www.revistageintec.net
ISSN: 2237-0722
revista GEINTEC Gestão, Inovação e Tecnologias

DETECÇÃO DE EXPRESSÕES FACIAIS: UMA ABORDAGEM BASEADA EM ANÁLISE DO FLUXO ÓPTICO

FACIAL EXPRESSION DETECTION: A TECHNIQUE FOR OPTICAL FLOW ANALYSIS

Leonardo Panta Leão[1]; Jonas Santos Bezerra[2]; Leonardo Nogueira Matos[3]; Maria Augusta Silveira Netto Nunes[4]

[1]Universidade Federal de Sergipe – UFS – São Cristóvão/SE – Brasil
leonardopspl@gmail.com,

[2]Universidade Federal de Sergipe – UFS – São Cristóvão/SE – Brasil
jonassantosbezerra@gmail.com

[3]Universidade Federal de Sergipe – UFS – São Cristóvão/SE – Brasil
lnmatos@ufs.br

[4]Universidade Federal de Sergipe – UFS – São Cristóvão/SE – Brasil
gutanunes@gmail.com

Resumo

Este artigo descreve um sistema em tempo real de reconhecimento de expressões faciais usando uma técnica para detecção automática de expressões faciais baseada no algoritmo Haartraning seguido pela colocação automática dos pontos responsáveis por fazer a leitura dos movimentos faciais, utilizando uma técnica de análise do fluxo óptico nestes pontos. Os movimentos faciais lidos são interpretados como action units, de acordo com a classificação proposta por Ekman. Como resultado foi gerado um sistema de detecção automática em tempo real de expressões faciais capaz de selecionar uma dentre seis emoções (raiva, medo, felicidade, tristeza, aversão e surpresa).

Palavras-chave: Processamento de imagens, Expressões Faciais, Emoção, Fluxo Óptico, Action Units.

Abstract

This article describes a real-time automatic facial expression recognition system using a technique for automatic facial expression detection based on Haartraning algorithm followed by automatic points placement that performs facial movements' reading using a technique for optical flow analysis on these points. The facial movements read are interpreted as action units, according to a classification presented by the researcher Ekman. As a result, an automatic real-time facial expression and correlated emotion detection system capable of selecting one among six expressions (anger, fear, happiness, sadness, disgust and surprise) was generated.

Key-words: Image processing, Facial Expression, Emotion, Optical Flow, Action Units.

1. Introdução

Há um interesse recente na melhoria da interação entre humanos e computadores. Para que uma efetiva interface humano-computador inteligente seja alcançada é necessário que o computador interaja naturalmente com o usuário, semelhante à maneira que os humanos interagem. Uma das formas que os pesquisadores encontraram para promover esse tipo de interfaceamento é através da Computação Afetiva.

A Computação Afetiva estuda como os computadores podem reconhecer, modelar e expressar as emoções (e outros aspectos psicológicos humanos) e como podem responder às mesmas (PICARD, 1997), contribuindo para o aumento da coerência, consistência, predicabilidade e credibilidade das reações e respostas computacionais providas durante a interação humana via interface humano-computador (NUNES, 2009).

Reeves & Nass (1998) mostraram que as pessoas tratam computadores, celulares e outras mídias eletrônicas como se estes fossem, também, humanos detentores de personalidade, emoções e até mesmo vontade próprias.

Entretanto, fazer com que uma máquina reconheça, modele e expresse emoções não é uma tarefa simples. Quando seres humanos interagem entre si, boa parte dessa interação é baseada na linguagem verbal e na utilização da linguagem corporal por meio de gestos e expressões faciais que carregam e transmitem as emoções dos interlocutores. A percepção das características fisiológicas, visuais e vocais que denotam as emoções é natural e inerente aos seres humanos, ocorrendo de forma quase imperceptível. Como então fazer com que o computador adquira tais habilidades?

Dentre os vários meios de que os seres humanos dispõem para demonstrar/perceber emoções, um dos mais importantes é através das expressões faciais, de forma que uma ferramenta de reconhecimento do estado emocional humano através da análise facial pode ser bastante valiosa nesta busca por uma interface humano-computador mais inteligente, gerando interações mais adaptáveis e personalizadas às demandas dos usuários.

Este trabalho descreve um sistema de detecção automática, em tempo real, que utiliza como entrada um vídeo ou as imagens capturadas por uma *webcam*. Inicialmente ocorre a detecção da face do indivíduo na tomada de vídeo e a marcação de pontos, denominados pontos de leitura, em regiões de interesse (olhos e boca, por exemplo). E baseado nos movimentos fa[1]ciais, ocorre a classificação da expressão facial e, consequentemente, da emoção apresentada pelo usuário.

Este artigo é organizado da seguinte forma: Na seção 2 é apresentada a fundamentação teórica, criando um apanhado histórico dos estudos que relacionam as expressões faciais humanas

1 . Esse artigo é uma extensão do artigo publicado em Leao, L. P. ; MATOS, L. N.;NUNES, M. A. S. N. Detecção de expressões faciais: uma abordagem baseada em análise do fluxo óptico. In: WTICG, 2011, Salvador.

com as emoções, bem como um levantamento de trabalhos relacionados e uma breve introdução à técnica deste trabalho. Na seção 3 é apresentada a metodologia utilizada, descrevendo-se com mais detalhes a técnica empregada, a implementação e os experimentos. Na seção 4 são apresentadas as conclusões e, por fim, segue-se as referências bibliográficas.

2. Fundamentação Teórica

Desde o início da Computação Afetiva, os pesquisadores tem buscado formas de permitir que o computador seja capaz de reconhecer e responder as emoções humanas. Encontrando-se hoje uma variedade de grupos de pesquisa na área, muitos deles com trabalhos específicos voltados ao reconhecimento das emoções, como MIT *Affective Computing Research Group* (2011), *Humaine* (2011) (*Human-Machine Interaction Network on Emotion*), GERG (2011) (*Geneva Emotion Research Group*), ERMIS (*Effective Reproducible Model of Innovation System*) e até mesmo grandes empresas como a IBM (BlueEyes, 2011) tem dado atenção a área.

Esse tipo de pesquisa tem base em trabalhos muito anteriores ao próprio advento do computador. Historicamente, o primeiro pesquisador a trabalhar com as expressões e emoções humanas foi o francês Duchenne de Boulogne. Duchenne (1862) utilizou choques elétricos objetivando causar contrações musculares para assim determinar como o rosto humano produzia as expressões faciais, as quais, acreditava ele, estavam ligadas à "alma" humana.

Em 1872, Darwin estudou a relação entre as expressões faciais e o estado emocional e afirmou que tanto os jovens quanto os idosos, sejam eles homens ou animais, expressam os mesmos estados mentais e emoções por meio dos mesmos movimentos e expressões faciais (DARWIN, 1872).

Já em 1978, o psicólogo Ekman demonstrou evidências sobre essa "universalidade" da manifestação de emoções através de expressões faciais, não só entre pessoas de diferentes idades como também de diferentes etnias - com diferentes tipos faciais - e diferentes culturas. Ekman elencou também um conjunto base de emoções, a saber: felicidade, surpresa, raiva, medo, tristeza e aversão (EKMAN & FRIESEN, 1978).

Ekman desenvolveu ainda um sistema de codificação para as expressões faciais, FACS - *Facial Action Code System* - onde os movimentos faciais são descritos como um conjunto de *action units* (AU), um conjunto de músculos faciais que geram uma ação quando estimulados (EKMAN & FRIESEN, 1982). Cada AU tem como base estudos relacionados à anatomia dos músculos faciais. A grande maioria dos pesquisadores de visão computacional inspira-se nessas AU fazendo o uso de

processamento de imagem e vídeo para automaticamente rastrear características faciais e então utilizá-las para categorizar as diferentes expressões.

Na literatura, encontra-se diversos trabalhos sobre o reconhecimento das emoções utilizando-se expressões faciais, alguns destes são unimodais, ou seja, utilizam apenas o reconhecimento de expressões faciais como base, outros procuram utilizar mais de um tipo de entrada, como por exemplo processamento de voz, movimentos corporais, etc.

Quanto ao reconhecimento das emoções através de análise de expressões faciais Fraponogos & Taylor (2005) dizem que há basicamente duas abordagens para o mapeamento das características faciais para as emoções. Na primeira, a análise da expressão facial é feita de maneira estatística em relação às extremidades de uma expressão, para assim encontrar pistas como rugas, posições e formas no rosto que ajudem a inferir qual o estado emocional de uma pessoa, entretanto, segundo os autores, poucos trabalhos baseados nessa técnica obtiveram sucesso, devido principalmente ao grande custo computacional que apresenta. Na segunda, mais utilizada e difundida na literatura, a análise é orientada a gestos, sendo necessária a extração de vários *frames* sucessivos da expressão facial dos quais são extraídos os gradientes e variações entre *frames* e destes são mapeados os estados emocionais.

Busso *et al.* (2004) conceberam um sistema de reconhecimento das emoções baseado em tanto na análise da expressão facial quanto no reconhecimento de timbres de voz, seu trabalho apresenta algumas das limitações a qual estas duas abordagens estão sujeitas quando utilizadas separadamente e demonstra a complementaridade das duas modalidades.

Ioannou *et al.* (2005) criaram um sistema de análise facial que extrai parâmetros de animação facial (FAP) a partir dos quais é criada uma rede neural que utiliza lógica fuzzy com regras baseadas na análise dos FAPs e permite que o computador aprenda e se adapte as características das expressões faciais específicas de cada usuário.

Hammal *et al*. (2005) construíram um sistema de classificação de expressões faciais baseado na Teoria da Crença - *Belief Theory* – podendo quantificar o nível de confiabilidade do reconhecimento de uma expressão e a intensidade/valência com que o usuário demonstra cada expressão.

Ioannou *et al*. (2007) apresentam um sistema de reconhecimento de expressões faciais baseado em uma fusão de diferentes técnicas, baseando-se no princípio de que não se pode controlar o ambiente humano em termos de luz ou qualidade de cor, bem como a expressividade e movimentos humanos, permitindo assim que o sistema consiga fazer o reconhecimento das expressões faciais e das emoções mesmo em ambientes adversos ou pouco usuais.

Castellano *et al*. (2008) utilizaram uma abordagem multimodal para reconhecimento das emoções unindo expressões faciais, fala, gestos e movimentos corporais. Utilizando-se de classificadores Bayesianos para cada modalidade e depois unindo os dados no nível de tomada de decisão os autores conseguiram aumentar em 10% o poder de reconhecimento geral das emoções.

Pantic & Rothkrantz (2000) e Tao & Tan (2005) fornecem um apanhado de abordagens diferentes utilizadas na tentativa de fazer o reconhecimento automático de expressões faciais: Discriminante Linear de Fisher (FDA), Redes Neurais, Modelos Ocultos de Markov, Análise de Componentes Principais e de Componentes Independentes (PCA e ICA, respectivamente), Modelos de Distribuição de Pontos (PDM), Fluxo Optico, Coeficientes DCT, etc. Essas abordagens têm algo similar quanto ao fato de que a maioria delas utilizam o rastreamento de características faciais utilizando algum modelo de movimento da imagem (fluxo óptico, coeficientes DC, etc). Baseado nessas características um classificador pode ser treinado. A principal diferença entre essas abordagens fica no conjunto de características extraídos das imagens de vídeo e do classificador utilizado.

3. Metodologia

O trabalho de (AZCARATE *ET AL*., 2005), assim como nosso trabalho, utiliza o detector de faces proposto por (VIOLA & JONES, 2001) para a detecção inicial da localização da face. Nesse trabalho é construída na face uma máscara que é deformada com os movimentos faciais que são realizados quando o usuário muda a expressão facial. Essa máscara é baseada na máscara que utiliza deformação dos volumes de Bézier proposta por (TAO & HUANG, 1998). Esse trabalho faz o uso de treinamento de classificadores baseado nos dados obtidos com a deformação dos volumes de Bézier. Os classificadores utilizados são o Naive Bayes e o *Tree-Augmented-Naive Bayes* (TAN).

Tabela 1. Action units utilizadas no detector de expressões faciais implementado

AU	Descrição
1+2	levantamento arqueado das sobrancelhas
1+2+4+5	levantar e aproximar as sobrancelhas
4	abaixar as sobrancelhas
5	levantamento da parte superior dos olhos
6	apertar os olhos
12+25	boca no formato de sorriso
18	espremer os lábios
20	esticar a boca horizontalmente
26	queda do queixo

Nosso trabalho não faz uso de treinamento de classificadores para a categorização das expressões faciais. Utilizamos as AU detectadas através da leitura do movimento de pontos colocados automaticamente na face do usuário utilizando inicialmente um detector de faces baseado no algoritmo Haartraining (VIOLA & JONES, 2001) e, em seguida, utilizando a técnica de fluxo óptico através do método Lucas-Kanade (LUCAS & KANADE, 1981). A categorização das expressões faciais ocorre com a aparição de conjuntos de AU, que caracterizam uma dentre as seis expressões faciais além da neutra, presentes simultaneamente num dado instante. A Tabela 1 mostra as AU utilizadas em nosso trabalho. Essas AU são listadas em (EKMAN & FRIESEN, 1978), a primeira linha da Tabela 1, por exemplo, mostra o resultado de duas action units distintas que podem aparecer ao mesmo tempo (AU 1+2), a AU 1 é o levantamento da parte interna da sobrancelha enquanto que a AU 2 é o levantamento da parte externa da sobrancelha, resultando no levantamento arqueado das sobrancelhas.

3.1. Detecção da Face

O método de segmentação proposto por Viola & Jones (2001) e discutido em Lienhart & Maydt (2002) é uma técnica de detecção baseada na aparência do objeto. Métodos desse tipo aprendem características a partir de conjuntos de imagens de treinamento que capturam a variedade da classe do objeto.

Como sugerido em Viola *et al*. (2003) esse método pode ser usado para detectar faces humanas. Nesse método é feito uma etapa inicial de treinamento onde ocorre a seleção de dois conjuntos de imagens, um positivo e outro negativo. O conjunto de imagens positivas contêm recortes onde está contido apenas o objeto a ser detectado. Já o conjunto de imagens negativas são imagens que não contêm o objeto, geralmente paisagens onde o objeto pode ser encontrado. Em seguida é feito um conjunto de exemplos onde as imagens positivas são combinadas com as negativas para formar imagens de provável aparição do objeto.

O processo de extração de características é baseado em máscaras convolucionais inspiradas em funções de Haar Gonzalez & Woods (2002) (Figura 1). Essas características são calculadas através da convolução na imagem do objeto baseadas numa decisão binária a partir de um *threshold*.

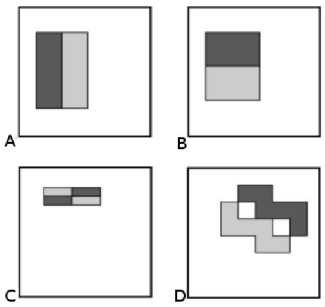

Figura 1. Exemplos de características retangulares
Fonte: Viola & Jones (2001)

Algumas características são mostradas na Figura 1, em (A) e (B), são características que capturam transições de nível de cinza verticais e horizontais respectivamente. Em (C) e (D) são mostradas máscaras que capturam características diagonais.

Na etapa de treinamento o método utiliza o AdaBoost (FREUND & SCHAPIRE, 1999) para construir o classificador. O objetivo desse algoritmo é construir um classificador eficiente a partir de uma série de classificadores fracos baseados em decisões binárias a partir da convolução dessas máscaras na imagem do objeto. O uso de características reduz a variedade de dados na classe e aumenta a variedade de dados fora da classe em comparação a dados de entrada crús, isto é, somente a informação contida nos *pixels* da imagem. Além disso, as características geralmente codificam conhecimento sobre o domínio, o que é difícil aprender a partir de um conjunto crú e finito de dados. O conjunto de características usados pelo algoritmo são características como as descritas acima em todas as escalas e enquadramentos possíveis. Para se ter uma ideia, numa imagem de 64 por 64 *pixels* temos um conjunto de mais de 120 mil características possíveis.

Para o cômputo rápido dessas características é utilizado uma imagem chamada imagem integral, descrita em detalhes no trabalho de Lienhart e Maydt (LIENHART & MAYDT, 2002). Essa imagem é construída utilizando o princípio da programação dinâmica, onde cada pixel (i, j) da imagem integral guarda a soma dos valores de todos os *pixels* do canto superior esquerdo (a origem) até o pixel corrente, isto é , o pixel (*i, j*). Usando a imagem integral é possível realizar o cálculo de uma característica em tempo constante.

O AdaBoost seleciona as características que melhor classificam os objetos e chama um classificador fraco repetidamente numa série de turnos. Para cada chamada, uma distribuição de pesos é atualizada para indicar a importância de alguns exemplos no conjunto de dados para a classificação. A cada turno, os pesos de cada exemplo incorretamente classificado são incrementados, de maneira que o novo classificador tenha um maior foco nesses exemplos. Assim, após selecionar um classificador ótimo, baseado nessas características selecionadas e nessa

distribuição de pesos, os exemplos que o classificador classifica incorretamente tem seus pesos aumentados e os exemplos classificados corretamente tem seus pesos diminuídos. Consequentemente, quando o algoritmo testa uma nova distribuição de pesos ele irá selecionar um classificador que melhor identifica esses exemplos que o classificador anterior errava. Cada classificador forte obtido a partir desses conjuntos de características é posto numa estrutura de cascata de rejeição (Figura 2) na ordem do menos complexo (menos características usadas) para o mais complexo (mais características usadas).

Na etapa de detecção utilizamos a arquitetura de cascata de rejeição. A Figura 2 ilustra a cascata de rejeição. A entrada é passada pelo primeiro classificador que decide entre verdadeiro ou falso (objeto encontrado ou não encontrado). Uma determinação de falso interrompe computação posterior e faz com que o detector retorne falso. Uma determinação verdadeira passa a entrada para o próximo classificador na cascata. Se todos os classificadores votarem em verdadeiro então a entrada é classificada como um exemplo verdadeiro. Dessa maneira, podemos economizar vários ciclos computacionais já que um dado de entrada, quando falso, requer a utilização de somente alguns nós da cascata. A detecção ocorre numa janela deslizante desde o canto superior esquerdo da imagem até o canto inferior direito. A cada término de deslizamento na imagem a janela de detecção é aumentada 20% do seu tamanho inicial até que a janela enquadre ou exceda o tamanho da imagem. Caso mais de uma face seja encontrada praticamente na mesma região, já que poderá ser encontrada tanto na janela de detecção menor quanto na próxima aumentada, essas detecções são mescladas e consideradas uma só.

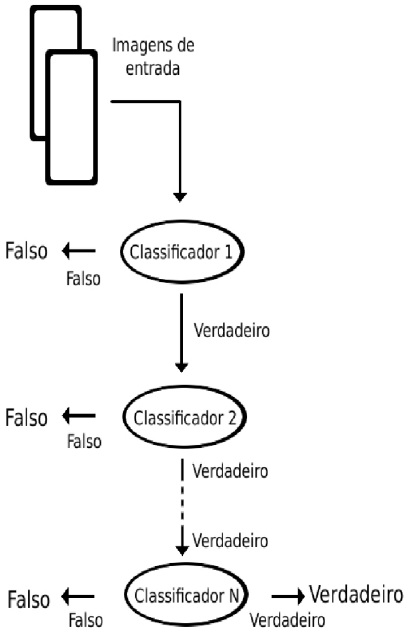

Figura 2. Modelo em cascata do algortimo

Fonte: Viola & Jones (2001)

3.2. Colocação dos Pontos de Leitura

A colocação automática dos pontos de leitura, isto é, pontos de marcação de estruturas de controle como olhos, boca e nariz, ocorre logo após a etapa inicial de detecção automática da face e das regiões de interesse na face. De fato, ela só ocorre após a segmentação dessas regiões de interesse. A técnica utilizada para a segmentação dessas regiões é a mesma utilizada para a detecção da face utilizando a face detectada como região de interesse.

Para reduzir o custo computacional e a possibilidade de falhas utilizamos como região de interesse para a detecção da boca apenas a metade inferior da face detectada, já pra as regiões dos olhos esquerdo e direito utilizamos a metade de cada lado respectivamente. A região entre as sobrancelhas é encontrada imediatamente após o término da região do olho esquerdo e tem o mesmo comprimento e metade da altura. A região da testa é encontrada imediatamente acima da região entre as sobrancelhas e tem a mesma dimensão. A região abaixo da boca é encontrada logo após a colocação do ponto da parte inferior da boca e tem o mesmo comprimento da região do olho esquerdo e 25% da altura da mesma região. Os pontos são colocados nas seguintes regiões:

- Cantos esquerdo, direito, superior e inferior da boca totalizando quatro pontos sendo colocados na região de interesse correspondente é boca.

- Cantos superior e inferior do olho e um ponto acima da sobrancelha (tanto no olho esquerdo, como no direito) totalizando três pontos sendo colocados em cada uma dessas regiões de interesse.
- Região entre as sobrancelhas.
- Região da testa.
- Região abaixo da boca.

A Figura 3 mostra a tela inicial do programa após a etapa inicial de detecção e segmentação dessas regiões de interesse.

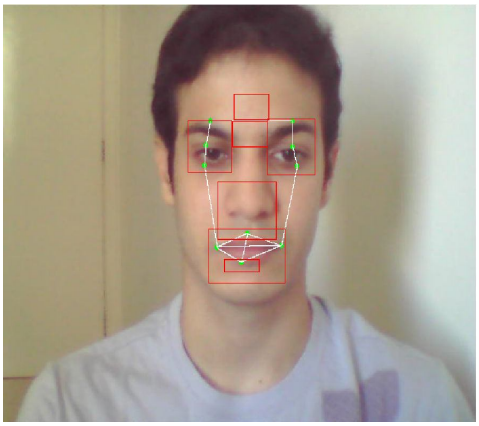

Figura 3. Regiões de interesse apontadas pelo sistema implementado na face inicialmente detectada

Na região da boca utilizamos o filtro de Sobel, que calcula a magnitude do gradiente da intensidade luminosa em cada pixel da imagem. Assim, obtemos uma noção de como varia espacialmente a luminância em cada ponto, se de modo mais suave ou abruptamente. Com isso, conseguimos estimar a presença de uma transição abrupta de nível de cinza. Filtros de Sobel direcionados permitem identificar a direção dessas variações (horizontal ou vertical) (GONZALEZ & WOODS, 2002). Como as variações intensas correspondem a fronteiras bem definidas entre objetos, conseguimos fazer a detecção de contornos.

Após a utilização do filtro configurado para detectar transições verticais (já que essa é a orientação perpendicular da boca) fazemos uma varredura da esquerda para a direita até que o primeiro canto seja encontrado para colocarmos o ponto esquerdo e uma varredura da direita para esquerda da mesma maneira para colocarmos o ponto direito após a binarização da imagem. A binarização acontece tornando branco *pixels* com tonalidade abaixo de um dado limiar e preto caso contrário. Assim, ficam evidentes as transições abruptas do nível de cinza. A mesma ideia é feita para encontrarmos o canto inferior e superior da boca utilizando o centro da boca como região inicial. A Figura 4 mostra a sequência da obtenção da posição inicial dos pontos na boca.

Figura 4. Na parte esquerda temos uma imagem em escala de tons de cinza que é utilizada para a convolução do filtro de Sobel, no centro temos a imagem após a convolução e binarização da imagem (para que fique evidenciado apenas as transições na horizontal) e na direita temos a colocação o inicial dos pontos nos cantos encontrados.

Para os olhos utilizamos o mesmo filtro e fazemos o mesmo tipo de varredura para colocarmos os pontos inferior e superior. Para o canto superior fazemos a varredura no sentido da esquerda para a direita e de cima para baixo partindo da metade da altura da região de interesse do olho, para o canto inferior fazemos o mesmo procedimento, porém partindo de baixo para cima. Para a sobrancelha fazemos a varredura de cima para baixo, partindo do meio até que a primeira transição seja encontrada.

Nas regiões de interesse da área entre as sobrancelhas, da testa e da área abaixo da boca, colocamos o ponto no centro da região de interesse. Esses pontos são utilizados para que seja possível rastrear essas regiões constantemente a um baixo custo computacional. No próximo tópico é mostrada a técnica utilizada para a leitura contínua da posição atual desses pontos. A Figura 3 mostra os pontos colocados.

3.3. Fluxo Óptico nos Pontos

Para o cálculo contínuo da posição atual dos pontos, foi utilizado o algoritmo Lucas-Kanade (LK) (LUCAS & KANADE, 1981) (BOUGUET, 2000). Esse método diferencial para o cálculo do fluxo óptico supõe que o fluxo é essencialmente constante numa vizinhança local do pixel em consideração. Assim, as equações básicas de fluxo óptico para todos os *pixels* naquela vizinhança são resolvidas utilizando o método dos mínimos quadrados (LUCAS, 1984).

O método LK supõe que o deslocamento do conteúdo da imagem entre dois instantes próximos é pequeno e aproximadamente constante com uma vizinhança do ponto p em consideração. Assim, podemos supor que a equação do fluxo óptico é mantida para todos os *pixels*

com uma janela centrada em p. Isto é, o vetor (V*x* , V*y*) deve satisfazer

$$I_x(q_1)V_x + I_y(q_1)V_y = -I_t(q_1)$$
$$I_x(q_2)V_x + I_y(q_2)V_y = -I_t(q_2)$$
$$\vdots$$
$$I_x(q_n)V_x + I_y(q_n)V_y = -I_t(q_n)$$

Onde q_1 , q_2 , ..., q_n são pixels dentro da janela, e I_x (q_i), I_y (q_i), I_t (q_i) são as derivadas parciais da imagem I em relação à posição x, y e o tempo t, calculados no ponto q_i e no tempo corrente. Essas equações podem ser escritas na forma da matriz A*v* = b, onde

$$A = \begin{bmatrix} I_x(q_1) & I_y(q_1) \\ I_x(q_2) & I_y(q_2) \\ \vdots & \vdots \\ I_x(q_n) & I_y(q_n) \end{bmatrix}, v = \begin{bmatrix} V_x \\ V_y \end{bmatrix}, \text{e } b = \begin{bmatrix} -I_t(q_1) \\ -I_t(q_2) \\ \vdots \\ -I_t(q_n) \end{bmatrix}$$

Esse sistema tem mais equações que incógnitas, assim ele geralmente é indeterminado. O método LK usa uma solução ajustada utilizando o princípio dos mínimos quadrados. Isto é, ele resolve o sistema 2 × 2

$$A^T A v = A^T b \text{ ou}$$

$$v = \frac{A^T b}{A^T A}$$

Onde A^T é a matriz transposta, ou seja, ela computa

$$\begin{bmatrix} V_x \\ Vy \end{bmatrix} = \begin{bmatrix} \sum_i I_x(q_i)^2 & \sum_i I_x(q_i) I_y(q_i) \\ \sum_i I_x(q_i) I_y(q_i) & \sum_i I_x(q_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i I_x(q_i) I_t(q_i) \\ -\sum_i I_y(q_i) I_t(q_i) \end{bmatrix}$$

A matriz $A^T A$ é chamada de tensor de estrutura da imagem no ponto *p*.

3.3.1. Janela Ponderada

A solução plana dos mínimos quadrados acima dá a mesma importância para todos os n *pixels* q_i na janela. Na prática é geralmente melhor dar mais peso para os *pixels* que são próximos do pixel central *p*. Para isso, utilizamos uma versão ponderada da equação dos mínimos quadrados,

$A^T W A v = A^T W b$ *ou*

$$v = \frac{A^T W b}{A^T W A}$$

onde W é uma matriz diagonal $n \times n$ contendo os pesos $W_{ii} = w_i$ a serem assinalados na equação do pixel q_i. Isto é, computamos

$$\begin{bmatrix} V_x \\ Vy \end{bmatrix} = \begin{bmatrix} \sum_i w_i I_x(q_i)^2 & \sum_i w_i I_x(q_i) I_y(q_i) \\ \sum_i w_i I_x(q_i) I_y(q_i) & \sum_i w_i I_x(q_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i w_i I_x(q_i) I_t(q_i) \\ -\sum_i w_i I_y(q_i) I_t(q_i) \end{bmatrix}$$

O peso w_i é geralmente ajustado como sendo uma função Gaussiana da distância entre q_i e p.

3.3.2. Método LK Piramidal

Para o cálculo do fluxo óptico dos pontos na imagem foi utilizada a versão "piramidal" do método LK, como descrito em Bradski & Kaehler (2008). A técnica consiste em criar uma pirâmide Gaussiana da imagem onde no topo da pirâmide temos a menor versão da imagem (com menos detalhes) até chegarmos à base onde atingimos os *pixels* crús da imagem (maior riqueza de detalhes). O fluxo óptico é calculado para os pontos no topo da pirâmide e os resultados do cálculo desse nível servem como ponto de partida para os cálculos do nível seguinte. Utilizar o método LK dessa forma nos permite que movimentos maiores sejam capturados por janelas locais, o que torna o método mais eficiente e robusto, pois janelas maiores ferem a suposição inicial do método de que o pixel rastreado e seus vizinhos pertecem à mesma superfície fechada (BRADSKI & KAEHLER, 2008). O método é ilustrado na Figura 5.

3.4 Implementação

Para a implementação do sistema proposto utilizamos a linguagem C++ com a biblioteca livre de visão computacional OpenCV (INTEL, 2001). Essa biblioteca possui diversos algoritmos e rotinas de processamento de imagem e visão computacional. Em nosso trabalho fizemos o uso de alguns desses algoritmos, assim como a captura de vídeo através da webcam que também é suportada por essa biblioteca.

3.4.1. Visualização

Para a utilização da implementação basta que o usuário execute a aplicação. A captura das imagens através da *webcam* ou do vídeo é iniciada e nessa etapa o algoritmo detector automático de expressões faciais, descrito na Seção 3.1, é utilizado até que a face do usuário seja detectada. Após isso, o mesmo algoritmo é executado na região da face encontrada para segmentar as regiões dos olhos esquerdo e direito, boca e nariz, como descrito na Seção 3.2. Ao término desta etapa esse algoritmo deixa de ser executado. Na próxima etapa o algoritmo de fluxo óptico, descrito na Seção 3.3, entra em execução para realizar a leitura contínua da movimentação dos pontos colocados nas diferentes regiões da face.

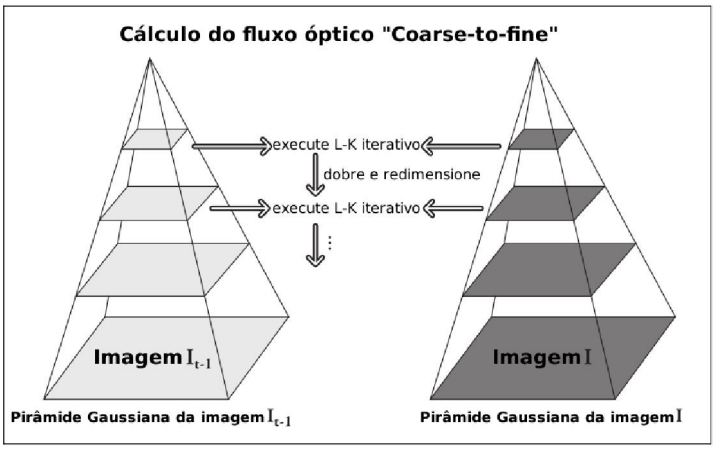

Figura 5. Fluxo óptico piramidal de Lucas-Kanade
Fonte: Bradski & Kaehler (2008)

As mudanças na movimentação dos pontos em relação à distância inicial entre esses pontos na face inicial, que é interpretada como neutra, caracteriza a presença de algumas AU, além da presença ou ausência de rugas, que são detectadas utilizando filtro de Sobel, nas regiões entre as sobrancelhas, testa e abaixo da boca. A combinação de algumas AU representam a aparição de uma emoão na face. A Figura 6 mostra o sistema sendo utilizado.

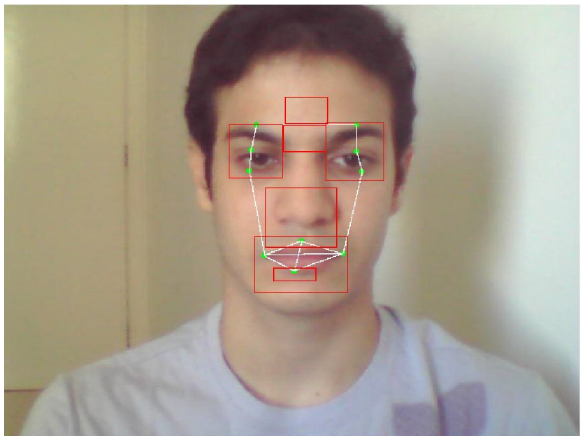

Figura 6. Exemplos da aplicação em execução

3.5. Experimentos

A etapa de experimentos consistiu na coleta de dados a partir do uso da aplicação por três usuários distintos. Nenhum deles é autor ou tinha conhecimento prévio do sistema. As expressões foram feitas de forma não espontânea, cinco vezes cada expressão facial para cada um dos três usuários e de forma consecutiva iniciando com a expressão facial neutra. A Tabela 2 mostra a matriz de confusão para as seis expressões faciais. Podemos ver na tabela a taxa de acerto (expressão facial detectada corretamente) em negrito em cada uma das linhas onde coincidem as expressões intencionadas pelos usuários com as colunas correspondentes às expressões detectadas pelo sistema.

Tabela 2. Matriz de confusão para as seis expressões faciais: felicidade (F), medo (M), surpresa (S), tristeza (T), raiva (R) e aversão (A). As linhas representam a expressão intencionada pelo usuário, já as colunas representam a expressão detectada pelo sistema. A precisão média foi de 68,89%

	F	M	S	T	R	A
F	73,3%	6,7%	0%	0%	0%	20%
M	0%	80%	6,7%	0%	0%	13%
S	0%	40%	60%	0%	0%	0%
T	0%	0%	0%	67%	0%	33%
R	6,7%	0%	0%	0%	53,3%	40%
A	0%	20%	0%	0%	0%	80%

A taxa de atualização dos quadros de vídeo após a detecção da face foi próxima de 30 quadros por segundo. O computador utilizado nos experimentos é constituído de um processador de 2 núcleos de 2.2 GHz e 2 GB de memória RAM. A *webcam* utilizada possui 2 MB *pixels* de resolução.

4. Conclusão

Apresentamos neste trabalho mais uma abordagem para a problemática da detecção automática de emoções através de expressões faciais baseada em algoritmos conhecidos de processamento de imagem e visão computacional. Essa abordagem tem como principal vantagem a não intervenção do usuário na utilização da aplicação. Nos experimentos foram obtidos uma taxa de atualização dos quadros de vídeo compatível com uma aplicação de tempo real. Porém, existem ainda algumas limitações que devem ser levadas em consideração nessa abordagem.

As expressões faciais são detectadas apenas com o indivíduo de frente para a câmera e sem objetos ou características muito distintas na face (óculos escuros, cabelo cobrindo os olhos, por exemplo), já que a técnica detectora de faces tem sua cascata de rejeição treinada para detectar apenas a face em posição frontal (faces de perfil ou parcialmente oclusas não são detectadas) sem peculiaridades na aparência. A luminosidade do ambiente deve ser constante para não ferir a suposição de constância do brilho da técnica proposta por Lucas-Kanade (BRADSKI & KAEHLER, 2008).

Para avaliar melhor a eficiência da detecção das expressões faciais, pretendemos realizar testes com uma base de vídeos extensa e com uma quantidade maior de usuários. Como trabalhos futuros pretende-se, além de superar as limitações dessa abordagem, fazer com que a aplicação exporte os dados sobre as emoções do usuário no formato da EmotionML (W3C, 2008), a linguagem de marcação das emoções padrão do W3C e futura integração com a PersonalityML (NUNES *ET* AL, 2012), o que permitirá integração da aplicação a outras que possam ser tanto aplicações que façam o reconhecimento automático de emoções e da personalidade através de outras modalidades, quanto aplicações que utilizem esses aspectos psicológicos no processo de tomada de decisão computacional, como por exemplo adaptação de interação, personalização de interfaces, recomendações em *e-commerce, e-learning, e-training,* dentre outros (NUNES *ET AL, 2009*).

Referências

AZCARATE, A.; HAGELOH, F.; SANDE, K.V.D. & VALENTI, R., **Automatic facial emotion recognition.** Universiteit van Amsterdam, 2005.

BLUEEYES, IBM RESEARCH.

Available at http://www.almaden.ibm.com/cs/BlueEyes/index.html, 2011.

BOUGUET, J.Y., **Pyramidal implementation of the lucas kanade feature tracker description of the algorithm.** Intel Corporation Microprocessor Research Labs, 2000.

BRADSKI, G. & KAEHLER, A., **Learning OpenCV: Computer Vision with the OpenCV Library.** 1a edição. O'Reilly Media, 2008.

BUSSO, C.; DENG, Z.; YILDIRIM, S.; BULUT, M.; LEE, M.C.; KAZEMZADEH, A.; LEE, S.; NEUMANN, U. & NARAYANAN, S., **Analysis of emotion recognition using facial expressions, speech and multimodal information.** Sixth International Conference on Multimodal Interfaces ICMI 2004, :205–2112004.

CASTELLANO, G.; KESSOUS, L. & CARIDAKIS, G., **Emotion recognition through multiple modalities: Face, body gesture, speech.** In: Peter, C. & Beale, R., (Eds.). Affect and Emotion in Human-Computer Interaction. Heidelberg: Springer-Verlag, v. 4868 de Lecture Notes in Computer Science, 2008. p. 92–103.

DARWIN, C., **The expression of the emotions in man and animals.** Philosophical Library, 1872.

DUCHENNE, G.B.A., **Mécanisme de la Physionomie Humaine: Où, Analyse Électro-Physiologique de L'expression Des Passions.** BiblioBazaar, 1862.

EKMAN, P. & FRIESEN, W.V., **Facial Action Coding System (FACS): Manual.** Consulting Psychologists Press, 1978.

EKMAN, P. & FRIESEN, W.V., **Emotion in the Human Face.** Cambrige Univ. Press, 1982.

FRAPONOGOS, & TAYLOR, **Emotion recognition in human-computer interaction.** Elsevier Neural Networks, 18:389–405, 2005.

FREUND, Y. & SCHAPIRE, R., **A short introduction to boosting**. Journal of Japanese Society for Artificial Intelligence, :1401–14061999.

GERG, GENEVA EMOTION RESEARCH GROUP. Available at http://www.unige.ch/cisa/gerg.html, 2011.

GONZALEZ, R.C. & WOODS, R.E., **Digital Image Processing.** 2nd Edition. Prentice Hall, 2002.

GROUP, M.A.C.R., MIT AFFECTIVE COMPUTING RESEARCH GROUP. Available at http://affect.media.mit.edu/, 2011.

HAMMAL, Z.; COUVREUR, L.; CAPLIER, A. & ROMBAUT, M., **Facial expression recognition based on the belief theory: Comparison with different classifiers.** In: Roli, F. & Vitulano, S., (Eds.). Image Analysis and Processing. Heidelberg: Springer-Verlag, v. 3617 de Lecture Notes in Computer Science, 2005. p. 743–752.

HUMAINE, HUMAN-MACHINE INTERACTION NETWORK ON EMOTION. Available at ttp://emotion-research.net/, 2011.

INTEL, C., **Open source computer vision library reference manual.** Intel Corporation Microprocessor Research Labs, 2001.

IOANNOU, S.V.; CARIDAKIS, G.; KARPOUZIS, K.C. & KOLLIAS, S.D., **Robust feature detection for facial expression recognition.** Journal on Image and Video Processing, 2, 2007.

IOANNOU, S.V.; RAOUZANIOU, A.T.; TZOUVARAS, V.A.; MAILIS, T.P.; KAROUZIS, K.C. & KOLLIAS, S.D., **Emotion recognition through facial expression analysis based on a neurofuzzy network.** Elsevier Neural Networks, 18:423–435, 2005.

LIENHART, R. & MAYDT, J., **An extended set of haar-like features for rapid object detection.** ICIP: 900–9032002.

LUCAS, B.D., **Generalized Image Matching by the Method of Differences.** Tese de doutorado, Robotics Institute, Carnegie Mellon University, 1984.

LUCAS, B.D. & KANADE, T., **An iterative image registration technique with an application to stereo vision.** Journal of Japanese Society for Artificial Intelligence, :121–1301981.

NUNES, M.A.S.N., **Recommender Systems based on Personality Traits: Could human psychological aspects influence the computer decision-making process?** 1a edição. Berlin: VDM Verlag, 2009.

NUNES, M. A. S. N.; BEZERRA, J. S.; OLIVEIRA, A. A. **Estendendo o conhecimento afetivo da EmotionML**. In: IHC, 2010, Belo Horizonte - MG. Anais do IX Simpósio de Fatores Humanos em Sistemas Computacionais, 2010. v. 1. p. 197-200.

NUNES, M. A. S. N.; BEZERRA, J. S.; OLIVEIRA, A. A. **Personalityml: A Markup Language To Standardize The User**

Personality In Recommender Systems. Revista GEINTEC – ISSN: 2237-0722. São Cristóvão/SE – 2012. Vol. 2/n. 3/ p.255-273

D.O.I: 10.7198/S2237-0722201200030006

PANTIC, M. & ROTHKRANTZ, L., **Automatic analysis of facial expressions: The state of the art**. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12), 2000.

PICARD, R.W., **Affective Computing**. 2nd edição. Massachusetts, USA: MIT Press, 1997.

REEVES, B. & NASS, C., **The media equation: how people treat computers, television, and new media like real people and places**. CSLI lecture notes. CSLI Publications, 1998.

TAO, H. & HUANG, T., **Connected vibrations: a modal analysis approach to non-rigid motion tracking**. Proceedings in IEEE Conference on CVPR, 1:735–740, 1998.

TAO, J. & TAN, T., **Affective computing: A review.** In: Tao, J.; Tan, T. & Picard, R., (Eds.). Affective Computing and Intelligent Interaction. Heidelberg: Springer-Verlag, v. 3784 de Lecture Notes in Computer Science, 2005. p. 981–995.

VIOLA, P. & JONES, M., **Rapid object detection using a boosted cascade of simple features.** Conference on Computer Vision and Pattern Recognition, 2001.

VIOLA, P.; JONES, M. & SNOW, D., **Detecting pedestrians using patterns of motion and appearance.** Proceedings of the Ninth IEEE International Conference on Computer Vision, 2003.

W3C, WORLD WIDE WEB CONSORTIUM. **Emotion Markup Language (EmotionML) 1.0.** Available at http://www.w3.org/2005/Incubator/emotion/XGR-emotionml-20081120/, 2008.