

Redução de dimensionalidade não-linear para seguimento de movimento¹

J. G. Silva^{*}, J. S. Marques[†], J. M. Lemos[‡]

^{*}ISEL, R. Conselheiro Emídio Navarro, 1950-062 Lisboa, Portugal
jgs@isiel.ipl.pt

[†]IST/ISR, Av. Rovisco Pais 1949-001, Lisboa, Portugal
jsm@isr.ist.utl.pt

[‡]INESC-ID/IST, R. Alves Redol, 9, 1000-029 Lisboa, Portugal
jlml@inesc.pt

Resumo: O problema da redução de dimensionalidade é relevante em diversas áreas da engenharia, nomeadamente em processamento de sinais e reconhecimento de padrões. Enquanto o caso linear pode ser resolvido com recurso à conhecida técnica de Análise em Componentes Principais, o caso não-linear é mais complexo. Recentemente, têm sido feitos avanços baseados na aprendizagem de variedades para aproximar os dados. Neste artigo apresenta-se um algoritmo designado Tangent Bundle Approximation (TBA), ou Aproximação do Fibrado Tangente, que segue a mesma abordagem, com vantagens computacionais relativamente a outras técnicas, para conjuntos de dados na ordem dos milhares de pontos. Outra vantagem do algoritmo TBA é fornecer uma estimativa da correcta dimensão intrínseca da variedade que aproxima os dados. Neste artigo apresenta-se uma descrição do TBA, ilustrada com resultados experimentais numa sequência vídeo.

Introdução

O recente interesse em algoritmos de aprendizagem de variedades deve-se, em grande medida, à multiplicação de conjuntos de dados de grandes dimensões, adquiridos e armazenados electronicamente, em diversos ramos da ciência e engenharia, desde a multimédia e processamento de sinais até à bioinformática [1].

Em aplicações destinadas ao seguimento de movimento, é frequente que as observações tenham dimensionalidade muito elevada. Isto sucede naturalmente em visão por computador - por exemplo, o número de *pixels* numa imagem é da ordem das centenas de milhar - mas também ocorre em problemas de outras áreas. Em situações onde os parâmetros de interesse têm relativamente poucos graus de liberdade e variam de forma suave, o problema pode simplificar-se consideravelmente através da utilização de informação geométrica.

Assumindo que as trajectórias observadas estão restritas a uma variedade (aparte ruído de observação), embebida no espaço das observações, a utilização de restrições geométricas deverá resultar em maior robustez e menor complexidade computacional [2]. Esta abordagem tem sido seguida noutros trabalhos, por exemplo o ISOMAP [3] e o LLE [4].

O presente artigo prossegue trabalho anterior ([5], [6], [7]), onde o algoritmo *Tangent Bundle Approximation* (TBA), ou Aproximação do Fibrado Tangente foi introduzido.

O algoritmo TBA é válido para qualquer topologia da variedade e para qualquer número de dimensões e co-dimensões da mesma. Assume-se que a variedade é orientável e compacta, ou seja, admite campos vectoriais normais e pode ser coberta por um atlas finito. Pode existir ou não fronteira, e a

¹ Este trabalho foi em parte suportado pelo projecto FCT POCTI 37844.

curvatura Gaussiana não precisa de ser nula. O fibrado tangente de uma variedade é o conjunto de todos os hiperplanos tangentes em todos os pontos dessa variedade [8]. Como existe um número infinito de pontos e hiperplanos, é necessária uma aproximação discreta e finita, a qual é fornecida pelo TBA. Isto simplifica a posterior estimação dos mapas da variedade.

Nas duas secções seguintes dá-se uma visão geral da abordagem seguida, bem como uma descrição mais detalhada do algoritmo. Depois apresentam-se resultados experimentais e conclusões.

Visão geral

Pretende-se, a partir de um conjunto de treino constituído por pontos não ordenados segundo qualquer grelha regular, seguir trajectórias pertencentes à mesma variedade que os pontos de treino. Mostra-se um exemplo na Figura 1, onde as observações estão dispostas num toro. A dimensão intrínseca é 2.

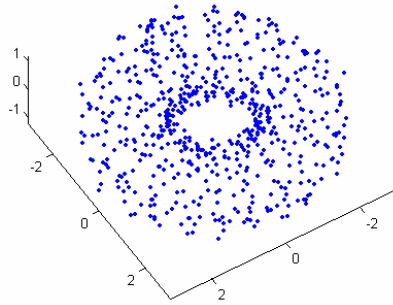


Fig. 1. Exemplo de um conjunto de treino: observações ruidosas y num toro.

Outro exemplo é o ilustrado na Figura 2, que consiste em imagens a níveis de cinzento de faces, todas da mesma pessoa, mas com diferentes poses e expressões faciais. As imagens foram recortadas a partir de uma sequência vídeo e têm resolução 71 por 71 *pixels*, o que resulta num espaço de observações com dimensão 5041. É possível reduzir esta dimensionalidade para valores entre 2 e 4. Este exemplo será também abordado na secção de resultados experimentais.



Fig. 2. Exemplo de um conjunto de treino: algumas imagens 71x71 de faces recortadas de um vídeo.

Estão envolvidos os seguintes passos:

- Aproximação do fibrado tangente à variedade;
- Estimação de parametrizações e mapas locais.

Cada um destes passos será abordado na secção respectiva.

A variedade M é modelada a partir de um conjunto de diffeomorfismos $g_i : U_i \subset R^n \rightarrow R^m$, onde n é a dimensão intrínseca, m é a dimensão do espaço envolvente e os U_i são conjuntos abertos. Estas funções são *mapas* da variedade. Juntas, as imagens $g_i(U_i)$, também chamadas *patches*, cobrem a variedade. Como diffeomorfismos, os mapas admitem inversas, g_i^{-1} , denominadas *parametrizações*.

Em primeiro lugar, são calculados os maiores valores e vectores próprios da matriz de covariância local numa vizinhança de cada um dos dados. Estes vectores próprios mais significativos definem um subespaço tangente à variedade, enquanto os restantes vectores próprios definem um subespaço normal.

Em segundo lugar, são encontrados *patches*. Isto é feito através de um procedimento de crescimento de região. Cada *patch* cresce por adição de todos os pontos vizinhos em que o subespaço tangente não

se desvie, em ângulo principal, mais do que um limiar (no máximo, $\pi/2$) em relação ao ponto inicial. Um dado ponto pode pertencer a mais do que um *patch*. Este procedimento resulta numa partição da variedade, a partir da qual se torna possível encontrar sistema de coordenadas locais, usando Análise em Componentes Principais (PCA: *Principal Component Analysis*).

Podem agora estimar-se mapas locais, o que neste trabalho é feito recorrendo a interpolação com *thin-plate splines*. Devido às restrições angulares introduzidas, garante-se que os mapas são bijetivos, e são também diferenciáveis, pelas propriedades das *thin-plate splines*. O seguimento das trajectórias pode agora ser feito nos domínios paramétricos de baixa dimensão.

Aproximação do fibrado tangente

Tal como o nome indica, este algoritmo estima subespaços tangentes a partir das observações $y_i=(y^1, \dots, y^m)$, permitindo dispôr de domínios paramétricos naturais para parametrização da variedade. Quando se parametriza uma variedade com topologia arbitrária, pode ser necessário mais do que um *patch* - trata-se do chamado “dilema do cartógrafo”.

Cada *patch* fica associado ao melhor (em termos de mínimo erro quadrático) subespaço tangente, estimado via PCA, e o conjunto de todos os subespaços tangentes serve como aproximação do fibrado tangente da variedade. Cada subespaço tangente serve como sistema de coordenadas locais.

A fim de tornar mais fácil a posterior estimação dos mapas, é conveniente que uma simples projecção forneça uma relação de um-para-um entre o subespaço tangente e o correspondente *patch* da variedade. Isto pode ser assegurado se não se permitir, dentro do *patch*, um desvio de ângulo principal dos subespaços tangentes superior a um determinado limiar τ , que nunca deve ser superior a $\pi/2$. É portanto necessário calcular os subespaços tangentes e os ângulos principais entre eles.

Começa-se por, para todos os pontos, encontrar os n vectores próprios mais significativos. Isto equivale a encontrar a dimensão intrínseca, o que só por si é um dos objectivos importantes deste trabalho. Devido à presença de ruído e também a facto de ser frequente a escassez de pontos em certas zonas do espaço de observações, optou-se por se calcular de forma automática o chamado “joelho” do gráfico de valores próprios na vizinhança de cada observação, ou seja o número n tal que o $(n+1)$ -ésimo valor próprio apresenta a maior descida, e depois calcular a mediana para todas as observações. O resultado é usado como estimativa da dimensão intrínseca da variedade.

Assume-se que as vizinhanças têm raio ε , que os dados são suficientemente densos para não deixar lacunas maiores que um valor δ e que o ruído não é maior do que σ . O raio ε deve ter em conta δ e σ . Após serem calculados os subespaços tangentes, pode aplicar-se o algoritmo descrito pelo seguinte pseudo-código:

```

while M not covered
  P = new patch
  y0 = choose a new seed from data points not in
        any patch
  {n0} = subspace basis at y0
  while NOT all points visited
    y1 = choose nearest neighbor
    {n1} = subspace basis at y1
    if principal_angle(n0,n1) < τ AND
       distance(y0,y1) < ε
      append y1 to P
    end if
  end while
end while

```

O resultado final é a cobertura de M por um número finito p de *patches*. Dentro de cada *patch*, o ângulo principal não varia mais do que τ , e o teste de distância permite assegurar que cada *patch* é conexo.

Mapas e parametrizações

Existem diversas técnicas para aproximação de funções não-lineares que permitem garantir continuidade e diferenciabilidade, como se pretende neste caso. Os resultados aqui apresentados baseiam-se na utilização de *thin-plate splines* [9], mas podem ser utilizadas outras técnicas, desde que não exijam a grelhas regulares. Funções de base radiais (RBFs: Radial Basis Functions) ou perceções multi-camada (MLPs: Multi-Layer Perceptrons) são alternativas possíveis.

Tendo conseguido previamente uma partição da variedade M em p *patches*, é possível encontrar p mapas $g_i(\mathbf{x})$. Em resumo, a partir de pontos $\mathbf{y}=(y_1, \dots, y_m)$ pertencentes a um dado *patch* i , e tendo efectuado PCA, ficam disponíveis uma matrix V_i de vectores próprios e um vector de média μ_i . Projectar \mathbf{y} no subespaço associado ao *patch* i é uma questão de efectuar:

$$\tilde{\mathbf{x}} = V_i^T (\mathbf{y} - \mu_i) \quad (1)$$

$$\mathbf{x} = \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_n \end{bmatrix}. \quad (2)$$

A equação (1) é uma isometria, neste caso uma translação seguida de uma rotação, enquanto (2) equivale a descartar as $m-n$ coordenadas correspondentes à co-dimensão de M . Quanto ao mapa g_i , o qual corresponde à inversa de (2), obedece à expressão

$$g_i(\mathbf{x}) = V_i \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ \tilde{g}(\mathbf{x}) \end{bmatrix} + \mu_i \quad (3)$$

Os restantes $m-n$ componentes de $\tilde{\mathbf{x}}$, em vez de serem postos a zero, o que resultaria numa aproximação linear por troços algo grosseira, são portanto preservados. Localmente, a variedade é parametrizada por

$$\mathbf{x} \rightarrow \begin{bmatrix} x \\ \tilde{g}(\mathbf{x}) \end{bmatrix}^T \quad (4)$$

com \tilde{g} aproximada por *thin-plate spline*.

Resultados experimentais

Partindo de um conjunto de treino constituído por $M=194$ imagens de níveis de cinzento com resolução 71×71 , recortadas de uma sequência de vídeo em formato PAL (768 por 576 *pixels*, 25 fotogramas por segundo). As imagens correspondem à face de uma mesma pessoa, como se mostra na Figura 2, e pretende-se aprender a variedade de faces admissíveis. Foi subtraída a imagem média, pelo que o conjunto de treino fica com média nula.

Cada imagem Φ tem $N=71 \times 71=5041$ *pixels*, o que resulta num problema a 5041 dimensões com apenas 194 observações. Felizmente, é possível aplicar o algoritmo *eigenfaces* [10], evitando o cálculo de matrices de covariância $N \times N$ e respectivos valores e vectores próprios.

Sucintamente, os M vectores \mathbf{u}_k mais relevantes da covariância

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = AA^T \quad (7)$$

com $A = [\Phi_1 \cdots \Phi_M]$, podem ser extraídos dos vectores próprios de $L=A^T A$, que é uma matrix $M \times M$. Obtém-se uma outra matrix U , de tamanho $N \times M$, cujas colunas são os \mathbf{u}_k .

Uma análise aos valores próprios revela que se podem representar todas as faces como combinações lineares dos primeiros $M'=15$ vectores próprios, (as chamadas *eigenfaces*) sem grande perda de informação. As 15 *eigenfaces* estão representadas na Figura 5.

Aplica-se uma transformação linear

$$W = U^T A \quad (9)$$

aos dados e guardam-se apenas as primeiras M' linhas de W reduzindo desde logo a dimensão das observações de 5041 para 15. As M colunas de W constituem os novos vectores de treino, aos quais se pode aplicar o TBA.



Fig. 5. As 15 *eigenfaces*.

Após o algoritmo TBA, resulta uma variedade com 2 *patches*, sendo a dimensão intrínseca estimada $n=3$. Para facilidade de representação, e por permitir ainda uma qualidade de reconstrução aceitável, usa-se $n=2$. A Figura 6 mostra as primeiras 3 coordenadas do *patch* que cobre mais amostras.

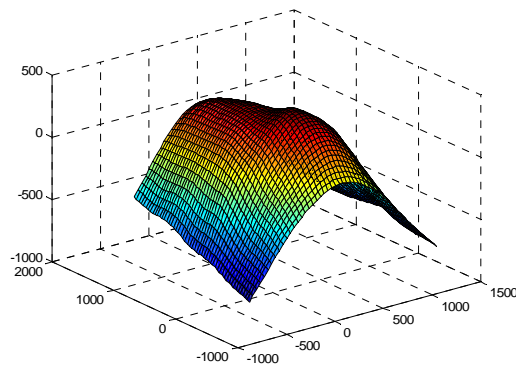


Fig. 6. First three dimensions for one of the manifold patches found by TBA.

O segundo *patch* cobre uma região com reduzido número de observações, relativas a imagens nas quais a apresentadora olha para baixo para consultar notas. Isto mesmo pode ver-se na Figura 7, que mostra as duas coordenadas reduzidas dos dados.

Na Figura 8 mostra-se uma trajectória projectada no espaço de baixa dimensão, correspondendo a 6 fotogramas consecutivos. É possível ver a sequência original e a sua reconstrução baseada apenas nas duas coordenadas fornecidas pelo TBA.

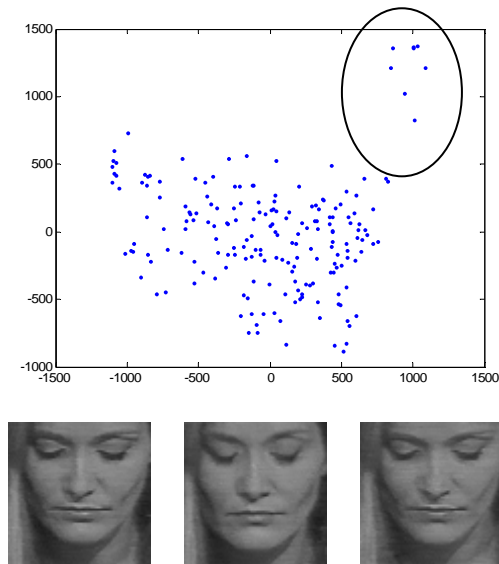


Fig. 7. Acima: Projecção do conjunto de treino nas duas dimensões reduzidas. O *cluster* marcado no topo direito representa imagens nas quais a apresentadora olha para as suas notas (como se mostra abaixo).

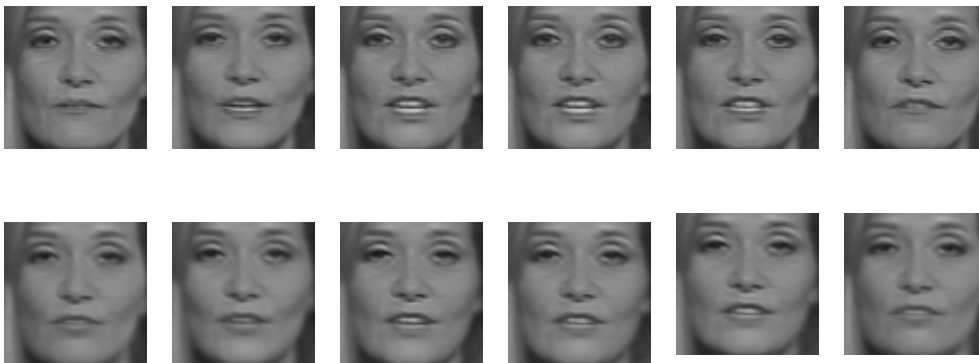
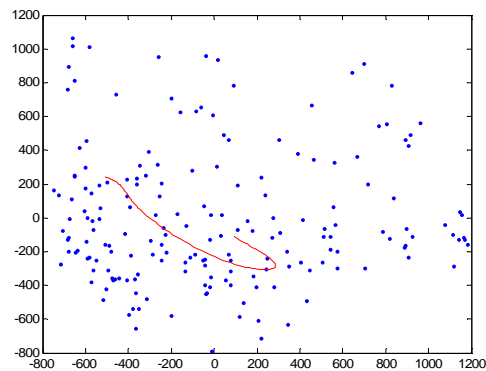


Fig. 8. Acima: Trajectória (intepolada). A meio: Sequência original. Abaixo: Sequência reconstruída.

Conclusões

Neste artigo apresenta-se uma aplicação do algoritmo TBA ao seguimento de movimento em seqüências vídeo. O algoritmo aproveita informação geométrica, conseguindo-se uma considerável redução da dimensionalidade do problema e mantendo uma qualidade de reconstrução aceitável. Apresentam-se resultados para um exemplo de seguimento de faces.

Esta abordagem permite utilização posterior de técnicas como filtragem de Kalman ou filtros de partículas para seguimento, reconstrução e síntese do movimento, num espaço de dimensão reduzida, o que permite ganhos computacionais e de robustez face ao ruído. Está em curso trabalho adicional relativo à identificação dos comportamentos dinâmicos no espaço de dimensão reduzida.

Referências

- [1] H. Ladesmäki, O. Yli-Harja, W. Zhang e I. Shmulevich, “Intrinsic Dimensionality in Gene Expression Analysis”, *GENSIPS 2005*.
- [2] J. Marques, J. M. Lemos e A. Abrantes, “Estimation of random trajectories on manifolds: Application to object tracking”, *Proceedings of the IEEE International Conference on Image Processing*, pp. 103-107.
- [3] J. B. Tenenbaum, V. de Silva e J. C. Langford, “A Global Geometric Framework for Nonlinear Dimensionality Reduction”, *Science*, 2000.
- [4] S. T. Roweis e L. K. Saul. “Nonlinear Dimensionality Reduction by Locally Linear Embedding”, *Science*, 2000.
- [5] J. Silva, J. Marques e J. M. Lemos. “A geometric approach to motion tracking in manifolds”. *Proceedings of the 13th IFAC Symposium on System Identification*, 2003.
- [6] J. Silva, J. Marques and J. M. Lemos. “Non-Linear Dimension Reduction with Tangent Bundle Approximation”. *ICASSP 2005*.

- [8] B. O’Neill. *Elementary Differential Geometry*, Academic Press, 1997.
- [9] T. Duchamp e W. Stuetzle. “Spline Smoothing on Surfaces”, *Proceedings of ACM Siggraph*, 2002.
- [10] M. Turk e A. Pentland. “Eigenfaces for recognition”, *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71-86.