

# Aprendizagem e Síntese de Variedades via Coordenadas Normais de Riemann Locais e Baricêntricas

Gastão F. Miranda Jr. <sup>1,2</sup>, Gilson Giraldi <sup>2</sup>, Carlos Thomaz <sup>3</sup>, Raul Daniel Millán <sup>2</sup>

<sup>1</sup>Departamento de Matemática – Universidade Federal de Sergipe (UFS)  
São Cristóvão – SE – Brasil

<sup>2</sup>Laboratório Nacional de Computação Científica (LNCC)  
Petrópolis – RJ – Brasil.

<sup>3</sup>Faculdade de Engenharia Industrial (FEI)  
São Bernardo do Campo – SP – Brasil.

gastao@ufs.br, gilson@lncc.br, cet@fei.edu.br, rdaniel.millan@gmail.com

**Abstract.** *Manifold learning has been widely exploited for dimensionality reduction in high dimensional database analysis with applications in pattern recognition, data mining and computer vision. In this work we propose a framework, called Local Riemannian Manifold Learning (LRML), which recovers the manifold topology and geometry through normal coordinate neighborhoods computed by the exponential map. Besides, we propose a simple strategy based on barycentric coordinates to map the feature space into the Riemannian manifold in order to perform the manifold synthesis. We apply the whole methodology for artificial face image database analysis to demonstrate its efficiency for dimensionality reduction and data visualization.*

**Resumo.** *Técnicas em aprendizado de variedades vêm sendo utilizadas para redução de dimensionalidade em aplicações envolvendo reconhecimento de padrões, mineração de dados e visão computacional. Neste trabalho, propomos uma metodologia denominada aprendizado local de variedades Riemannianas, a qual recupera a topologia e geometria da variedade utilizando sistemas locais de coordenadas normais computadas via aplicação exponencial. Além disto, é proposto também uma estratégia simples, baseada em coordenadas baricêntricas, para sintetizar pontos da variedade a partir de novos pontos do espaço de características. Nós aplicamos a metodologia proposta para análise de imagens sintéticas de faces para demonstrar a eficiência do método para redução de dimensionalidade e visualização de dados.*

## 1. Introdução

Classificação, reconhecimento e agrupamento automático de padrões são problemas importantes em áreas como engenharia, psicologia, biologia, medicina, visão computacional e inteligência artificial [Ma and Fu 2012]. No caso de imagens de faces humanas em tons de cinza, com resolução  $m \times n$  pixels, preferencialmente normalizada com mesma escala e orientação, observamos uma quantidade relevante de propriedades redundantes, como a posição dos olhos, boca e nariz. Assim, se representarmos as imagens como pontos em um espaço Euclidiano  $\mathbb{R}^D$  de dimensão  $D = m \cdot n$ , é intuitivo

supor que possamos representar o conjunto destas imagens com uma quantidade  $d < D$  de coordenadas, isto é, efetuando uma redução de dimensionalidade.

Supondo que as imagens originais possam ser representadas por uma variedade diferenciável (generalização de superfície suave)  $\mathcal{M}^d$  imersa no espaço  $\mathbb{R}^D$ , podemos tratar a redução de dimensionalidade utilizando técnicas de aprendizado de variedades para estimar uma aplicação entre o espaço  $\mathbb{R}^d$  e a variedade  $\mathcal{M}^d$ . A maioria das técnicas de aprendizado de variedades concentra-se em mapear a variedade que contém os dados por uma única aplicação, ou mais especificamente, por uma parametrização global  $\varphi : \mathbb{R}^d \rightarrow \mathcal{M}^d \subset \mathbb{R}^D$ . Na prática, este problema é tratado da seguinte forma: dada uma amostra  $\mathcal{D} = \{p_1, p_2, \dots, p_N\} \subset \mathcal{M}^d$  o objetivo é estimar parâmetros  $\tilde{\mathcal{D}} = \{z_1, z_2, \dots, z_N\} \subset \mathbb{R}^d$ , tais que  $\varphi(z_i) = p_i, i = 1, 2, \dots, N$ , sem de fato construir explicitamente a aplicação  $\varphi$ .

Esta filosofia está na base de métodos tradicionais de aprendizado de variedades, tais como o *Local Tangent Space Alignment* (LTSA) [Zhang and Zha 2005]. Neste caso, para cada ponto  $p_i$  da base de dados, a solução proposta determina inicialmente os  $KNN$  vizinhos mais próximos,  $KNN(p_i) = \{p_{i_1}, p_{i_2}, \dots, p_{i_K}\}$ . Em seguida, computa-se uma aproximação afim,  $d$ -dimensional para cada  $KNN(p_i)$ . Finalmente, aplica-se um processo de otimização para compatibilizar as representações lineares locais obtidas gerando as coordenadas globais  $z_i, i = 1, 2, \dots, N$ . O *Locally Linear Embedding* (LLE) [Roweis and Saul 2000] e *Local Generative Units and Global Affine Transformation* (LGGTA) [Huang et al. 2009] também utilizam reconstruções lineares locais. O primeiro utiliza representação via combinação convexa enquanto que o segundo aplica decomposição SVD (*Singular Value Decomposition*) para as vizinhanças  $KNN(p_i)$ , guardando apenas os  $d$  valores singulares principais. Aspectos diretamente ligados a geometria intrínseca dos dados podem ser melhor abordados incorporando métodos da geometria Riemanniana ao aprendizado, o que é parte dos fundamentos do *Riemannian Manifold Learning* (RML) e LOGMAP [Lin and Zha 2008, Brun et al. 2005].

Este trabalho tem a sua fundamentação teórica na técnica de RML [Lin and Zha 2008] e LOGMAP [Brun et al. 2005], e diretamente ligado a um dos problemas principais, e ainda sem solução geral na área de aprendizado de variedades: Como preservar distâncias no espaço de parâmetros, ou espaço de características? Estas técnicas procuram representações em dimensão reduzida para os dados via isometrias (aplicações bijetoras e diferenciáveis que preservam produto interno, e conseqüentemente distâncias, sobre a variedade). Porém, na forma proposta, estas técnicas contemplam apenas isometrias globais.

No entanto, como desconhecemos a estrutura da variedade suporte dos dados, não podemos partir desta hipótese. Por exemplo, se os dados são uma amostra da esfera  $S^2$ , é um resultado conhecido na geometria diferencial que uma única parametrização não é suficiente para cobrir o conjunto [do Carmo 1988]. Assim, necessita-se buscar uma família de parametrizações locais, cujas imagens gerem uma cobertura para a variedade que representa os dados, e com isso ter garantia de recuperar a geometria intrínseca que os dados carregam.

Com estes pensamentos, propomos neste trabalho uma nova metodologia, denominada aprendizado local de variedades Riemannianas (*Local Riemannian Manifold*

*Learning* - LRML), a qual recupera a topologia e geometria da variedade utilizando sistemas locais de coordenadas normais computadas via aplicação exponencial. O método proposto tem como entrada elementos da base de dados e executa os seguintes passos: (a) Determinar a topologia dos dados utilizando a metodologia do RML; (b) Estimar a dimensão  $d$  da variedade; (c) Gerar um sistema de vizinhanças; (d) Computar coordenadas normais de Riemann para cada vizinhança obtida via LOGMAP.

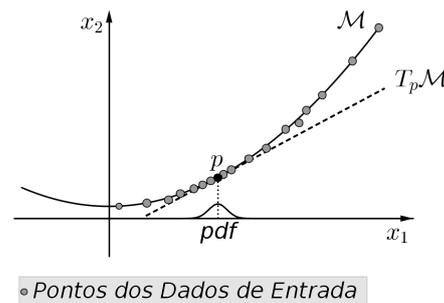
Contudo, o RML e LOGMAP não tratam o problema de síntese de novos dados a partir do espaço de características, problema este tratado no contexto de outras técnicas de aprendizagem de variedades [Álvarez-Meza et al. 2011][Huang et al. 2009]. Neste trabalho, a síntese de novas imagens é implementada via coordenadas baricêntricas [Allgower and Georg 1990]. Especificamente, dado um novo ponto  $z \in \mathbb{R}^d$  no espaço de parâmetros, primeiramente identificamos um  $d$ -simplexo que o contém, com vértices em pontos do conjunto  $\tilde{\mathcal{D}}$ . Em seguida, computamos as coordenadas baricêntricas de  $z$ . Finalmente, estas coordenadas são utilizadas para interpolar os pontos correspondentes em  $\mathbb{R}^D$  para estimar (sintetizar) o novo ponto da variedade. A extensão da técnica RML para uma família de parametrizações locais e a técnica proposta para síntese são as principais contribuições deste artigo.

O texto está organizado como segue. Na Seção 2, descrevemos os elementos de geometria diferencial e projetiva presentes neste trabalho. O método LRML é apresentado na Seção 3. Os experimentos computacionais são discutidos na Seção 4, onde ilustramos uma aplicação em imagens de faces humanas de modo que a geometria de aquisição permite estimar propriedades da variedade que representa os dados. Finalmente, na Seção 5, apresentamos as conclusões e trabalhos futuros.

## 2. Fundamentação Teórica

A representação de um banco de imagens via variedades diferenciáveis permite reunir ferramentas geométricas, algébricas e estatísticas em um mesmo formalismo. Isto pode ser observado na Figura 1, a qual representa um conjunto de dados, a variedade (superfície) suporte correspondente e seu plano (espaço) tangente em um ponto  $p$ , formado por todos os vetores tangentes a  $\mathcal{M}$  em  $p$ .

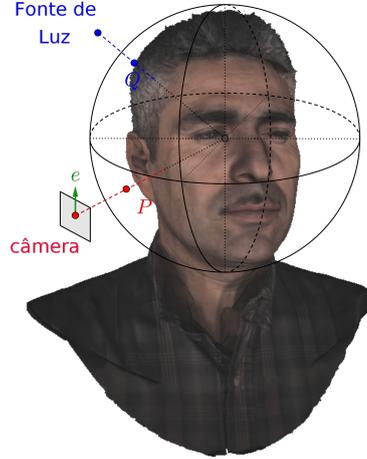
Estes dados são uma amostra da população, a qual segue alguma densidade de probabilidade representada abaixo do espaço tangente. Se a curvatura da variedade for desprezível, então o espaço tangente fornece uma representação eficiente para o conjunto de dados. Este espaço pode ser obtido por técnicas tais como *Principal Component Analysis* (PCA) e *Multidimensional Scaling* (MDS) [Engel et al. 2012] [Lee and Verleysen 2007]. Se a curvatura da variedade não for desprezível, então o espaço tangente  $T_p\mathcal{M}$  é apenas uma representação local do espaço



**Figura 1. Modelo de dados: sistema de coordenadas original  $(x_1, x_2)$ , variedade  $(\mathcal{M})$ , espaço tangente  $T_p\mathcal{M}$  em um ponto  $p \in \mathcal{M}$ , função de densidade de probabilidade (*pdf*).**

de imagens em questão. Aspectos estatísticos relacionados à densidade de probabilidade associada devem ser considerados para estabelecer condições para densidade de amostras.

Uma das grandes dificuldades em aprendizado de variedades é justificar se os dados de entrada realmente se ajustam a uma variedade suave. Para imagens de faces humanas podemos utilizar a representação mostrada na Figura 2 para a geometria de aquisição destas imagens. Nesta figura representamos um indivíduo e a esfera  $S^2 \subset \mathbb{R}^3$ , utilizada para descrever o posicionamento da câmera e da fonte de luz. Diferentes poses são obtidas movendo a câmera enquanto que o rosto permanece fixo. Assim, vamos supor que a distância da câmera para o rosto e o comprimento focal da câmera sejam fixos de modo que as imagens adquiridas tenham escalas semelhantes. O eixo da câmera é definido passando pela origem da esfera. A iluminação é modelada por uma fonte de luz fixa. Se a orientação da câmera também não variar ao longo das aquisições, é natural considerar que o espaço das imagens  $\mathcal{M}$  gerado pode ser parametrizado por uma aplicação:



**Figura 2. Modelo de aquisição das imagens (Fonte: [www.cyberware.com](http://www.cyberware.com)).**

onde  $(\phi, \theta)$  representa um ponto de intersecção de  $S^2$  com o eixo da câmera. Esta expressão, juntamente com a geometria de aquisição descrita, de modo intuitivo, mostram que podemos considerar que as imagens de uma face fixa estão contidas em uma superfície de dimensão  $d = 2$ , uma vez que o espaço de parâmetros  $U$  é bidimensional [Lin and Zha 2008]. Esta superfície estaria imersa em um espaço de dimensão maior  $\mathbb{R}^D$ . Supondo que a equação (1) seja diferenciável então podemos calcular vetores tangentes em um ponto  $p$  genérico da superfície, bem como definir o espaço formado por estes vetores, o espaço tangente  $T_p\mathcal{M}$ . Esta ideia pode ser generalizada, supondo-se que temos um conjunto  $\mathcal{M} \subset \mathbb{R}^D$ , e uma família de aplicações biunívocas  $x_\alpha : U_\alpha \subset \mathbb{R}^d \rightarrow \mathcal{M}$  de abertos  $U_\alpha$  de  $\mathbb{R}^d$  em  $\mathcal{M}$  tais que:

$$\varphi : U \subset \mathbb{R}^2 \rightarrow \mathcal{M}, U = \left\{ (\phi, \theta); 0 \leq \phi < \pi, -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2} \right\} \quad (1)$$

(i)  $\cup_\alpha x_\alpha(U_\alpha) = \mathcal{M}$ ,  
(ii) Para todo par  $\alpha, \beta$ , com  $W = x_\alpha(U_\alpha) \cap x_\beta(U_\beta) \neq \emptyset$ , os conjuntos  $x_\alpha^{-1}(W)$  e  $x_\beta^{-1}(W)$  são abertos em  $\mathbb{R}^d$  e as aplicações  $x_\beta^{-1} \circ x_\alpha^{-1}$  são diferenciáveis.

Neste caso, dizemos que  $\mathcal{M}$ , ou  $\mathcal{M}^d$ , é uma variedade diferenciável de dimensão  $d$ . Cada par  $(U_\alpha, x_\alpha)$  da família é chamado uma parametrização local de  $\mathcal{M}^d$ . A propriedade (i) tem o significado óbvio de garantir a *cobertura* completa do conjunto  $\mathcal{M}$  pela família de aplicações. A propriedade (ii) implica que na região  $W \subset \mathcal{M}$  é indiferente utilizar a parametrização  $x_\alpha$  ou  $x_\beta$  para descrever as propriedades geométricas de  $\mathcal{M}^d$ , uma vez que esta propriedade permite *mudar* suavemente de uma parametrização para outra. Este modelo pode ser estendido para definir uma noção de distância entre imagens que

seja coerente com a geometria da variedade associada. Para isto, devemos acrescentar o conceito de produto escalar (aplicação em  $\mathcal{M}^d$  que a cada ponto  $p \in \mathcal{M}^d$  associa uma forma bilinear  $\langle \cdot, \cdot \rangle_p$ , simétrica e positivo definida em  $T_p \mathcal{M}^d$ ) à definição de variedade, gerando a noção de variedade Riemanniana [do Carmo 1988].

Com estes elementos, dados dois pontos  $p, q \in \mathcal{M}^d$ , e uma curva  $\alpha : \mathbb{R} \rightarrow \mathcal{M}^d$  passando por estes pontos, podemos computar o comprimento do arco que liga  $p$  e  $q$ , denotado por  $d_\alpha(p, q)$ . Se  $\alpha$  for uma geodésica (curva de comprimento mínimo sobre  $\mathcal{M}^d$ , ligando  $p$  e  $q$ ), então dizemos que  $d_\alpha(p, q)$  é a distância geodésica entre os dois pontos, a qual será denotada por  $d(p, q)$  para simplificar a notação.

Podemos definir também a aplicação exponencial, a qual faz parte dos fundamentos do RML. Seja  $p \in \mathcal{M}^d$ , e  $\alpha$  uma geodésica passando pelo ponto  $p$ , com direção tangente  $\mathbf{v} \in T_p \mathcal{M}^d$  em  $p$ . Então definimos a aplicação exponencial  $\exp$  em  $p$  como sendo a função:

$$\begin{aligned} \exp_p : U \subset T_p \mathcal{M}^d &\rightarrow \mathcal{M}^d, \\ \mathbf{v} &\mapsto q = \exp_p(\mathbf{v}) \Leftrightarrow d_\alpha(p, q) = \langle \mathbf{v}, \mathbf{v} \rangle_p, \end{aligned}$$

onde  $\langle \mathbf{v}, \mathbf{v} \rangle_p$  é a norma de  $\mathbf{v}$  dada pelo produto interno  $\langle \cdot, \cdot \rangle_p$  sobre a variedade e  $U$  é um aberto em  $T_p \mathcal{M}^d$ .

Se identificarmos  $T_p \mathcal{M}^d$  com  $\mathbb{R}^d$  então pode-se considerar a função  $\exp_p$  como uma parametrização local de  $\mathcal{M}^d$ . Pode-se mostrar ainda que a aplicação  $\exp_p$  é uma isometria local em  $p$ , ou seja, preserva distâncias sobre a variedade. Dada uma base ortonormal  $\{\mathbf{e}_i\}$  de  $T_p \mathcal{M}$ , as componentes de um vetor  $\mathbf{v}$  nesta base, ou seja, o vetor  $(u_1, u_2, \dots, u_d)$  tal que  $\mathbf{v} = \sum_{i=1}^d u_i \mathbf{e}_i$ , definem as **coordenadas normais de Riemann** do ponto  $q = \exp_p(\mathbf{v})$  em relação à base  $\{\mathbf{e}_i\}$ .

### 3. Método LRML Proposto

Nossa proposta, denotada por LRML, consiste em utilizar o algoritmo RML [Lin and Zha 2008] para determinar a topologia dos dados, efetuar a redução de dimensionalidade através de uma adaptação local do algoritmo de LOGMAP [Brun et al. 2005], e sintetizar novos pontos no espaço original dos dados utilizando coordenadas baricêntricas.

#### 3.1. Seleção de vizinhanças

O RML [Lin and Zha 2008] tem como entrada a base de dados  $\mathcal{D} = \{p_1, \dots, p_N\} \subset \mathbb{R}^D$ , o número de vizinhos para cada ponto  $p_i$  da base e um parâmetro  $\rho$  utilizado para a seleção destes vizinhos, ambos escolhidos por tentativa e erro. Inicialmente, utiliza-se o algoritmo de Dijkstra para computar os caminhos mínimos sobre o grafo completo  $G(\mathcal{D})$  cujos vértices são os elementos de  $\mathcal{D}$ .

- 1 Para cada elemento  $p_i \in \mathcal{D}$  usamos o algoritmo  $KNN$  para escolher o conjunto  $KNN(p_i)$  formado pelos  $K \leq N$  vizinhos mais próximos do ponto  $p_i$  sobre  $G(\mathcal{D})$ .
- 2 Determinar o conjunto  $VN(p_i)$ , formado pelos vizinhos visíveis de  $p_i$ , definidos por

$$VN(p_i) = \{p \in KNN(p_i); \langle p_i - p, p_i - q \rangle \geq 0, \forall q \in KNN(p_i)\}.$$

- 3 Obter o conjunto dos vizinhos seguros, denotado por  $SN(p_i)$ : Primeiro, ordenamos os vetores  $\mathbf{e}_j = p_j - p_i$ ,  $p_j \in VN(p_i)$  de forma crescente em relação a seus comprimentos, obtendo o conjunto  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_K\}$ . Em seguida, usamos o PCA para estimar a dimensão intrínseca local,  $d_j$  do subespaço gerado pelos primeiros  $j$  elementos de  $\{\mathbf{e}_1, \dots, \mathbf{e}_K\}$ , sendo  $1 \leq j \leq K$ . Então, temos uma lista de dimensões  $\{d_1, \dots, d_K\}$  para o conjunto  $\{\mathbf{e}_1, \dots, \mathbf{e}_K\}$ . Se  $d_j > d_{j-1}$ , calcule o incremento,  $\|\mathbf{e}_j\| - \|\mathbf{e}_{j-1}\|$ . Se este incremento é maior que um limiar dado por  $\rho \mu(\|\mathbf{e}_j\|)$ , onde  $\rho > 0$  real e  $\mu(\|\mathbf{e}_j\|)$  é a média dos comprimentos dos vetores  $\{\mathbf{e}_1, \dots, \mathbf{e}_K\}$ , então o ponto  $p_j$  é removido.

Ao final destes passos, teremos associada uma dimensão  $d_i$  para cada  $SN(p_i)$ . Com isto, podemos estimar a dimensão da variedade utilizando uma média sobre os  $d_i$ 's. No entanto, os conjuntos  $SN(p_i)$  definem uma topologia sobre o conjunto de dados, a qual permite definir sistemas de vizinhanças sobre a variedade de interesse, como veremos a seguir.

### 3.2. Construindo a Cobertura dos Dados

Com o objetivo de efetuarmos um tratamento local, descrevemos a construção de uma cobertura formada por vizinhanças *seguras*  $\mathcal{D}_i \subset \mathcal{D}$  tais que  $\mathcal{D} = \cup_{i=1}^L \mathcal{D}_i$ . Para isto, vamos agora considerar o sub-grafo  $\tilde{G}$  de  $G$ , cujas arestas interligam apenas vizinhos seguros e utilizar o algoritmo de Dijkstra para computar os caminhos mínimos sobre o grafo  $\tilde{G}$ . Na literatura de aprendizado de variedades [Lin and Zha 2008], estes caminhos são considerados aproximações das distâncias geodésicas sobre a variedade que contém os dados.

- 1 Escolhemos um elemento  $p \in \mathcal{D}$  e fazemos  $W \leftarrow \mathcal{D}$  e  $i \leftarrow 1$ .
- 2 Início do laço principal: enquanto  $W \neq \emptyset$ :
  - 3  $q_i \leftarrow p$ ,
  - 4 Seja  $V(q_i)$  o conjunto dos  $KNN$  vizinhos mais próximos de  $q_i$  em  $\tilde{G}$ .
  - 5 Seja  $q'$  o elemento de  $V(q_i)$  mais distante de  $q_i$  e  $r_i = d(q_i, q')$  a distância geodésica de  $q_i$  a  $q'$ . Em seguida toma-se todos os pontos seguros cuja distância geodésica a  $q_i$  é menor ou igual a  $r = \alpha r_i$ , onde  $\alpha > 1.0$ . Seja  $\mathcal{D}_i$  o conjunto obtido.
  - 6 Escolhemos  $p$  como o ponto mais distante de  $q_i$  em  $\mathcal{D}_i$ ,
  - 7 Se  $W \neq W - \mathcal{D}_i$ , então  $W \leftarrow W - \mathcal{D}_i$ , caso contrário escolhemos  $p \in W$ ,
  - 8  $i \leftarrow i + 1$ ,

O algoritmo acima tem como saída os pontos  $q_i \in \mathcal{D}$ , denominados pontos base, e os conjuntos  $\mathcal{D}_i$  e  $V(q_i)$ ,  $i = 1, \dots, L$ , onde  $L$  é o número de vezes que é executado o laço principal.

### 3.3. Coordenadas Normais de Riemann Locais

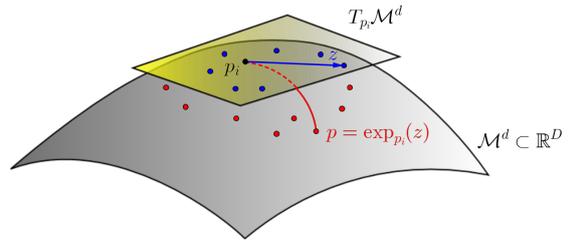
Os conjuntos  $\mathcal{D}_i$  obtidos fornecem uma cobertura da amostra  $\mathcal{D}$  a qual, juntamente com a dimensão  $d$  da variedade, será utilizada como entrada para a versão local do LOGMAP, dada abaixo:

- 1 Consideramos o conjunto  $V(p_i)$  e utilizamos PCA para determinar uma base ortonormal  $\{\mathbf{e}_j^i\}_{j=1}^d$  para o espaço tangente. Todos os pontos  $p \in V(p_i)$

são projetados em  $T_{p_i}\mathcal{M}$ , utilizando a base gerada pelo PCA, obtendo-se as coordenadas locais  $y = (y_1, \dots, y_d) \in \mathbb{R}^d$  correspondentes. Procedemos de modo análogo para  $p \in \mathcal{D}_i - V(p_i)$  obtendo-se  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ . Sejam  $Y_i$  e  $X_i$  os conjuntos de vetores de coordenadas assim formados.

- 2 Usando um processo de mínimos quadrados padrão, aproximamos localmente o quadrado da função distância geodésica  $d(x, y)$  por uma função  $f_i : X_i \times Y_i \rightarrow \mathbb{R}$  onde  $f_i(x, y) = (x - y)^T A_i (x - y)$ , sendo  $A_i \in M_{d \times d}(\mathbb{R})$  simétrica e positiva definida.
- 3 Dado  $p \in D_i$ , para calcular a direção da geodésica ligando este ponto ao ponto  $p_i$ , estimamos numericamente  $\mathbf{g} = \sum_{i=1}^d g^j \mathbf{e}_j = \nabla_y d^2(x, y)|_{y=p_i}$  usando a função  $f_i$  obtida no passo anterior.
- 4 Computamos as coordenadas normais de Riemann para cada ponto  $p \in \mathcal{D}_i$  pela expressão:  $z = d(p, p_i) \frac{\mathbf{g}}{\sqrt{\langle \mathbf{g}, \mathbf{g} \rangle}}$ , onde distância  $d(p, p_i)$  é estimada pelo caminho mínimo no grafo  $\tilde{G}$  que liga os pontos  $p$  e  $p_i$ .

Pelo item 4 acima, a origem do sistema de coordenadas normais obtido,  $z = 0$ , será o vetor de coordenadas normais de  $p_i$ . Ao final deste processo obtemos o conjunto  $\tilde{\mathcal{D}} = \{z_1, z_2, \dots, z_N\}$  onde  $z_i$  representa as coordenadas normais (locais) do ponto  $p_i$  do conjunto de dados original  $\mathcal{D}$ . Por outro lado, temos também:  $\langle z, z \rangle = \left\langle d(p, p_i) \frac{\mathbf{g}}{\sqrt{\langle \mathbf{g}, \mathbf{g} \rangle}}, d(p, p_i) \frac{\mathbf{g}}{\sqrt{\langle \mathbf{g}, \mathbf{g} \rangle}} \right\rangle = (d(p, p_i))^2$ , ou seja, obtemos uma representação em dimensão reduzida que preserva as distâncias entre os pontos do conjunto de dados original. Esta é uma propriedade importante pois distância entre pontos (imagens) está diretamente ligada à noção de similaridade entre imagens; daí, a importância de preservá-la na representação em dimensão reduzida. A Figura 3 agrupa os elementos básicos da exposição acima e sumariza o mapeamento obtido pelo LRML, via aplicação exponencial.



**Figura 3. [Aplicação exponencial]** O espaço tangente  $T_{p_i}\mathcal{M}^d$  é levado em  $\mathcal{M}^d \subset \mathbb{R}^D$ , onde  $z$  representa um vetor em  $T_{p_i}\mathcal{M}^d$  com norma  $\sqrt{\langle z, z \rangle} = d(p, p_i)$  e  $p = \exp_{p_i}(z)$ .

### 3.4. Síntese de Novos Dados

Após efetuar a redução de dimensionalidade, necessitamos de uma técnica para efetuar a síntese de novos dados a partir de novos pontos no espaço de dimensão reduzida  $\mathbb{R}^d$ . Especificamente, dado um novo ponto  $\tilde{z}$  no espaço  $\mathbb{R}^d$ , com  $\tilde{z} \notin \tilde{\mathcal{D}}$ , como computar a imagem correspondente na variedade  $\mathcal{M}^d$  que está imersa em  $\mathbb{R}^D$ ? Neste trabalho, vamos tratar este problema utilizando coordenadas baricêntricas. Vejamos alguns conceitos fundamentais inicialmente [Allgower and Georg 1990].

**Definição 1** Um conjunto de pontos  $\{v_0, v_1, \dots, v_n\} \subset \mathbb{R}^m$  é dito **afim-independente** se as diferenças  $v_1 - v_0, v_2 - v_0, \dots, v_n - v_0$  formam um conjunto de vetores linearmente independente no  $\mathbb{R}^m$ .

**Definição 2** O fecho convexo  $\{v = \sum_{i=0}^n \alpha_i v_i \mid \alpha_i \geq 0, i = 0, 1, \dots, n \text{ e } \sum_{i=0}^n \alpha_i = 1\}$ , com vértices dados por  $n+1$  pontos afim-independentes  $\{v_0, v_1, \dots, v_n\}$  do  $\mathbb{R}^m$ , é chamado um **n-simplexo** e será denotado por  $[v_0, v_1, \dots, v_n]$ .

Dado um novo  $\tilde{z} \in \mathbb{R}^d$ , o objetivo da síntese é encontrar uma representação  $q \in \mathcal{M}^d \subset \mathbb{R}^D$  para  $\tilde{z}$ . Para isso, escolhemos um  $d$ -simplexo  $[\tilde{z}_0, \tilde{z}_1, \dots, \tilde{z}_d]$  contendo  $\tilde{z}$ , com vértices pertencentes ao conjunto  $\tilde{\mathcal{D}}$ . Neste caso, a equação  $\tilde{z} = \sum_{j=0}^d \alpha_j \tilde{z}_j$ , com a restrição  $\sum_{j=0}^d \alpha_j = 1$ , tem solução única  $(\alpha_0, \alpha_1, \dots, \alpha_d)$ . Uma vez que  $\tilde{z}_i \in \tilde{\mathcal{D}}$ , o  $d$ -simplexo em  $\mathbb{R}^d$  tem um correspondente direto  $[\tilde{p}_0, \tilde{p}_1, \dots, \tilde{p}_d]$ , com vértices em  $\tilde{p}_i \in \mathcal{D}$ , o que permite escrever  $q = \sum_{j=0}^d \alpha_j \tilde{p}_j$  como sendo uma estimativa (síntese) linear para a nova amostra  $q$  do espaço dos dados.

#### 4. Experimentos Computacionais

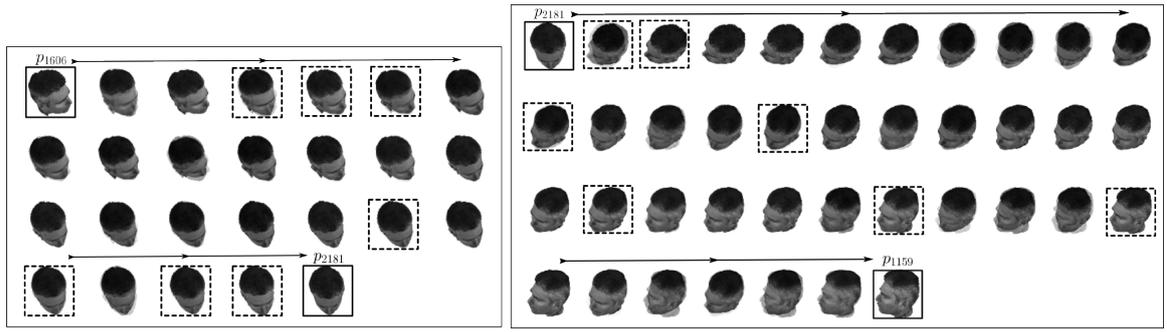
Dadas as restrições de orientação da câmera e da fonte luminosa fixada, descritas na Seção 2, montamos um banco de imagens sintéticas com diferentes poses da pessoa representada na Figura 2. As imagens são geradas utilizando o sistema **ParaView**<sup>1</sup>, tendo como entrada o modelo geométrico/textura correspondente à Figura 2. Neste caso, sabemos que  $d = 2$  pela discussão da Seção 2. Fazendo uma discretização regular do espaço  $(\phi, \theta)$ , foram geradas 2196 imagens com resolução de  $70 \times 80$  pixels, as quais compõem o conjunto de dados  $\mathcal{D}$ . Aplicamos o algoritmo LRML proposto (Seção 3) o qual gerou uma cobertura  $\{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_L\}$  com  $L = 74$ .

Para os experimentos que realizamos, usamos para a seleção de vizinhos descrita na Seção 3.1 um limiar de corte igual a 0.14 para autovalores fornecidos pelo PCA e  $\rho = 0.9$  (passo 3 da seleção de vizinhos). Após determinar as distâncias geodésicas sobre o grafo  $G$ , usamos o algoritmo da Seção 3.2 com  $K = 50$  vizinhos e  $\alpha = 1.5$  para os passos 4 e 5 na construção dos conjuntos  $\mathcal{D}_i$ .

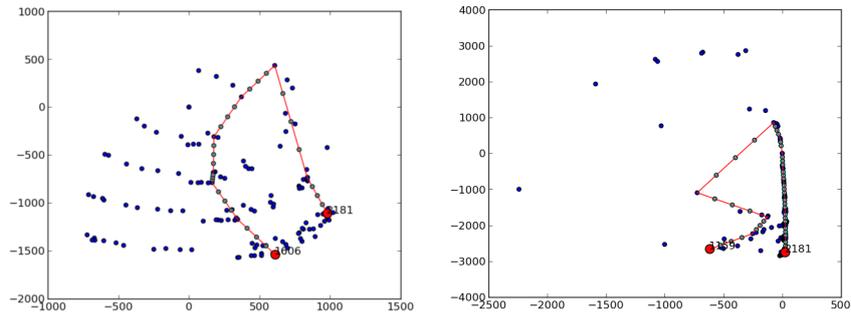
Foram tomadas as imagens  $p_{1606}, p_{1159} \in \mathcal{D}$  ilustradas na Figura 4(a) e um caminho geodésico  $c$  no grafo completo  $G(\mathcal{D})$ , ligando os pontos correspondentes. Inicialmente, vamos sintetizar pontos intermediários ao longo de  $c$ , para verificar a qualidade das novas imagens geradas pelo LRML. Neste processo, precisamos escolher um subconjunto  $\{\mathcal{D}_{i_1}, \mathcal{D}_{i_2}, \dots, \mathcal{D}_{i_\beta}\} \subset \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_L\}$  que cubra o caminho  $c$ . Por razões computacionais, escolhemos as vizinhanças  $\mathcal{D}_{i_j}$  de tal forma que  $\beta$  seja mínimo (obtivemos  $\beta = 2$  neste caso). Nas Figuras 4(a) e 4(b) que ilustram os resultados obtidos, as imagens destacadas nos extremos pertencem ao banco de imagens original, todas as outras são geradas utilizando a técnica descrita na Seção 3.4. Observa-se um resultado visualmente satisfatório para as imagens geradas. As Figuras em 4(b) mostram a disposição dos pontos no espaço de parâmetros  $\mathbb{R}^2$  onde cada ponto desta figura corresponde a uma imagem do banco  $\mathcal{D}$  e foi obtido pelo algoritmo da Seção 3.3.

Para ilustrar o erro da reconstrução neste experimento, tomamos os vértices do grafo  $G(\mathcal{D})$  que pertencem ao caminho  $c_0 \cup c_1$  da Figura 4 e, a partir das suas representações no espaço de características,  $\mathbb{R}^2$ , calculamos suas coordenadas baricêntricas. A partir destas coordenadas executamos a síntese de novas imagens  $\{\bar{p}_1, \dots, \bar{p}_{s-1}, \bar{p}_s\}$ , onde  $s = 21$  neste caso. Assim, podemos computar o erro relativo

<sup>1</sup>Kitware. *ParaView Visualization System*. <http://www.paraview.org>



(a)



(b)

**Figura 4. Trechos do caminho de  $p_{1606}$  à  $p_{1159}$  em duas vizinhanças distintas: em (a), à esquerda, subcaminho  $c_0$  ligando  $p_{1606}$  e  $p_{2181}$  e, à direita, subcaminho  $c_1$  de  $p_{2181}$  à  $p_{1159}$ ; em (b) a representação dos caminhos  $c_0$  e  $c_1$  no espaço de dimensão reduzida e suas respectivas vizinhanças em  $\mathbb{R}^2$ .**

para cada imagem gerada, dado por:

$$E_{\text{reconst}}(\bar{p}_{k_j}) = \frac{\|p_{k_j} - \bar{p}_{k_j}\|}{\|p_{k_j}\|}, \quad (2)$$

onde  $\|\cdot\|$  denota a norma Euclidiana em  $\mathbb{R}^D$ , espaço ambiente da imagens originais. Obtemos então um erro relativo médio de 0.1323 com desvio padrão de 0.0879, que é um resultado satisfatório. Na Figura 5 ilustramos as imagens sintetizadas e originais com o erro máximo e mínimo obtidas usando o LRML.

Comparamos ainda este resultado com o obtido pelo método RML rápido ilustrado na Figura 6, o qual computa uma parametrização global a partir dos dados. Esta

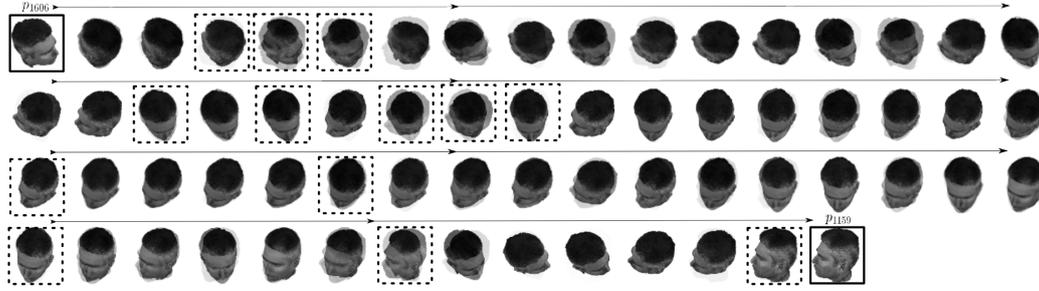


(a)

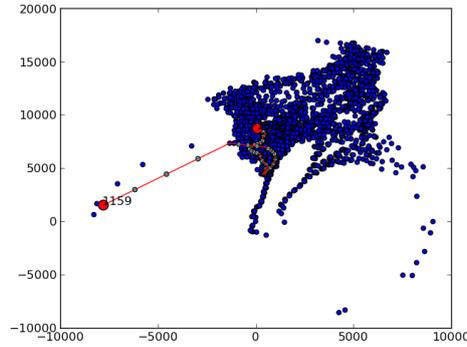
(b)

**Figura 5. Imagem reconstruída (esquerda) e imagem original (direita): (a) Imagens para o caso com menor erro relativo ( $E_{\text{reconst}} \approx 0.04966$ ); em (b) imagens com maior erro relativo ( $E_{\text{reconst}} \approx 0.38407$ ).**

comparação é importante pois a geração de um modelo a partir de dados discretos possui uma imprecisão inerente a todo processo de recuperação de um objeto contínuo a partir de uma amostra. No caso do aprendizado de variedades, técnicas envolvendo parametrizações locais podem minimizar este problema por diminuir a propagação de erros ao longo do processo de reconstrução.



(a)

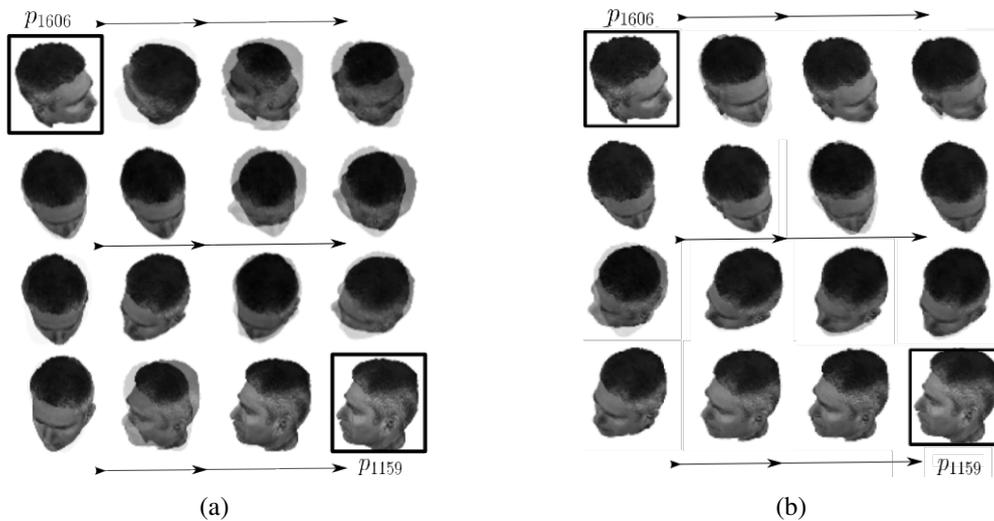


(b)

**Figura 6. Caminho  $c$  gerado pelo RML rápido: (a) ilustra o caminho  $c$  ligando  $p_{1606}$  e  $p_{1159}$  no espaço das imagens; em (b) a representação do caminho  $c$  no espaço de dimensão reduzida,  $\mathbb{R}^2$ .**

Para verificar este fato, observemos a Figura 6(a) a qual mostra o resultado obtido quando geramos as imagens intermediárias ao longo do caminho  $c$  utilizado no experimento anterior. Destacamos algumas das imagens obtidas (quadrado tracejado) para o LRML na Figura 4(a) e as suas correspondentes no RML rápido na Figura 6(a), as quais podem ser melhor visualizadas nas Figuras 7(a) e 7(b) facilitando uma comparação visual. Observamos que o RML rápido fornece um resultado com perda de qualidade visual para algumas imagens. Calculando o erro relativo dado na equação (2) para cada vértice no caminho  $c$ , obtivemos um erro relativo médio de 0.2174 e desvio padrão de 0.107 para a reconstrução usando uma parametrização global, que são maiores se comparadas como os valores obtidos pelo LRML. Este fato, indica que técnicas envolvendo parametrizações locais podem minimizar o erro inerente ao processo de reconstrução da variedade a partir de uma amostragem. Na Figura 6(b) temos a representação dos dados no espaço de dimensão reduzida  $\mathbb{R}^2$  fornecida pelo RML rápido.

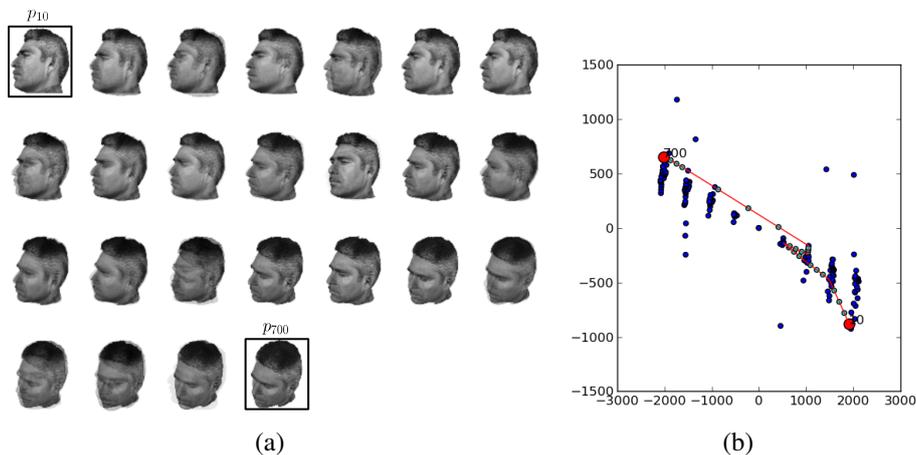
O próximo experimento tem o objetivo de avaliar a sensibilidade do LRML em relação à densidade da amostra. Assim, vamos re-amostrar a variedade gerando agora 1098 imagens, e executar o LRML com os mesmos parâmetros anteriores. Neste caso



**Figura 7. Imagens destacadas das Figuras 6(a) e 4(a): (a) imagens destacadas na Figura 6(a), obtidas através do método RML rápido por uma parametrização global; em (b) imagens destacadas de 4(a), obtidas pelo método LRML.**

obtivemos uma cobertura para as amostras com  $L = 45$  conjuntos  $\mathcal{D}_i$ 's.

Tomamos agora as imagens  $p_{10}$  e  $p_{700}$  ilustradas na figura 8(a) e um caminho geodésico  $c$  no grafo  $G(\mathcal{D})$ , ligando os pontos correspondentes. A Figura 8(a) também mostra as imagens geradas pelo método de síntese proposto ao longo deste caminho. Novamente, observamos um resultado visual satisfatório, apesar da redução da densidade da amostragem.



**Figura 8. Reconstrução com quantidade de amostras reduzida: em (a) ilustramos o trecho do caminho entre  $p_{10}$  e  $p_{700}$ ; em (b) representamos a vizinhança que contém o caminho no espaço de dimensão reduzida.**

## 5. Conclusões e Trabalhos Futuros

Neste trabalho, propomos uma metodologia para recuperar a topologia e geometria da variedade, a partir de um conjunto de amostras, utilizando sistemas locais de coordenadas normais computadas via aplicação exponencial. Os experimentos mostram

que a técnica é promissora, indicando superioridade em relação a parametrização global obtida pelo método RML rápido [Brun et al. 2005]. O método foi testado em base de imagens sintéticas, geradas a partir de um modelo tradicional para aquisição de fotografias de faces humanas. Assim, o espaço de parâmetros é bidimensional, o que simplifica o processo de síntese, pois necessitamos apenas de 3 pontos no espaço de dimensão reduzida (ou seja um 3-simplexo em  $\mathbb{R}^2$ ) para a geração da nova imagem. Para trabalhos futuros vamos aplicar a técnica para variedades diversas de dimensão  $d > 2$  e em bases de imagens reais. Custo computacional e análise de complexidade serão também considerados nos próximos trabalhos.

## Referências

- [Allgower and Georg 1990] Allgower, E. L. and Georg, K. (1990). *Numerical Continuation Methods: An Introduction*. Springer-Verlag Berlin Heidelberg.
- [Álvarez-Meza et al. 2011] Álvarez-Meza, A. M., Valencia-Aguirre, J., Daza-Santacoloma, G., Acosta-Medina, C. D., and Castellanos-Domínguez, G. (2011). Image synthesis based on manifold learning. volume 6855 of *Lecture Notes in Computer Science*, pages 405–412. Springer.
- [Brun et al. 2005] Brun, A., Westin, C., Herberthson, M., and Knutsson, H. (2005). Fast manifold learning based on riemannian normal coordinates. In *Image Analysis*, volume 3540 of *LNCS*, pages 920–929. Springer.
- [do Carmo 1988] do Carmo, M. (1988). *Geometria Riemanniana*. Projeto Euclides. Instituto de Matemática Pura e Aplicada.
- [Engel et al. 2012] Engel, D., Hüttenberger, L., and Hamann, B. (2012). A Survey of Dimension Reduction Methods for High-dimensional Data Analysis and Visualization. In *Proceedings of IRTG 1131 Workshop 2011*, volume 27, pages 135–149, Germany. Schloss Dagstuhl.
- [Huang et al. 2009] Huang, D., Yi, Z., and Pu, X. (2009). Manifold-based learning and synthesis. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 39(3):592–606.
- [Lee and Verleysen 2007] Lee, J. and Verleysen, M. (2007). *Nonlinear Dimensionality Reduction*. Information Science and Statistics. Springer.
- [Lin and Zha 2008] Lin, T. and Zha, H. (2008). Riemannian manifold learning. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(5).
- [Ma and Fu 2012] Ma, Y. and Fu, Y. (2012). *Manifold Learning Theory and Applications*. Taylor & Francis Group.
- [Roweis and Saul 2000] Roweis, S. and Saul, L. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326.
- [Zhang and Zha 2005] Zhang, Z. and Zha, H. (2005). Principal manifolds and nonlinear dimension reduction via local tangent space alignment. *SIAM J. Scientific Computing*, 26(1):313–338.