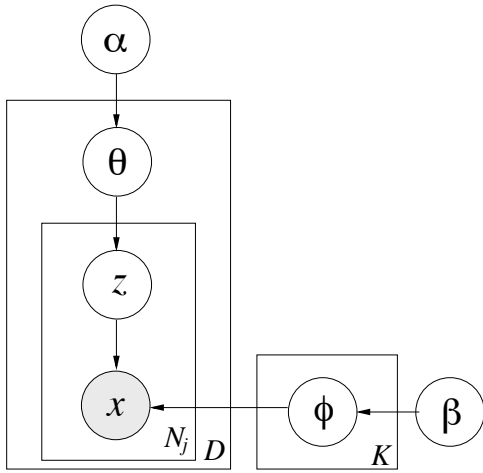
Fig. 6. Graphical representation of the LDA generative model of code-words associated with symbolic outline shapes.

parameter α. For the i-th code-word of S_j, a topic z_{ij} is drawn with topic k chosen with probability θ_{jk}. Then, the code-word x_{ij} is drawn from the z_{ij}th topic, with x_{ij} taking value w with probability ϕ_{kw}. Finally, a symmetric Dirichlet prior with parameter β is placed on the topic parameters $\phi_k = \{\phi_{kw}\}$. The LDA graphical model is shown in Fig. 6, where $\boldsymbol{z} = \{z_{ij}\}$, $\boldsymbol{x} = \{x_{ij}\}$, $\boldsymbol{\theta} = \{\theta_j\}$, and $\boldsymbol{\phi} = \{\phi_j\}$, and N_j is the number of code-words for the j-th symbolic outline shape. The graphical model in Fig. 6 uses plate notation.

The application to this model to shape representation is motivated by an intuitive argument. An outline's code-words are formed by concatenating symbolic labels (which really represent chord-length distributions) of adjacent outline vertices. Consequently, code-words tend to encode correlation patterns of local geometric features (CLDs) of an outline shape. We assume that such correlations are preserved across instances of a shape class. We also presume that a particular shape class may be a mixture of different correlation patterns that appear with distinctive frequencies (mixing proportions $\boldsymbol{\phi}$) across class instances. Obviously, different classes may share the same local correlation patterns, but the point is that their frequency of occurrence (geometric topic mixing proportions $\boldsymbol{\theta}$) differs significantly between classes. Such differences would reflect, to some extent, global shape information describing how local geometric features are organized within a shape class.

Assuming that α and β are given, the joint distribution of all parameters and variables of the model is

$$
\begin{aligned}
p(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{\phi} | \alpha, \beta; K) &= \prod_{j=1}^{D} \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \prod_{k=1}^{K} \theta_{jk}^{n_{jk.}+\alpha-1} \\
&\times \prod_{k=1}^{K} \frac{\Gamma(W\beta)}{\Gamma(\beta)^W} \prod_{w=1}^{W} \phi_{kw}^{n_{.kw}+\beta-1}(1)
\end{aligned}
$$

where $n_{jkw} = \#\{i : x_{ij} = w, z_{ij} = k\}$, and dot means the corresponding index is summed out: $n_{.kw} = \sum_j n_{jkw}$, and $n_{jk.} = \sum_w n_{jkw}$. Given the observed code-words $\boldsymbol{x}$ the task of Bayesian inference is to compute the posterior distribution over the latent geometric topic indices $\boldsymbol{z}$, the

mixing proportions $\boldsymbol{\theta}$, and the geometric topic parameters $\boldsymbol{\phi}$.

By applying collapsed Gibbs sampling [8] to perform inference, we construct a Markov Chain that converges to the posterior distribution on $\boldsymbol{z}$, and then use the results to infer $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. To apply this algorithm, we need the full conditional distribution $p(z_{ij} = k | \boldsymbol{z}^{\neg ij}, \boldsymbol{x}; K)$, where the superscript $\neg ij$ means the corresponding variables or counts x_{ij} and z_{ij} are excluded. This conditional distribution is computed in two steps. First, the marginal distribution $p(\boldsymbol{z}, \boldsymbol{x}; K)$ is obtained by marginalizing out $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ in (1) by applying conjugate analysis [1]. The collapsed marginal distribution over $\boldsymbol{x}$ and $\boldsymbol{z}$ is

$$
\begin{aligned}
p(\boldsymbol{x}, \boldsymbol{z} | \alpha, \beta; K) &= \prod_{k=1}^{K} \frac{\Gamma(W\beta)}{\Gamma(n_{.k.} + W\beta)} \prod_{w=1}^{W} \frac{\Gamma(n_{.kw} + \beta)}{\Gamma(\beta)} \\
&\times \prod_{j=1}^{D} \frac{\Gamma(K\alpha)}{\Gamma(n_{j..} + K\alpha)} \prod_{k=1}^{K} \frac{\Gamma(n_{jk.} + \alpha)}{\Gamma(\alpha)}(2) \\
&= p(\boldsymbol{x} | \boldsymbol{z}; K) \times p(\boldsymbol{z}; K)
\end{aligned}
$$

where Γ is the gamma function. Second, cancellation of terms in (2) yields the desired result

$$
p(z_{ij} = k | \boldsymbol{z}^{\neg ij}, \boldsymbol{x}, \alpha, \beta; K) = \frac{n_{.kw}^{\neg ij} + \beta}{n_{.k.}^{\neg ij} + W\beta} \cdot \frac{n_{jk.}^{\neg ij} + \alpha}{n_{j..}^{\neg ij} + K\alpha}. \quad (3)
$$

Having obtained the full conditional distribution, the Gibbs sampling algorithm is simple. The z_{ij} variables are initialized to values in $\{1, \cdots, K\}$, determining the initial state of the Markov chain. The chain is then run for a number of iterations, each time finding a new state by sampling each z_{ij} from the distribution in (3). After enough iterations for the chain to approximate the target distribution, the current value of the z_{ij} variables is recorded. Subsequent samples are taken after an appropriate lag to ensure that their autocorrelation is low [8]. The predictive values for $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ given $\boldsymbol{z}$ and $\boldsymbol{x}$ can be estimated from the chain samples by

$$
\hat{\theta}_{jk} = \frac{n_{jk.} + \alpha}{n_{j..} + K\alpha} \quad (4)
$$

$$
\hat{\phi}_{kw} = \frac{n_{.kw} + \beta}{n_{.k.} + W\beta} \quad (5)
$$

Note from (3) that $p(z_{ij} = k | \boldsymbol{z}^{\neg ij}, \boldsymbol{x}) \propto (n_{.kw}^{\neg ij} + \beta)(n_{.k.}^{\neg ij} + W\beta)^{-1}(n_{jk.}^{\neg ij} + \alpha)$; consequently, z_{ij} depends on $\boldsymbol{z}^{\neg ij}$ only through the counts $n_{.kw}^{\neg ij}$, $n_{.k.}^{\neg ij}$, and $n_{jk.}^{\neg ij}$. Specifically, the dependence of z_{ij} on any particular other variable $z_{i'j'}$ is very weak for large data sets. For this reason, the convergence of collapsed Gibss sampling is expected to be quick [9].

The symbolic shape descriptors for shape $S_j \in \mathcal{S}$ are defined by the estimated parameters $\hat{\theta}_j$ of the LDA model as the vector

$$
\Theta_j = (\Theta_{j1}, \cdots, \Theta_{jK})
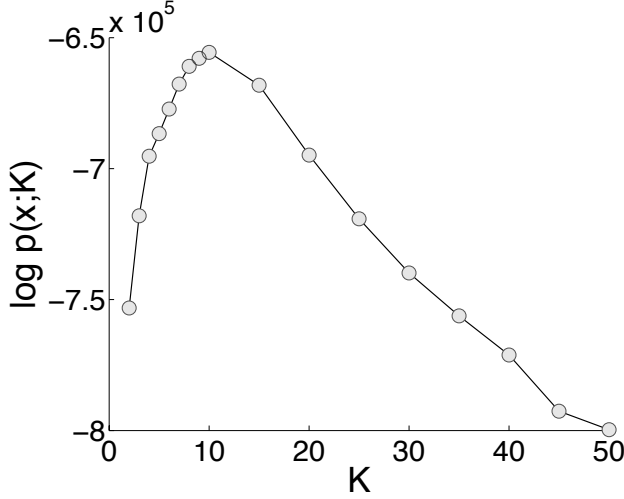$$

where $\Theta_{jk} = \hat{\theta}_{jk}$.

Fig. 7. Log-likelihood of the data for different settings of the number of geometric topics, K for $\alpha = 50$ and $\beta = 1$. Standard errors for each point in the plot were smaller than the plot symbols.

	LDA-SVM-I			LDA-SVM-II		
	S	M	C	S	M	C
S	96.7	0	5.66	95.6	0	1.89
M	2.2	94.74	1.89	3.3	94.74	3.77
C	1.1	5.26	92.46	1.1	5.26	94.34
	PLSA-SVM-I			PLSA-SVM-II		
	S	M	C	S	M	C
S	95.6	0	5.66	96.7	0	3.77
M	2.2	94.74	1.89	2.2	94.74	3.78
C	2.2	5.26	92.45	1.1	5.26	92.45
	FD-PCA			FD-RP		
	S	M	C	S	M	C
S	93.4	2.63	5.66	95.6	0	5.66
M	2.2	60.53	86.79	1.1	55.26	13.21
C	4.4	36.8	86.79	3.3	34.21	71.7

TABLE I

CONFUSION MATRICES (%) FOR LDA-SVM-I-, LDA-SVM-II-, PLSA-SVM-I-, PLSA-SVM-II-, FD-PCA- AND FD-RP-BASED DESCRIPTORS. KEY: SAGITTAL (S), METOPIC (M) AND CONTROL (C). DIMENSIONALITY REDUCTION RATE WAS SET TO 60:1.

3) **LDA model selection:** The LDA model depends on α, β and an unknown parameter K. Following [9], our strategy in this work is to fix α and β and explore the consequences of varying K. Given the values of α and β, the problem of choosing the appropriate K is a problem of model selection that we approach by computing the likelihood of $p(\boldsymbol{x}; K)$. We approximate $p(\boldsymbol{x}; K)$ by taking the harmonic mean of a set of values $p(\boldsymbol{x}|\boldsymbol{z}; K)$ [12], when $\boldsymbol{z}$ is sampled from the posterior $p(\boldsymbol{z}|\boldsymbol{x}; K)$:

$$p(x|K) = \left[\frac{1}{m} \sum_{i=1}^{m} p(\boldsymbol{x}|\boldsymbol{z}^{(i)}; K)^{-1} \right]^{-1},$$

where $\{\boldsymbol{z}^{(i)}, i = 1, \cdots, m\}$, is a sample taken from the Gibbs sampler described above. The value of $p(\boldsymbol{x}|\boldsymbol{z}, K)$ can be computed from the first term in right-hand-side of (2):

$$p(\boldsymbol{x}|\boldsymbol{z}; K) = \prod_{k=1}^{K} \frac{\Gamma(W\beta)}{\Gamma(n_{.k.} + W\beta)} \prod_{w=1}^{W} \frac{\Gamma(n_{.kw} + \beta)}{\Gamma(\beta)}.$$

4) **Classification functions:** We classify our skull shapes by applying support vector machines (C-SVMs) to numeric and symbolic shape descriptors [22]. All the free parameters of our approach (k, W_s, α, β, and C) were computed using a cross-validation technique in order to minimize the leave-one-out classification error [22]. We construct our SVMs with two kinds of kernels. A standard dot-product kernel (linear SVM) is used for both, numeric and symbolic descriptors. In addition, we apply a kernel that yields an isometric embedding of the SSDs into the unit sphere. More specifically, let Θ_i and Θ_j be the SSDs for S_i and S_j, respectively. The kernel is defined as $k(\Theta_i, \Theta_j) = \sum_{k=1}^{K} \Theta'_{ik}\Theta'_{jk}$, where $\Theta'_{ik} = \sqrt{\Theta_{ik}}$ and $\Theta'_{jk} = \sqrt{\Theta_{jk}}$. Note that $||\Theta'_i|| = ||\Theta'_j|| = 1$.

III. RESULTS

We conducted a comparative study of numeric and symbolic shape descriptors in a classification task of synostotic and normal control head shapes. Our data set consisted of 100 skull shapes with sagittal synostosis, 40 with metopic synostosis, 16 with right unilateral coronal synostosis, and of 60 normal controls [20].

The FDs of section II-B.1 were used as our numeric descriptors with $N = 200$ and $L = 3$ that yield 600-D vectors. We applied PCA and RP techniques to the FDs in order to reduce their dimensionality to $K = 10$, thus achieving a 60-fold dimensionality reduction rate (DRR). Classification (cross-validation) error rates for FD-PCA and FD-RP were 15.38% and 19.78%, respectively.

We also tested PLSA-based and LDA-based symbolic shape descriptors in various classification runs with linear (LDA-SVM-I and PLSA-SVM-I) and nonlinear (LDA-SVM-II, PLSA-SVM-II) kernels (see section II-D.4). The classification errors were comparable: 5.5% for linear SVMs and 4.95% for non-linear SVMs. Symbolic-descriptors were also 10-dimensional vectors. This value of K was chosen using the model selection techniques described in section II-D.3 (Fig. 7). We performed experiments with $\beta \in [1, 5]$ and $\alpha \in [1, 50]$ and obtained error rates that are comparable to those reported above and in Table I.

Confusion matrices for all the algorithms tested (FD-PCA, FD-RP, LDA-SVM-I, LDA-SVM-II, PLSA-SVM-I and PLSA-SVM-II) show that symbolic descriptors outperform numeric ones at the given DRR (Table I). Overall, LDA- and PLSA-based descriptors have comparable performance. Classification rates are slightly better for non-linear SVMs (LDA-SVM-II and PLSA-SVM-II). Similar results to those of Table I were obtained for PLSA- and LDA-based descriptors with DRRs of 75:1 and 85:1 The performance of PCA and RP-based descriptors deteriorated significantly at these rates. Computation of LDA parameters for our population

sample took roughly 1 (s) and 20 (s) on average for LDA and PLSA, respectively. Training and testing of the SVMs using cross-validation took about 1 (min) for both LDA and PLSA techniques.

IV. DISCUSSION

Single-suture craniosynostosis constitute an important area of research that requires the creation of new methods to characterize cranial anatomy. Quantitative methods for skull-shape analysis are crucial in the study of cranial abnormalities and their relation to the general physiological state of affected individuals. A step in that direction is the development of shape descriptors that enable accurate of classification craniosynostosis malformations.

We have compared the classification performance of LDA-based SSDs with and numeric descriptors of reduced dimensionality and PLSA-based SSDs that have been previously applied to classify synostotic head shapes. Symbolic descriptors achieve lower error rates than those of numeric descriptors at the given DRR (85:1, 75:1, and 60:1). Lower error rates can be achieved by the numeric shape descriptors than those reported here [15]. However, such performance increase comes at the expense of much larger descriptor sizes and increased probability of classification over-fitting. Random projection and PCA produced similar error rates in all our experiments.

LDA-based and PLSA-based SSDs showed comparable classification behavior. However there are many reasons why LDA-based descriptors are better suited for applications in craniosynostosis research. Probabilistic latent semantic analysis (PLSA) was developed in the context of document modeling [10] and has been successfully used in a variety of applications in areas such as object class learning and recognition [6][7], scene analysis [25], and shape classification [15]. However, some studies have shown that PLSA has severe over-fitting problems [2] [17].

A further difficulty with PLSA in applications to craniosynostosis research is that the parameters of the model are computed by applying a maximum likelihood estimation approach (often by means of the EM algorithm), which produces a local maxima solution of the problem (that is to say, a non-unique set of PLSA-based SSDs). Different runs of the estimation algorithm will generally provide different solutions. This is due to the fact that the initial guess of the model parameters are often set at random values.

Although we have shown that the multiplicity of local solutions is not an issue in relation to classification performance, it may be a problem in terms of the shape representation. Symbolic shape descriptors are not only applied to head shape classification. They are also used to characterize possible associations (via regression models) between cranial shape and a diversity of genetic, physiological or neuro-physiological variables, representing the status of an affected patient. For this reason we suggest that the use of PLSA-based SSDs to shape representation should be carefully examined. On the other hand, LDA is a well defined generative

model that generalizes easily to new symbolic outline shapes [2]. Our implementation of the LDA model uses a Monte Carlo procedure that provides model parameters computed by averaging the local solutions of the likelihood function [9], thus avoiding the problem of multiple of solutions and over-fitting.

We are aware of popular shape descriptors and template-matching techniques for computer vision that are related to our work. An example is the well known *shape context* algorithm developed in [3]. Although the shape context has been successfully in 2-D object recognition, it is difficult to apply to our classification problem since we have several outlines representing a skull shape and the shape context algorithm can only deal with a single outline at a time. Moreover, the shape-context representation leads to high-dimensional shape descriptors. The shape context of a skull shape typically is of size $O(LNM)$ with $M \sim 10^2$.

It is worth mentioning that we found error rates for unicoronal synostosis were on the order of 30% (data not shown). This is likely due to the fact that shape descriptors were computed from CT image slices (planes A, F and M) that were originally selected to represent sagittal and metopic synostosis but may not be representative of unicoronal malformation. Also, the population size for these group of patients is small (16 infants). Further work and a larger sample size is required to improve the statistical significance of this finding.

Finally we note that in all our experiments, classification results obtained with the non-linear SVM appeared to be higher than those obtained with the linear SVM. Therefore we suggest the square root parameterization described in section II-D.4 may be directly used to construct the alternative descriptors Θ', which could be used for classification and other problems related to craniosynostosis research.

V. CONCLUSION

We have presented an approach that uses a hierarchical Bayesian model and SSDs to classify synostotic head shapes. We conducted a comparative study with previously reported numeric and PLSA-based SSDs and showed that our proposed descriptors better characterize craniosynostosis malformations and provides low error rates compare to those of traditional numeric descriptors. Our results would be strengthened in future by larger data sets and by applications beyond classification. Nevertheless, this exploratory analysis suggests that the LDA-based SSDs will be a useful tool in the study of single-suture craniosynostosis.

VI. ACKNOWLEDGMENTS

REFERENCES

[1] J. M. Bernardo and M.F.A. Smith, M. F. A. *Bayesian theory*. New York: Wiley 1994.

[2] D. M. Blei, A. Y. Ng, M.I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 2003, pp 993-1022.

[3] S. Bolongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on PAMI*, vol. 24, no. 24, 2002, pp 509-522.

[4] M. M. Cohen and M.C. MacLean. *Craniosynostosis: Diagnosis, Evaluation, and Management*, 2nd Ed. Oxford: Oxford University Press; 2000.

[5] S. Dasgupta, D.J. Hsu, and N. Verma. "A concentration theorem for projections", *Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI)*, 2006.

[6] R. Fergus, P. Perona, and A Zisserman. "A visual category filter for Google images". In *Proc. ECCV*. Springer, 2004.

[7] R. Fergus, L. Fei-Fei, P. Perona, and A Zisserman. "Learning object categories from google's image search ". In *Tenth IEEE International Conference on Computer Vision*, vol. 2, 2005, pp 1816-1823.

[8] W.R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in practice*. London: Chapman and Hall, 2005.

[9] T. L. Griffiths and M. Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, vol. 101, 2004, pp 5228-5235.

[10] T. Hofmann, Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, vol. 42, 2001, pp 177-196.

[11] K. A. Kapp-Simon, A. Figueroa, C. A.Jocher, and M. Schafer. Longitudinal assessment of mental development in infants with nonsyndromic craniosynostosis with and without cranial release and reconstruction. *Plast. Reconstr. Surg*. vol. 92 num. 5, pp 831-9.

[12] R. E. Kass and A. E. Raftery, Bayes factors. *Journal of the American Statistical Association*, vol. 90, no. 430, 1995, pp 773-795.

[13] E. Lajeunie, M. Le Merrer, C. Marchac and D. Renier. Genetic study of scaphocephaly. *Am. J. Med. Gene*. vol. 62, 1996, pp 282-285.

[14] S.R. Lele and J.T. Richtsmeier. *An invariant approach to the statistical analysis of shapes*. New York: Chapnan and Hall/CRC; 2001.

[15] H. J. Lin, S. Ruiz-Correa, L. G. Shapiro, A. V. Hing, M. L. Cunningham, M. L. Speltz and R.W. Sze. "Symbolic shape descriptors for classifying craniosynostosis deformations from skull imaging", *IEEE Engineering in Medicine and Biology Society (EMBS), Annual International Conference*, 2005, pp 6325-6331.

[16] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vettering. *Numerical Recipes in C: The Art of Scientific Computing*, 3rd ed., Cambridge University Press, 2002.

[17] A. Popescul, L. Ungar, D. Pennock, and S. Lawrence. "Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments". In *Proceedings of the seventh Conference Uncertainty in Artificial Intelligence* 2001, pp 437-444.

[18] R.M. Rangaraj. *Biomedical image analysis*, CRC Press, 2006.

[19] S. Ruiz-Correa, R. W. Sze, H. J. Lin, L. G. Shapiro, M.L. Speltz and M. L. Cunningham. "Classifying craniosynostosis deformations from skull shape imaging". *Computer-Based Medical Systems (CBMS). The 18th IEEE Symposium*, 2005, pp 335-340.

[20] S. Ruiz-Correa, R. W. Sze, J. R. Starr, H. J. Lin, M. L. Speltz, M. L. Cunningham and A. V. Hing. New scaphocephaly severity indices of sagittal craniosynostosis. A quantitative study with cranial index quantifications. *The American Cleft Palate-Craniofacial Association Journal*, vol. 43, num. 2, 2006, pp 211-221.

[21] S. Ruiz-Correa, J. R. Starr, H. J. Lin, K. Kapp-Simon, R. W. Sze, R. G. Ellenbogen, M. L. Speltz and M. L. Cunningham. New severity indices for quantifying single suture metopic craniosynostosis. Accepted for publication. *Neurosurgery*, 2008.

[22] B. Scholköpf and A. Somola. *Learning wit kernels*. The MIT Press, 2002.

[23] A. Shuper, P. Merlob, M. Grunembaum, and S.H Reisner. Reisner. The incidence of isolated craniosynostosis in the newborn infant. *Am J Dis Child* vol. 139, no. 1, 1985, pp 85-86.

[24] M.L. Speltz, K.A. Kapp-Simon, M.L. Cunningham, J. Marsh, and G. Dawson. Single-suture craniosynostosis: a review of neurobehavioral research and theory. *Journal of Pediatric Psychology*, vol. 29, num. 8, 2004, pp 651-668.

[25] P. Quelhas, F. Monay, J.-M. Odobez, D. Gatica-Perez, T. Tuytelaars, and L. Van Gool "Modeling Scenes with Local Descriptors and Latent Aspects", in *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, Beijing, 2005.