Boletim Técnico da Escola Politécnica da USP
Departamento de Engenharia de Telecomunicações
e Controle

ISSN 1517-3550

BT/PTC/0010

Reconhecimento Automático de Ações Faciais usando FACS e Redes Neurais Artificiais

Alexandre Tornice
Euvaldo F. Cabral Júnior

São Paulo – 2000

O presente trabalho é um resumo da dissertação de mestrado apresentada por Alexandre Tornice sob orientação do Prof. Dr. Euvaldo F. Cabral Júnior.: "Extração e Avaliação de Atributos do Eletrocardiograma para Classificação de Batimentos Cardíacos", defendida em 28/03/00, na Escola Politécnica.

A íntegra da dissertação encontra-se à disposição com o autor e na Biblioteca de Engenharia Elétrica da Escola Politécnica/USP.

FICHA CATALOGRÁFICA

Tornice, Alexandre
Reconhecimento automático de ações faciais usando FACS e redes neurais artificiais / A. Tornice, E.F. Cabral Júnior. -- São Paulo : EPUSP, 2000.
p. -- (Boletim Técnico da Escola Politécnica da USP, Departamento de Engenharia de Telecomunicações e Controle, BT/PTC/0010)

1. Redes neurais (Inteligência artificial) 2. Ações faciais I. Cabral Júnior, Euvaldo Ferreira II. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Telecomunicações e Controle III. Título IV. Série

ISSN 1517-3550

CDD 006.3
621.367

Reconhecimento Automático de Ações Faciais usando FACS e Redes Neurais Artificiais

Alexandre Tornice
USP
Escola Politécnica
Laboratório de Comunicação e Sinais
aletornice@ig.com.br

Euvaldo F. Cabral Jr.
USP
Escola Politécnica
Laboratório de Comunicação e Sinais
euvaldo@lcs.poli.usp.br

Abstract

The main goal of this work is to propose a way of automating facial actions recognition on images using artificial neural networks and FACS (Facial Action Coding System), a widely used tool that describes facial muscles motion. An image database consisting of 80 pictures was used to train the system and another 180 sample images selected from a 400 image samples database was used to test the system. Each image was first evaluated by a FACS coder expert and, once the system was trained, the automatic classification was compared to the one provided by a specialist.

A known method of feature extraction for frontal face images was used in combination to the spatial location of the head. That spatial location was obtained from the eyes and nose plain image coordinates, as well as the considered camera geometry. This new proposed technique has proved to be more robust than the simple feature extraction approach, but nevertheless the results were still not satisfactory.

We considered that the most common feature extraction methods presented in the literature have not shown good results because such chosen features were not optimized to detect facial actions. Instead of using the so called wrinkle measures of gray level segments at certain facial positions, bi-dimensional regions proved to be a good choice and seem to be perfectly applicable to face analysis. The comparison results between human and automatic classification is presented in the conclusions, using the improved new features together with the Multilayer Perceptron Neural Network, in the search for better classification rates.

Keywords: image, Facial Action Coding System, basic emotion, Multi-Layer Perceptron, facial action.

I. Introdução

O reconhecimento automático de expressões faciais em imagens digitais vem chamando a atenção de muitos pesquisadores da área de visão computacional e afins, devido às várias possibilidades existentes para aplicações nas comunicações, interação homem-máquina, medicina e psicologia.

Este trabalho tem como objetivo o desenvolvimento e o teste de desempenho de um método de análise facial automática utilizando um banco de imagens experimentais estáticas de expressões faciais, de modo a classificar movimentos musculares da face. Especificamente, foi estudada a região superior da face, ou seja, a região acima da base do nariz. Essa separação pode ser feita, pois os movimentos faciais superiores e inferiores apresentam uma certa independência, podendo assim, ser estudados separadamente.

I.1. Motivação e Justificativa

Emoções, ânimos e expressões faciais. Todos esses fenômenos, que ocorrem fisicamente a partir de uma dinâmica cerebral complexa, se apresentam como um mistério para qualquer instrumento de medição conhecido. As emoções e os estados de ânimo, por serem não mensuráveis diretamente e por serem algo não localizável, são chamados eventos privados. Apesar disso, pode-se ter uma idéia do estado de ânimo de um sujeito a partir de indicadores, como relatos verbais e ações faciais ou corporais.

A motivação desse trabalho está em se classificar um desses indicadores: o conjunto de movimentos faciais, que caracterizam as expressões faciais. Para isso, utiliza-se um sistema de codificação de ações faciais (FACS - *Facial Action Coding System*) [Ekman, Friesen, 78], que descreve os movimentos musculares faciais visíveis de intensidade suficiente ou acima de um certo nível de aceitação. Ou seja, trata-se de um sistema puramente descritivo da dinâmica muscular facial.

A partir de uma imagem do rosto de uma pessoa, pode-se identificar movimentos musculares, ou ações faciais, que ocorrem no momento em que foi obtida essa imagem. As ações faciais são identificadas sempre em relação a uma imagem-base da face do sujeito, anteriormente estabelecida, representando um momento no qual a pessoa não

esteja apresentando ação facial visível alguma. Esta imagem é chamada face neutra. Um movimento muscular da face ou movimento da cabeça ou olhos que pode ser identificado isoladamente é denominado "unidade de ação" (ver tabela em anexo). O conjunto das unidades de ação faciais presentes numa foto ou imagem de um rosto constitui o código FACS daquela face.

De posse da informação estrutural dada pelo código, pode-se interpretar as ações faciais como expressões faciais de emoções. Além disso, cada emoção pode ser expressada através de mais de uma ação facial ou combinações delas, ou ainda de expressões faciais semelhantes [Ekman, Friesen, Ancoli, 80]. Por isso, é importante que se use ações faciais ao invés de expressões faciais genéricas arbitrárias para se analisar a face.

I.2. Metodologia

O processo da análise facial automática das imagens consiste no seguinte:

1) Aquisição ou captura de imagens de expressões faciais em laboratório;
2) Preprocessamento das imagens: filtragem de ruído e normalização de brilho;
3) Segmentação facial ou, mais especificamente, localização dos pontos dos extremos externos dos olhos e do ponto subnasal, sobre o plano da imagem;
4) Localização de segmentos em regiões faciais onde ocorrem movimentos musculares e enrugamentos da pele (sobre a testa e regiões próximas aos olhos), com base nas coordenadas tridimensionais dos pontos mencionados no item anterior;
5) As chamadas medidas de enrugamento são extraídas de cada segmento facial, obtendo-se um vetor de atributos;
6) Classificação dos atributos, através de uma rede neural artificial, em unidades de ação facial.

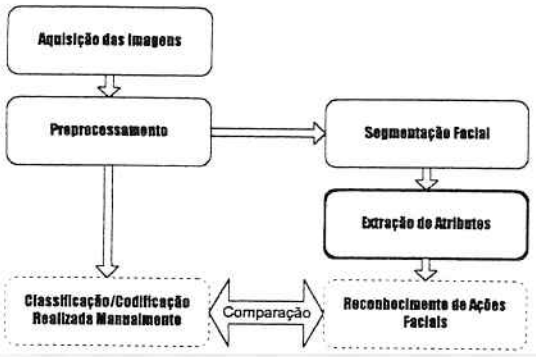

Fig. I: Diagrama descritivo da estrutura do trabalho.

Ao final do processo, os resultados obtidos com a análise automática são comparados com os códigos FACS determinados por um observador especializado. Todo o processo está ilustrado na figura I.1, e será detalhado adiante.

A comparação entre a imagem da face neutra e a de uma imagem de face expressiva de um sujeito é realizada através de processamentos que envolvem os ajustes, a normalização e, se necessário, algumas transformações geométricas para compensar a posição da cabeça, caso não esteja de frente.

II. Processo de Aquisição das Imagens

Para se construir um sistema de reconhecimento de ações faciais baseado em uma rede neural artificial, necessita-se que esse sistema seja previamente treinado com informações vindas de *imagens de treino*, para depois ser testado, com *imagens de teste*.

II.1. Imagens de Teste

As imagens de ações faciais de quarenta sujeitos de variados sexo, idade e etnia, foram capturadas de gravações em videoteipe, realizadas com uma câmera oculta atrás de um espelho unidirecional [Gil, 93]. O experimento consistiu em fazer com que cada sujeito, deixado sozinho numa sala, respondesse a uma Lista de Estados de ânimos Presentes (LEP) [Engelmann, 86], cujas locuções, como por exemplo "estou triste" ou "estou surpreso", eram projetadas de maneira que o sujeito, ao lê-las, olhasse em direção à câmera. As imagens foram capturadas, ou digitalizadas, nos instantes em que a expressão facial estava no ápice, após ou durante a leitura da locução.

Para cada sujeito foi preestabelecida uma imagem de sua face neutra, para que fosse feita a codificação de suas imagens expressivas. A comparação de cada par de imagens, face expressiva-face neutra, foi realizada pelo especialista com a ajuda do computador, de modo a superpor as imagens e dispô-las lado a lado, vertical e horizontalmente.

Todo esse processo foi realizado para cada uma das 10 imagens de expressões faciais de cada um dos quarenta sujeitos, sendo identificados os 400 códigos correspondentes. Selecionou-se, a partir disso, as imagens correspondentes às expressões cujos movimentos faciais apresentavam maior intensidade, totalizando 180.

As resoluções finais das imagens foram reduzidas para 440x330 pixels, com 256 níveis de cinza. Esse número de pontos foi escolhido dentre as

opções existentes de modo a oferecer informações visuais razoáveis do rosto focalizado, possibilitando a detecção de detalhes da movimentação facial. As larguras das faces dos sujeitos nessas imagens ficaram em torno de 100 pixels, o que é aceitável, pois para se realizar uma análise facial automática, seriam necessários pelo menos 50 pixels EKMAN et al. (1993).

II.2. Imagens de Treino

As imagens utilizadas no treinamento do sistema foram obtidas de maneira semelhante às de teste, com relação ao método de captura, porém os movimentos faciais do material são propositais, ao contrário das imagens de teste. Alguns quadros de um videoteipe (vídeo demonstrativo - apêndice 2), com demonstrações dos movimentos de cada unidade de ação facial e de algumas combinações importantes delas (apêndice 2), foram capturados ou digitalizados. As imagens finais têm 256x256 pontos, possuindo 256 níveis de cinza. As larguras das faces ficaram em torno de 200 pixels, portanto o dobro do tamanho das imagens de teste.

Os quadros do videoteipe foram capturados de maneira a formarem uma seqüência de imagens para cada movimento facial. O total de imagens em cada seqüência ficou em torno de 5, mostrando a evolução do movimento da musculatura desde a face neutra até o ápice.

As imagens de treino não foram escolhidas para este fim casualmente. Elas estão em melhores condições do que as imagens de teste, pois a face do sujeito está sempre de frente e os movimentos musculares são controlados. Por isso, pôde-se capturar imagens com ações faciais desde pouco até bastante intensas, possibilitando a construção do conjunto de treinamento para a RNA mais completa. Inclusive, devido à maior proximidade do sujeito da câmera, a resolução é maior. Como foi citado, as faces desse conjunto têm o dobro do tamanho, em média, das do conjunto de teste. As unidades de ação abrangidas pelo conjunto de imagens de treino não são numerosas (apêndice 2), mas devido às várias imagens que formam as seqüências temporais das movimentações, ao todo formam um conjunto de 83 imagens.

III. Preprocessamento

Em cada imagem da base de dados, retira-se os pixels ao seu redor, de maneira a reduzir sua área original a uma imagem quadrada de 256x256 pixels. Tal mudança mantém a região central da imagem, que é de maior importância, facilitando o processamento e armazenamento.

Outras transformações como filtragens de ruído e normalização são tratamentos básicos que devem ser realizados, antes da realização de qualquer processamento mais complexo.

Para a filtragem do ruído presente nas imagens adotou-se um filtro espacial com máscara Gaussiana, de equação

$$G(\sigma, x, y) = \sigma^2 \exp\left(-\left(x^2 + y^2\right)/2\sigma^2\right) \qquad (1)$$

O processamento das imagens de treino teve bom desempenho utilizando σ=1.4 e para as imagens de teste, que inclusive têm resolução um pouco menor, foi feita uma filtragem de menor intensidade, com σ=1.2. Para os dois conjuntos de imagens foi usada uma máscara de tamanho 9x9.

IV. Localização e Extração de Atributos

IV.1. Localização da Cabeça

A partir do momento que as imagens estejam preprocessadas, pode-se identificar nas imagens posições estratégicas, como por exemplo os pontos do chamado triângulo canônico, formado pelos extremos externos dos olhos direito e esquerdo mais o ponto subnasal, localizado na base do nariz $\overline{O_d O_e Sn}$.

Tais pontos foram escolhidos como referência por serem localizados em posições razoavelmente rígidas das faces. O que não ocorre, por exemplo, com a boca, que é totalmente flexível, assim como as regiões pertencentes ao maxilar e à testa.

Esses pontos poderiam ser localizados automaticamente através de algum método de reconhecimento de elementos faciais na imagem, porém, assume-se que esses pontos já estejam determinados em cada uma das imagens do banco, já que o objetivo se concentra na classificação dos atributos de movimentação muscular.

Sendo assim, o problema se resume em, a partir dos pontos do triângulo de referência da face, com coordenadas

$$\begin{aligned} O_d &= A = (x_A, y_A, 0) \\ O_e &= B = (x_B, y_B, 0) \\ Sn &= C = (x_C, y_C, 0) \end{aligned} \qquad (2)$$

no plano da imagem $z = 0$, estimar a posição da cabeça no espaço e, assim, localizar pontos de interesse na superfície facial e encontrá-los na imagem analisada.

Assim pode-se, a partir de um modelo da cabeça genérico (figura II) e de parâmetros de calibragem da câmera de vídeo, localizar com precisão razoável pontos em qualquer posição da

face que se queira, assumindo que o triângulo canônico seja rígido em relação à ela. O método foi também usado em COHN et al. (1999).

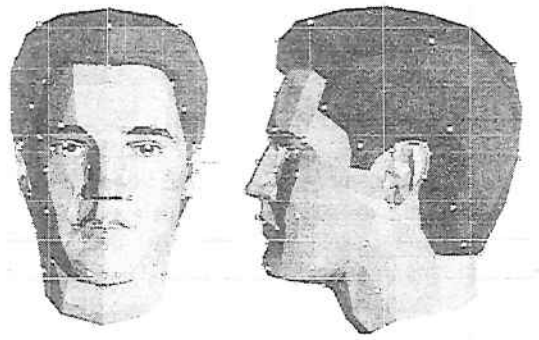

Fig. II: Modelo facial genérico segundo PARRAMON (1994) (*MIRALab Copyright © Information* 1998, Universidade de Geneva) para localização de pontos relativos na face.

Dadas as projeções cônicas, no plano da imagem, dos pontos do triângulo canônico A, B, C e o ponto de foco da câmera ou de projeção cônica

$$O = (0, 0, \lambda), \qquad (3)$$

sendo λ a distância entre O e o plano da imagem, obtém-se a posição real do triângulo no espaço: os pontos A', B' e C', dados pelas equações paramétricas

$$\begin{cases} A' = (0,0,\lambda) + (-x_A, -y_A, \lambda)t \\ B' = (0,0,\lambda) + (-x_B, -y_B, \lambda)u \\ C' = (0,0,\lambda) + (-x_C, -y_C, \lambda)v \end{cases} \qquad (4)$$

tal que

$$|A'-B'| = d_1 \text{ e } |A'-C'| = |B'-C'| = d_2, \qquad (5)$$

sendo as medidas d_1 e d_2 baseadas no modelo da figura II. Tais medidas evidentemente podem variar de pessoa para pessoa, mas assume-se que as proporções sejam as mesmas, portanto não há perda de generalidade. A projeção em perspectiva cônica do triângulo está ilustrada na figura III.

A partir de (4) e (5) obtém-se o sistema não linear

$$\begin{cases} k_1 t^2 - 2k_2 tu + k_3 u^2 = d_1^2 \\ k_1 t^2 - 2k_4 tv + k_5 v^2 = d_2^2 \\ k_3 u^2 - 2k_6 uv + k_5 v^2 = d_2^2 \end{cases} \qquad (6)$$

onde as constantes k têm os seguintes valores:

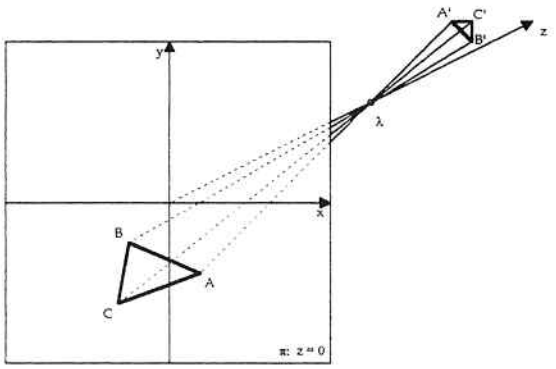

Fig. III: Projeção cônica de A'B'C' no plano da imagem z=0. Resolvendo o problema inverso (inferindo a posição do triângulo no espaço a partir de suas projeções A, B e C).

$$\begin{aligned} k_1 &= x_A^2 + y_A^2 + \lambda^2 \\ k_2 &= x_A x_B + y_A y_B + \lambda^2 \\ k_3 &= x_B^2 + y_B^2 + \lambda^2 \\ k_4 &= x_A x_C + y_A y_C + \lambda^2 \\ k_5 &= x_C^2 + y_C^2 + \lambda^2 \\ k_6 &= x_B x_C + y_B y_C + \lambda^2 \end{aligned} \qquad (7)$$

O sistema é resolvido numericamente e, caso haja mais de uma solução, seleciona-se aquela cuja posição correspondente da cabeça esteja aproximadamente de frente da câmera, ou seja,

$$\angle\left(\overline{A'C'} \times \overline{A'B'}, \overline{\text{ref}}\right) < \theta_{\max} \qquad (8)$$

sendo $\overline{\text{ref}}$ o vetor normal ao triângulo canônico correspondente a uma cabeça considerada de frente para a câmera. Em geral $\overline{\text{ref}} \neq (0,0,1)$, pois quando o sujeito está olhando de frente para a câmera, o triângulo canônico pertence a um plano não paralelo ao da imagem.

Caso, não hajam soluções, é realizada uma espécie de varredura nas vizinhanças dos pontos A, B e C, variando suas coordenadas horizontais e verticais na imagem dentro do limite de erro de ±2 pixels, até que uma solução seja encontrada dentro das condições preestabelecidas.

O processo de localização de pontos específicos na cabeça ou na face se torna simples a partir dessas ferramentas, pois sabendo-se sua posição relativamente ao triângulo canônico, basta projetá-lo no plano da imagem.

IV.2. Extração de Atributos

Segundo KWON; LOBO (1999), pode-se classificar os métodos de análise facial em dois grandes paradigmas. O primeiro é baseado em atributos, tais como posições de olhos, nariz, etc., e relações geométricas entre eles. O segundo, trata uma imagem inteira, ou parcial, de uma face como

um vetor de entrada, e realiza a análise e reconhecimento através de transformações algébricas nele aplicadas.

O método escolhido para realização da análise facial do sistema se encaixa na primeira abordagem. Trata-se de um método simples de se medir a movimentação muscular facial: a análise das intensidades dos pixels pertencentes a segmentos determinados sobre a face na imagem BARTLETT et al. (1996). Com essa ferramenta não há necessidade de se reconstruir a superfície tridimensional da face para a interpretação de seus movimentos, bastando apenas a localização de alguns pontos-chave. Tais segmentos, aqui chamados segmentos de Bartlett (figura III), são definidos em locais onde as movimentações musculares caracterizam as unidades de ação, no caso, da face superior. Além disso, são locais que podem ser analisados de maneira relativamente fixa na face, ou seja, não são afetados pelos movimentos faciais bruscos, como por exemplo a movimentação elástica das sobrancelhas, dos olhos ou da boca.

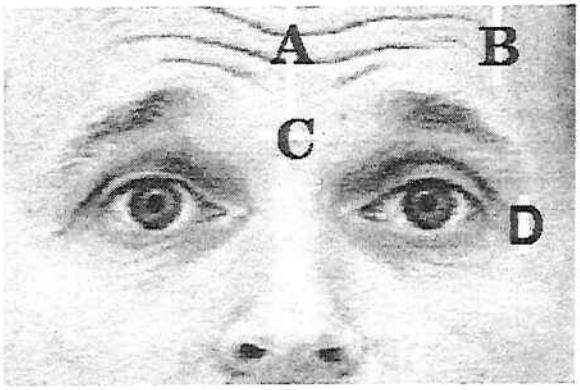

Fig. III: segmentos de Bartlett

A idéia original de Bartlett, para classificação de unidades de ação facial, consiste no seguinte procedimento:

1) A posição dos olhos é determinada manualmente e a região da face superior é selecionada.
2) As imagens passam por uma filtragem de ruído e ajustes de escala e rotação, mantendo-se os olhos na horizontal.
3) O brilho das imagens preprocessadas é normalizado.
4) A partir das coordenadas dos olhos, calcula-se as coordenadas dos segmentos A, B, C e D (figura III).
5) Do conjunto de pontos seqüenciais $((x_1,y_1),...,(x_i,y_i),...,(x_n,y_n))$, que compõem cada segmento discreto na imagem f, obtém-se seus níveis de cinza $(f(x_1,y_1),...,f(x_i, y_i),...,f(x_n, y_n)) = (I_1,...,I_i,...,I_n)$, sendo, assim, calculada a medida de enrugamento

$$P = \sum_{i=2}^{n} (I_i - I_{i-1})^2 \qquad (9)$$

obtendo-se P_A, P_B, P_C e P_D.

6) Mede-se a abertura dos olhos subtraindo-se a área da esclera visível da face expressiva, da área da face neutra.
7) Faz-se a classificação de 6 unidades de ação (AU1, AU2, AU4, AU 5, AU 6, AU 7 — vide apêndice 1) a partir dos 5 atributos extraídos através de um MLP de 3 camadas. A rede, então testada, teve melhor desempenho com 15 unidades na camada interna (figura IV). A unidade de ação classificada corresponde à saída de maior valor da rede (*winner-takes-all*).

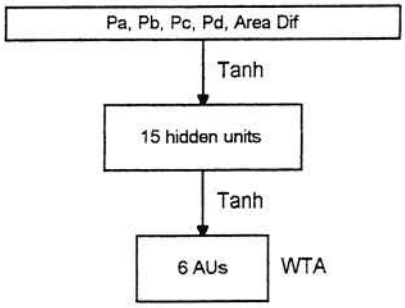

Fig. IV: Estrutura do MLP usado na classificação dos atributos faciais por BARTLETT et al. (1996).

O desempenho desse método, sobre o banco de imagens utilizado por Bartlett, foi de 57% BARTLETT et al. (1996). Tal resultado, revisado em 1999 BARTLETT et al. (1999), aparentemente não teve melhorias e está longe de ser satisfatório.

O procedimento de análise de enrugamento pode ser estendido de forma a aumentar a capacidade do método de retratar as movimentações musculares nas faces. Algumas melhorias podem ser realizadas, como por exemplo:

Linear Time Warping (LTW)

De imagem para imagem a proximidade da face à câmera muda, assim, sempre há diferença de resolução nos detalhes. Conseqüentemente, o tamanho dos segmentos correspondentes em cada imagem serão sempre diferentes. Logo, podem ter mais ou menos pontos.

Para comparar medidas de enrugamento - equação (9) - de segmentos correspondentes, assume-se que eles tenham o mesmo tamanho, pois o enrugamento, como é definido, depende do número de pontos do segmento.

Para resolver este problema, utiliza-se um método de interpolação linear, reduzindo o vetor de níveis de cinza dos segmentos para um vetor de tamanho fixo. Trata-se do chamado *Linear Time Warping*, que corresponde ao seguinte algoritmo:

Deseja-se transformar uma seqüência de valores em forma de um vetor $\mathbf{v}=(v_0,v_2,...,v_m)$ numa seqüência semelhante, porém com resolução $n \neq m$: $\mathbf{u}=(u_0,u_2,...,u_n)$. Para isso, faz-se o seguinte: para cada índice i, de 0 a n-1,

i. Calcula-se $j=i\left(\frac{m-1}{n-1}\right)$

ii. Se j for inteiro, considera-se $u_i = v_j$

iii. Caso contrário, interpola-se:

$u_i=(1-fp)\cdot v_{ip}+fp\cdot v_{ip+1}$, onde ip é a parte inteira de j e fp é a parte fracionária.

Já que o tamanho médio dos segmentos das imagens de teste é de 30 pixels, decidiu-se reduzir o número de pontos de todos eles para este valor.

Modificação na equação (9)

Para que houvesse uma variabilidade mais regular, decidiu-se adotar uma medida de enrugamento baseada em somatório de valores absolutos, ao invés de quadrados:

$$P=\sum_{i=2}^{n}|I_i-I_{i-1}| \qquad (10)$$

Enrugamento relativo

As medidas de enrugamento são afetadas de sujeito para sujeito, pois os de maior idade possuem maior número de rugas, o que não significa que estejam expressando algo na face. Assim, decidiu-se considerar a medida de enrugamento relativa, que corresponde a calcular a diferença, para um mesmo sujeito, entre o valor de enrugamento analisado e o enrugamento mínimo atingido nas imagens amostradas.

Após os devidos ajustes dos detalhes básicos do atributo de enrugamento, pode-se fazer uma extensão desse conceito. A equação (10), não caracteriza a maneira como as rugas estão dispostas espacialmente. Por isso, o atributo adotado, em substituição a este, não será mais um valor escalar, mas envolverá um vetor de enrugamentos locais dado por um conjunto de segmentos de Bartlett numa região de interesse.

As aqui chamadas regiões bidimensionais de Bartlett correspondem aos conjuntos relacionados a um determinado tipo de segmento, formando pequenas grades de segmentos ao redor de cada segmento A, B, C e D, ilustrados na figura III. Assim, obtém-se não mais os segmentos, mas as regiões A, B, C e D.

A análise dessas áreas tem o objetivo de se retirar informações da superfície da pele nessas regiões críticas, ao invés de se classificar apenas informações unidimensionais. Simples segmentos podem não apresentar mudanças relevantes, dependendo da expressão facial. A idade da pessoa pode influenciar, por exemplo, caso ela tenha mais rugas. E, por esse mesmo motivo, a classificação não deve ser feita sobre a região correspondente da imagem, simplesmente. Esse processo deve ser realizado sobre a diferença que ocorreu nas regiões correspondentes nas faces neutra e expressiva. A diferença entre as regiões pode ser quantificada através da própria diferença de brilho entre uma e outra subimagem, ou de uma função conveniente que caracterize a mudança ocorrente entre uma e outra superfície.

V. Utilização das redes neurais artificiais

Para se classificar os atributos obtidos a partir das técnicas acima citadas, notou-se ser conveniente a utilização de uma rede neural artificial, treinando-a com as principais variações de enrugamentos presentes nas imagens de treino (vídeo demonstrativo do material disponível sobre o FACS, contendo imagens de movimentos faciais isolados ou combinações deles). O modelo de rede que mais se adaptou a essa metodologia foi uma rede perceptron multicamada *feedforward*, com *tanh* como função de ativação para cada neurônio. A rede é treinada, através do algoritmo *backpropagation* HERTZ (1991), para reconhecer os padrões das principais variações.

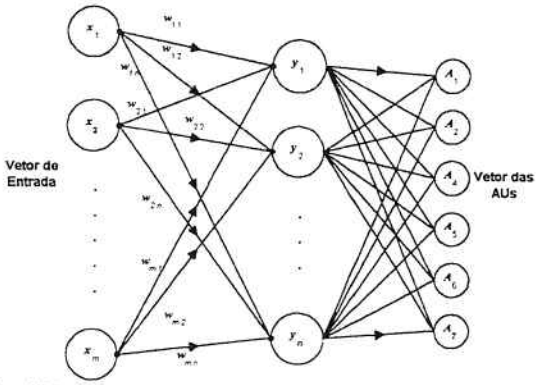

Fig. IV: Esquema do MLP utilizado na classificação dos atributos extraídos das imagens (padrão x) em unidades de ação facial (AUs).

O vetor de entrada da rede neural artificial (RNA) corresponde a um atributo ou padrão extraído da imagem. Suas entradas podem conter, por exemplo, as medidas de enrugamento de uma dada região.

Observou-se que, para se treinar a RNA com maior número de vetores e mais rapidamente, é conveniente uma quantização dos vetores de entrada. Isso pode ajudar, pois os valores de enrugamento têm uma dispersão de erro, que pode ser reduzida, restringindo sua variabilidade a intervalos de

ocorrência. Ou seja, associa-se um valor de enrugamento a uma classe de freqüência.

Analisando o histograma dos valores dos enrugamentos dos segmentos de todas as imagens do banco, notou-se uma certa constância. Assim, optou-se por escolher intervalos das classes igualmente espaçados. Um número de classes cujo resultado foi eficiente foi 7, sendo ímpar, com intuito de associar a classe central ao valor médio dos enrugamentos para um determinado conjunto de segmentos.

Para se construir a tabela de treinamento da RNA, adicionou-se associações, sempre relacionando a imagem de uma seqüência de movimentos no tempo à intensidade correspondente de ação facial existente na face. Além de unidades de ação isoladas, tem-se também combinações delas (vide apêndice 2). Nesse caso, pode ocorrer que uma ação facial evolua mais rapidamente do que outras que compõem uma combinação ao final da seqüência. Além disso, podem ocorrer assimetrias faciais, sendo que um lado da face pode apresentar uma ação facial e outro não, ou haver diferentes intensidades de uma mesma unidade de ação em cada lado da face. Em todas essas situações, procurou-se associar um vetor de saída contendo as intensidades das ações faciais correspondentes.

VI. Resultados

Dentre o total de 400 imagens capturadas em laboratório usadas para o teste da RNA, foram escolhidas 180 imagens. Os critérios usados nessa escolha foram os seguintes. Foram descartadas:

· As imagens com sujeitos com cabelo na testa, total ou parcialmente
· As imagens dos sujeitos que não apresentaram movimentos faciais significativos em nenhuma de suas imagens

Dessas imagens, os atributos são extraídos da mesma forma como para as imagens de treino, e estes são aplicados à RNA, já treinada, obtendo-se a classificação das AUs na saída.

No caso dos segmentos simples (A, B, C e D), as AUs 6 e 7, que envolvem a movimentação dos olhos, não foram classificadas com êxito (o classificador de Bartlett não detectou movimento por parte dessas AUs), o que indica a ineficiência do método de Bartlett em imagens apresentando ruído e de baixa resolução. Assim, considera-se apenas a classificação das AUs 1, 2 e 4, que retratam a movimentação da testa.

83.9 % das imagens testadas eram simplesmente faces neutras (relativamente às ações faciais superiores).

Dentre as faces que apresentavam maior intensidade nas ações faciais (7.2 % do total), obteve-se:

	AU1	AU2	AU4	Combinações
Segmentos de Bartlett	51.7	30.5	x	30.7
Região A	x	x	x	40.2
Região A incluindo "face neutra"	38.5	75.4	x	53.1
Níveis de cinza da região A	x	x	x	18.5

Tab. I: tabela de desempenho de reconhecimento (%) para as imagens de ações faciais mais intensas.

Dentre o total de 180 faces do banco de imagens, tem-se:

	AU1	AU2	AU4	Combinações
Segmentos de Bartlett	82.1	83.9	x	76.6
Região A	x	x	x	56.2
Região A incluindo "face neutra"	81.7	76.6	26.7	x
Níveis de cinza da região A	x	x	x	x

Tab. II: tabela de desempenho de reconhecimento (%) para o total de imagens.

As tabelas de percentuais mostram as taxas de classificação correta das AUs individualmente ou considerando as combinações de ações faciais como um todo. Ou seja, considerando, ao mesmo tempo, a resposta da RNA para as 3 unidades de ação. O "x" mostra situações em que o resultado não foi aceitável (inferior a 20%).

Foram testadas algumas modalidades de distribuição dos segmentos sobre a face. Dentre elas tem-se os próprios segmentos de Bartlett e a chamada região A, que corresponde a um conjunto de segmentos formando a região que varre a testa.

A primeira linha da tabela corresponde ao treinamento da RNA com os segmentos simples da face (A, B, C e D). As linhas posteriores correspondem a treinamentos de RNAs com os seguintes dados: enrugamentos dos segmentos da região A; enrugamentos dos segmentos da região A com um vetor de padrões representando uma face superior neutra (estimada sobre o mínimo dos enrugamentos observados nas imagens); níveis de cinza sem transformações dos segmentos aglomeradas num único vetor de padrões.

Observou-se que o uso da região A não gerou resultados aceitáveis, a não ser no caso em que se considerou a "face superior neutra", porém, o resultado para as combinações de AUs ainda deixa a desejar. Possivelmente, ao se considerar um conjunto de segmentos ao invés de apenas alguns poucos, e

em regiões estratégicas da face, informações espúrias podem estar atrapalhando a classificação.

VII. Conclusões

Sendo que a posição da cabeça nas imagens têm uma variação de posição, notou-se que a localização de regiões ou segmentos na superfície facial, a partir da estimação de sua posição no espaço, foi mais eficiente do que transformar a escala, rotação, translação da imagem da face, e ainda deformá-la para que se encaixasse nos moldes das regiões de interesse, onde se quer medir o enrugamento. Tais transformações só foram utilizadas na fase de classificação pelo especialista, com o intuito de facilitar a visualização.

O método de extração de atributos adotado neste trabalho se mostrou ser simples, rápido de ser processado, e ao mesmo tempo versátil, ou seja, pode ser aplicado a várias resoluções de imagens. Porém, apesar do desempenho não ter sido tão grande quanto era esperado, um teste mais abrangente, com um banco de imagens maior e mais abrangente é uma coisa interessante a ser feita, para que a eficácia da técnica seja realmente posta à prova.

vantagens:

- A técnica independe de resolução e posição da cabeça

VIII. Bibliografia

BARTLETT, M. S.; LARSEN, J.; HAGER, J. C.; EKMAN, P. Classifying Facial Action Advances. In Neural Information Processing Systems 8 MIT Press, Cambridge, MA, p.823-829, 1996.

BARTLETT, M. S.; HAGER, J. C.; EKMAN, P.; SEJNOWSKI, T. J. Measuring Facial Expressions by Computer Image Analysis. Psychophysiology, 36: (2) p.253-263, Mar. 1999.

COHN, J. F.; ZLOCHOWER, A. J.; LIEN, J.; KANADE, T. Automated Face Analysis by Feature Point Tracking Has High Concurrent Validity with Manual FACS Psychophysiology, 36: (1), p. 35-43, Jan. 1999.

Ekman, P. (1972)
Universals and Cultural Differencesin Facial Expression of Emotions
J. Cole (Ed), Nebraska Simposium on Motivation
Lincoln, University of Nebraska Press

Ekman, P.; Friesen, W. V. (1975)
Unmasking the Face
Englewood Cliffs, N. J., Prentice-Hall

Ekman, P.; Friesen, W. V. (1978)
Facial Action Coding System
California, Consulting Psychologists Press

Ekman, P.; Friesen, W. V.; Ancoli, S. (1980)
Facial Signs of Emotional Experience
Journal of Personality and Social Psychology, 6, 1125-1134.

EKMAN, P.; HUANG, T. S.; SEJNOWSKI, T. J.; HAGER J. Final Report to NSF of the Planning Workshop on Facial Expression Understanding. Human Interaction Lab., UCSF, CA 94143, 1993. (Relatório Técnico National Science Foundation)

Engelmann, A. (1986)
LEP- Uma lista de Origem Brasileira, para Medir a Presença de Estados de Ânimo no Momento em que está sendo Respondida
Ciência e Cultura, 38, 121-146

GIL, I. A. Relações entre Ações Faciais e Relatos Verbais de Estados Subjetivos de Emoções e Eventos Correlatos São Paulo, 1993. 183p. Tese (Doutorado) - Insituto de Psicologia, Universidade de São Paulo.

Hertz, J.; Krogh, A.; Palmer, R. G. (1991)
Introduction to the Theory of Neural Computation
Santa Fe Institute

KWON, Y. H.; LOBO, N. V. Age Classification from Facial Images. Computer Vision and Image Understandin, v. 74, n. 1, April, p. 1-21, 1999.

IX. Apêndice 1

Tabela das unidades de ação do FACS
(face superior)

00	Ausência de açào facial observável
01	Levantador de testa interno
02	Levantador de testa externo
04	Abaixador de testa
05	XYZ Levantador de pálpebra superior
06	Levantador de bochecha
07	Apertador de pálpebra
41	Abaixador de pálpebra
42	Abridor de olhos
43	Fechador de olhos
44	Envesgador (*Squint*)
45	Piscador (*Blink*)
46	Piscador (*Wink*)
62	XYZ Virador de olhos para a direita
63	Levantador de olhos

64	Abaixador de olhos
65	*Walleye*
66	Envesgador de olhos (*Crosseye*)
70	Testa não visível

X. Apêndice 2

A demonstração das ações faciais está disponível em videoteipe, sendo feita pelo próprio Paul Ekman. Os códigos FACS, utilizados no trabalho, de cada movimentação contida no vídeo são, na ordem em que se apresentam, os seguintes:

0	
1	10+15+17
2	
5	10+14
7	
6	10+20+25
46	11
4+5	12
5+7	
1+4	6+12
1+2	13
1+2+4	10+12+25
1+2+5	12+16+25
6+43	10+12+16+25
7+43	12+16
9	12+27
10	12+17
...	12+15
9+16+25	6+15
10+16+25	12+15+17
	10+23+25
9+17	12+23
	12+24
10+15	12+17+23
10+17	10+17+23
	10+22+25

BOLETINS TÉCNICOS - TEXTOS PUBLICADOS

BT/PTC/9901 – Avaliação de Ergoespirômetros Segundo a Norma NBR IEC 601-1- MARIA RUTH C. R. LEITE, JOSÉ CARLOS TEIXEIRA DE B. MORAES

BT/PTC/9902 – Sistemas de Criptofonia de Voz com Mapas Caóticos e Redes Neurais Artificiais – MIGUEL ANTONIO FERNANDES SOLER, EUVALDO FERREIRA CABRAL JR.

BT/PTC/9903 – Regulação Sincronizada de Distúrbios Senodais – VAIDYA INÉS CARRILLO SEGURA, PAULO SÉRGIO PEREIRA DA SILVA

BT/PTC/9904 – Desenvolvimento e Implementação de Algoritmo Computacional para Garantir um Determinado Nível de Letalidade Acumulada para Microorganismos Presentes em Alimentos Industrializados – RUBENS GEDRAITE, CLÁUDIO GARCIA

BT/PTC/9905 - Modelo Operacional de Gestão de Qualidade em Laboratórios de Ensaio e Calibração de Equipamentos Eletromédicos – MANUEL ANTONIO TAPIA LÓPEZ, JOSÉ CARLOS TEIXEIRA DE BARROS MORAES

BT/PTC/9906 – Extração de Componentes Principais de Sinais Cerebrais Usando Karhunen – Loève Neural Network – EDUARDO AKIRA KINTO, EUVALDO F. CABRAL JR.

BT/PTC/9907 – Observador Pseudo-Derivativo de Kalman Numa Coluna de Destilação Binária – JOSÉ HERNANDEZ LÓPEZ, JOSÉ JAIME DA CRUZ, CLAUDIO GARCIA

BT/PTC/9908 – Reconhecimento Automático do Locutor com Coeficientes Mel-Cepstrais e Redes Neurais Artificiais – ANDRÉ BORDIN MAGNI, EUVALDO F. CABRAL JÚNIOR

BT/PTC/9909 – Análise de Estabilidade e Síntese de Sistemas Híbridos – DIEGO COLÓN, FELIPE MIGUEL PAIT

BT/PTC/0001 – Alguns Aspectos de Visão Multiescalas e Multiresolução – JOÃO E. KOGLER JR., MARCIO RILLO

BT/PTC/0002 – Placa de Sinalização E1: Sinalização de Linha R2 Digital Sinalização entre Registradores MFC- PHILLIP MARK SEYMOUR BURT, FERNANDA CARDOSO DA SILVA

BT/PTC/0003 – Estudo da Técnica de Comunicação FO-CDMA em Redes de Fibra Óptica de Alta Velocidade – TULIPA PERSO, JOSÉ ROBERTO DE A. AMAZONAS

BT/PTC/0004 – Avaliação de Modelos Matemáticos para Motoneurônios – DANIEL GUSTAVO GOROSO, ANDRÉ FÁBIO KOHN

BT/PTC/0005 – Extração e Avaliação de Atributos do Eletrocardiograma para Classificação de Batimentos Cardíacos – ELDER VIEIRA COSTA, JOSÉ CARLOS T. DE BARROS MORAES

BT/PTC/0006 – Uma Técnica de Imposição de Zeros para Auxílio em Projeto de Sistemas de Controle – PAULO SÉRGIO PIERRI, ROBERTO MOURA SALES

BT/PTC/0007 – A Connected Multireticulated Diagram Viewer – PAULO EDUARDO PILON, EUVALDO F. CABRAL JÚNIOR

BT/PTC/0008 – Some Geometric Properties of the Dynamic Extension Algorithm – PAULO SÉRGIO PEREIRA DA SILVA

BT/PTC/0009 – Comparison of Alternatives for Capacity Increase in Multiple-Rate Dual-Class DS/CDMA Systems – CYRO SACARANO HESI, PAUL ETIENNE JESZENSKY