

Homomorfismo entre Grafos para Segmentação e Reconhecimento de Características Faciais em Sequências de Vídeo: Resultados Preliminares

ANA BEATRIZ V. GRACIANO¹, ROBERTO MARCONDES CESAR JR.¹, ISABELLE BLOCH³

¹Departamento de Ciência da Computação, IME, Universidade de São Paulo. São Paulo, Brasil.
{cesar, abvg}@ime.usp.br

²Departamento TSI, CNRS URA 80, École Nationale Supérieure des Télécommunications. Paris, França.
Isabelle.Bloch@enst.fr

Abstract. Este artigo apresenta uma metodologia baseada em homomorfismo entre grafos para segmentação e reconhecimento de características faciais em sequências de vídeo. Tal metodologia estende uma técnica previamente proposta para imagens estáticas de modo a aplicá-la a sequências de vídeo. Essa generalização se dá através da incorporação do aspecto temporal inerente a tais sequências. Características faciais são representadas por *grafos relacionais atribuídos* e um modelo de referência é utilizado para assinalar as características de interesse. O processo de reconhecimento, por sua vez, consiste na busca por um homomorfismo ótimo entre os grafos correspondentes à imagem modelo e ao quadro corrente da sequência de vídeo.

1 Introdução

Este artigo aborda a segmentação e o reconhecimento de características faciais em sequências de vídeo digitais através de homomorfismos entre grafos. A técnica proposta consiste nos primeiros passos para a extensão da metodologia apresentada para imagens estáticas em [1, 2] e foi motivada pelo potencial de aplicação apresentado, tais como em biometria, bancos de dados multimeios e interação humano-computador inteligente. Outras abordagens para o problema incluem o uso de fluxo óptico [9], redes de Gabor Wavelets [6] e outras redes neurais [8].

De modo geral, características faciais são modeladas através de um *grafo relacional atribuído*, ou ARG (do inglês, attributed relational graph). Esses grafos são obtidos a partir da imagem do gradiente supersegmentado derivado de cada quadro da sequência. Já o processo de reconhecimento das características faciais consiste em buscar um homomorfismo apropriado entre o grafo de um quadro da sequência e aquele de uma imagem modelo em que as características faciais de interesse foram previamente escolhidas.

A Seção 2 aborda os aspectos de representação da face. A Seção 3 explica o passo de busca por homomorfismos entre ARGs, enquanto a Seção 4 explora o processo de rastreamento de características faciais ao longo da sequência de vídeo. Por fim, a Seção 5 apresenta alguns resultados e a Seção 6, conclusões.

2 Representação da Face

Grafos Relacionais Atribuídos. Neste trabalho, um grafo dirigido será representado por $\tilde{G} = (N, E)$, onde N representa o conjunto de vértices de \tilde{G} e $E \subseteq N \times N$, seu conjun-

to de arestas. Dois vértices a, b de N são ditos adjacentes se $(a, b) \in E$. Se todos os vértices de \tilde{G} são adjacentes entre si, então \tilde{G} é denominado completo. Ainda, $|N|$ e $|E|$ denotam, respectivamente, o número de vértices e de arestas de G .

Um *grafo relacional atribuído* é um grafo a cujos vértices e arestas são associados vetores de atributos. Formalmente, um ARG é definido como $G = (N, E, \mu, \nu)$, onde N e $E \subseteq N \times N$ representam, respectivamente, o conjunto de vértices e de arestas de G . Ainda, $\mu : N \rightarrow L_N$, também chamado de atributo de objeto, é associado a cada vértice de G , enquanto $\nu : E \rightarrow L_E$, ou atributo relacional, é associado a cada aresta.

Para os fins deste trabalho, vértices representam regiões conexas de características faciais e arestas denotam relações entre as mesmas, ou seja, carregam informações sobre o arranjo espacial de tais características.

Atributos. Para a discussão a seguir, consideremos um ARG $G = (N, E, \mu, \nu)$ e dois vértices quaisquer a, b de N .

O **atributo de objeto** $\mu(a)$ é definido como:

$$\mu(a) = (g(a), w(a), l(a)) \quad (1)$$

O termo $g(a)$ indica o nível de cinza médio da região da imagem associada ao vértice a , enquanto $w(a)$ é um coeficiente obtido através da aplicação de uma wavelet de Morlet. Além disso, $g(a)$ e $w(a)$ são normalizados entre 0 e 1 em relação ao maior nível de cinza possível. Finalmente, $l(a)$ é um rótulo de região.

O **atributo relacional** $\nu(a, b)$, para a, b em E , é defi-

nido como:

$$\nu(a, b) = (\vec{v}, \text{sym}(a, b)). \quad (2)$$

O atributo \vec{v} denota o vetor $(p_b - p_a)/2d_{max}$, em que d_{max} é a distância máxima entre quaisquer dois vértices do grafo da imagem de entrada enquanto p_a e p_b são os centróides das regiões às quais os vértices a e b correspondem. O termo $\text{sym}(a, b)$ representa uma simetria reflexiva calculada como descrito em [3].

O Modelo de Referência da Face. Uma imagem-modelo da face é usada como referência para reconhecer características faciais de interesse. Tipicamente, é utilizado o primeiro quadro da sequência a ser processada, que é segmentado manualmente em regiões relevantes. Ainda, seu respectivo ARG é obtido e deve conter vértices associados a cada característica facial que se deseja rastrear (ex: íris, lábios, sobrancelhas).

3 O Processo de Reconhecimento de Características Faciais

Homomorfismo entre Grafos. Considere dois ARGs $G_1 = (N_1, E_1, \mu_1, \nu_1)$, proveniente de um quadro da sequência, e $G_2 = (N_2, E_2, \mu_2, \nu_2)$, derivado do modelo. Eles serão chamados, respectivamente, de grafo de entrada e grafo do modelo, e terão seus vértices, arcos e atributos indexados pelos mesmos valores que os identificam.

Um *grafo de associação* \tilde{G}_A entre G_1 e G_2 é definido como o grafo completo $\tilde{G}_A = (N_A, E_A)$, onde $N_A = N_1 \times N_2$ e $E_A = E_1 \times E_2$.

Um *homomorfismo* h entre G_1 e G_2 é um mapeamento $h: N_1 \rightarrow N_2$ tal que, $\forall a_1 \in N_1, \forall b_1 \in N_1$, se $(a_1, b_1) \in E_1 \Rightarrow (h(a_1), h(b_1)) \in E_2$. Essa definição pressupõe que todos os vértices em G_1 serão mapeados em G_2 . No caso deste trabalho, tal mapeamento corresponde à fusão de subregiões coerentes na imagem de entrada supersegmentada e, portanto, a obtenção de um homomorfismo apropriado entre G_1 e G_2 é essencial.

Como proposto em [1], uma solução para encontrar um homomorfismo entre G_1 e G_2 pode ser definida como um sub-grafo completo $\tilde{G}_S = (N_S, E_S)$ do grafo de associação \tilde{G}_A , em que $N_S = \{(a_1, a_2), a_1 \in N_1, a_2 \in N_2\}$ é tal que $\forall a_1 \in N_1, \exists a_2 \in N_2, (a_1, a_2) \in E_S$, e $\forall (a_1, a_2) \in E_S, \forall (a_1', a_2') \in E_S, a_1 = a_1' \Rightarrow a_2 = a_2'$, garantindo que cada vértice do grafo de entrada corresponda a exatamente um vértice do grafo do modelo e $|N_S| = |N_1|$. Note que tal solução considera apenas as estruturas de G_1 e G_2 e que há outros homomorfismos possíveis.

Função Critério. Para avaliar a qualidade e a adequação de um dado homomorfismo entre os grafos de entrada e do

modelo, é necessário definir uma função critério que considere não somente as estruturas dos grafos, mas também as relações entre os atributos das características faciais. Neste trabalho, a adequação de um homomorfismo é analisada minimizando-se a seguinte função critério:

$$f(\tilde{G}_S) = \frac{\alpha}{|N_S|} \sum_{(a_1, a_2) \in N_S} c_N(a_1, a_2) + \frac{(1-\alpha)}{|E_S|} \sum_{e \in E_S} c_E(e) \quad (3)$$

onde c_N e c_E são as seguintes medidas de dissimilaridade:

$$c_N(a_1, a_2) = \left\{ \begin{array}{l} \gamma_N |g_1(a_1) - g_2(a_2)| + \\ (1 - \gamma_N) |w_1(a_1) - w_2(a_2)|, \\ \text{se } l(a_1) = l(a_2) \\ \infty, \text{ c.c} \end{array} \right\} \quad (4)$$

$$c_E(e) = \gamma_E \phi_v + (1 - \gamma_E) \phi_{sym} \quad (5)$$

e ϕ_v, ϕ_{sym} são definidos como:

$$\phi_v = \gamma_v \left(\|\vec{v}_1\| - \|\vec{v}_2\| \right) + (1 - \gamma_v) \frac{|\cos \theta - 1|}{2} \quad (6)$$

$$\phi_{sym} = |\text{sym}(a_1, b_1) - \text{sym}(a_2, b_2)|.$$

Nesse caso, $\cos \theta = \frac{\vec{v}_1 \cdot \vec{v}_2}{\|\vec{v}_1\| \|\vec{v}_2\|}$ e os valores γ_N, γ_E e γ_v são pesos.

Encontrando um Homomorfismo. Neste trabalho, a procura por um homomorfismo adequado entre os grafos de entrada e do modelo é realizada a partir de um algoritmo de otimização baseado em busca em árvores. Outras alternativas possíveis são abordadas em [1], [4] e [5].

Em linhas gerais, tal algoritmo gera uma árvore de busca em que cada vértice representa um par de vértices (k, l) , $k \in N_1$ e $l \in N_2$. O vértice raiz é rotulado com $(0, 0)$ e expandido em $|N_2|$ filhos rotulados com $(1, l)$, $l = 1 \dots |N_2|$. A cada passo k do algoritmo, o filho que minimiza a função critério, digamos (k, l_{min}) , é escolhido e também expandido em $|N_2|$ filhos $(k+1, i)$, $i = 1 \dots |N_2|$. Esse processo é repetido até que um vértice $(|N_1|, l)$ seja alcançado, o que garante que todos os vértices de G_1 tenham sido mapeados em um vértice de G_2 , estabelecendo um homomorfismo entre os grafos de entrada e do modelo.

4 O Processo de Rastreamento

Nesta seção, serão apresentados o processo geral para rastreamento de características faciais ao longo de uma sequência de vídeo, bem como as contribuições deste trabalho para a generalização do que foi proposto em [1].

Processo Geral. A Figura 1 apresenta a seqüência geral de etapas para a realização da segmentação e do reconhecimento de características faciais num quadro qualquer da seqüência de vídeo.

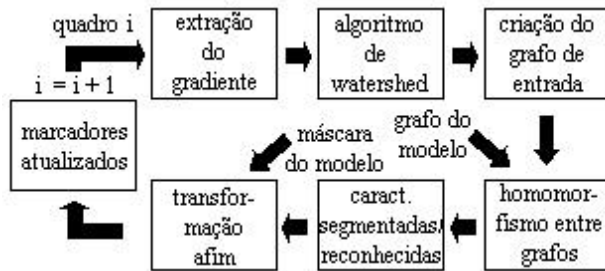


Figura 1: Esquema geral do processo de rastreamento.

Inicialmente, são localizados marcadores aproximados (*landmarks*) para as características faciais de interesse no quadro, os quais serão utilizados posteriormente para restringir a região em que a imagem deverá ser supersegmentada. Tais marcadores são obtidos através do uso de uma rede de Gabor wavelets (GWN - *Gabor Wavelet Networks*) [6], caracterizada como um processo de rastreamento global de faces. O passo seguinte consiste na aplicação do algoritmo proposto para imagens estáticas.

Uma das contribuições mais relevantes à metodologia está relacionada à atualização dos marcadores que serão usados no quadro subsequente da seqüência de vídeo, permitindo a independência do processo em relação ao uso da GWN.

Como uma seqüência de vídeo possui um caráter dinâmico e suas imagens, em geral, variam em função do tempo, a atualização dos marcadores mostra-se importante para a manutenção da qualidade da segmentação e para o rastreamento adequado das características faciais.

Atualização dos Marcadores. Embora a técnica GWN possa ser aplicada a cada quadro da seqüência para atualizar os marcadores, seria interessante aproveitar as informações obtidas diretamente dos grafos e da imagem-modelo. Para tanto, uma *transformação afin* é aplicada para mapear a imagem-modelo ao quadro atual baseando-se nas características faciais reconhecidas e, então, atualizando-se os marcadores.

No primeiro quadro da seqüência, os marcadores previamente obtidos relativos ao modelo são também usados para essa imagem. Nos quadros subsequentes, após o término do processo de reconhecimento, são calculados os centróides das regiões das características faciais de interesse. Além disso, são calculados os centróides das regiões pré-definidas do modelo. Então, a transformação afin que me-

lhor mapeia o conjunto de centróides do modelo para o conjunto relativo ao quadro considerado é estimada e aplicada usando-se a seguinte fórmula [7]:

$$\vec{q} = \alpha(A\vec{s} + \vec{b}) \quad (7)$$

onde A corresponde a uma matriz 2×2 não-singular representando a transformação procurada, α é um valor escalar qualquer, e \vec{q} , \vec{s} são os vetores das coordenadas dos centróides do quadro e do modelo respectivamente.

Essa transformação afin permite, além da atualização dos marcadores, a projeção da imagem-modelo sobre o quadro segmentado e reconhecido, permitindo uma avaliação visual do processo de reconhecimento.

Possíveis Extensões. Embora a mudança supracitada torne a metodologia mais robusta para ser aplicada a seqüências de vídeo, nossa pesquisa em andamento objetiva um maior aproveitamento das informações possivelmente redundantes presentes em quadros distintos.

Possíveis abordagens incluem a inserção de arestas temporais à estrutura dos ARGS, as quais representariam transições e relações entre quadros consecutivos da seqüência e permitiriam recalcular atributos de objeto e relacionais, ou ainda adicionar novas características faciais ao rastreamento durante o processo, tornando o modelo desnecessário.

Além disso, o resultado obtido na busca de um homomorfismo entre o grafo modelo e aquele correspondente ao do i -ésimo quadro poderia ser reutilizado como solução inicial da busca a ser realizada no quadro $i + 1$, reduzindo-se a expansão dos nós e considerando-se transições suaves entre quadros consecutivos.

5 Resultados

Nesta seção, mostramos alguns dos primeiros resultados obtidos a partir da aplicação dos novos passos apresentados na seção anterior.

Os testes foram realizados a partir de seqüências de vídeo com mudanças consideráveis na face ao longo do tempo, tais como sorrisos, piscar de olhos, movimentação da cabeça, entre outras.

A Figura 2 exhibe os resultados obtidos da projeção da máscara do modelo sobre alguns quadros de uma seqüência composta por 96 imagens coloridas de tamanho 512×512 , que foram convertidas para níveis de cinza para o propósito do algoritmo. O tempo de processamento da seqüência completa foi de, aproximadamente, duas horas e meia num computador com processador de 1.8Ghz e 1Gb de RAM. Como esta primeira etapa pretendia avaliar a metodologia em relação à qualidade dos resultados, o desempenho computacional não recebeu atenção especial. Porém, esse aspecto será considerado na continuação do trabalho.

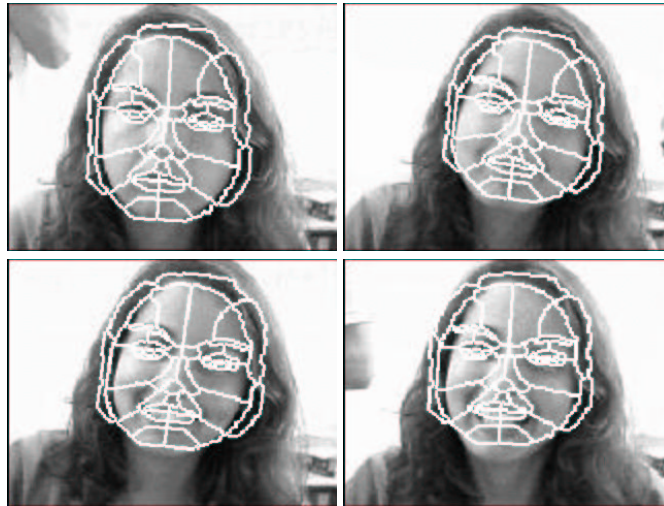


Figura 2: Máscaras do modelo projetadas sobre seus respectivos quadros segmentados e reconhecidos, de acordo com as transformações afins encontradas.

É possível verificar que a máscara do modelo é satisfatoriamente projetada sobre a face, permitindo o rastreamento da mesma ao longo da sequência. Ainda, grande parte das características faciais que podem ser de interesse (ex: sobrancelhas, narinas, nariz, lábios) são corretamente rastreadas, comprovando a validade do processo de reconhecimento e da atualização dos marcadores.

No entanto, certos refinamentos ainda são necessários à metodologia, especialmente em relação a mudanças bruscas entre quadros e às ações que devem ser tomadas quando novas características faciais são introduzidas na sequência.

6 Conclusões

Neste artigo, foi proposto um primeiro passo para a generalização da metodologia proposta em [1]. Os resultados preliminares mostraram que é possível refletir as mudanças nas características faciais entre quadros através de transformações afins. Apesar dos resultados encorajadores provenientes dessa modificação, as outras possibilidades listadas na Seção 4 são o tópico de nossa pesquisa em andamento.

Agradecimentos

Este trabalho foi parcialmente realizado com o auxílio de uma verba CAFES-COFECUB. A. Graciano recebeu apoio do CNPQ (132562/2003-0). R. Cesar recebeu apoio da FAPESP (99/12765-2) e do CNPQ (300722/98-2).

Referências

- [1] R. Cesar, E. Bengoetxea and I. Bloch. Inexact graph matching using stochastic optimization techniques for facial fea-

ture recognition. In *16th International Conference on Pattern Recognition*, volume 2, pages 465-468, August 2002.

- [2] R.M. Cesar Jr. and I. Bloch. First results on facial feature segmentation and recognition using graph homomorphisms. In *Proc. VI Simpósio Ibero-Americano de Reconhecimento de Padrões*, pages 95-99, Florianópolis, Brazil, 2001.
- [3] O. Colliot, A.V. Tuzikov, R.M. Cesar Jr. and I. Bloch. Approximate reflectional symmetries of fuzzy objects with an application in model-based object recognition. In Press. *Fuzzy Sets and Systems*, 2003.
- [4] A. Perchant and I. Bloch. Fuzzy morphisms between graphs. In *Fuzzy Sets and Systems*, 128(2):149-168, 2002.
- [5] R.C. Wilson and E.R. Hancock. A Bayesian compatibility model for graph matching. In *Pattern Recognition Letters*, 17(3):263-276, 1996.
- [6] V. Kruger and G. Sommer. Affine real-time face tracking using a wavelet network. In *Proc. of ICVV'99 Workshop Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, pages 141-148, Corfu, Greece, 1999.
- [7] R.M. Cesar Jr. and L. da F. Costa. *Shape Analysis and Classification - Theory and Practice*. CRC Press, 1 edition, 2001.
- [8] S.H. Lin, S.Y. Kung, and L.J. Lin. Face Recognition/Detection by Probabilistic Decision-Based Neural Network. In *IEEE Transactions on Neural Networks*, 8(1):114-132.
- [9] Y. Yacoob and L.S. Davis. Computing Spatio-Temporal Representations of Human Faces. In *CVPR94*, pages 70-75, 1994.