

Detecção de *Landmarks* Faciais Usando SVM

Gabriel M. Araújo, Waldir S. S. Júnior, Eduardo A. B. da Silva, Siome K. Goldenstein

Resumo—Este artigo aborda o problema de detecção de *landmarks* faciais. Neste contexto, nós apresentamos um sistema de detecção de *landmarks* baseado em SVM (*Support Vectors Machine*) com kernel gaussiano. O método proposto é comparado com outros encontrados na literatura, sendo a avaliação feita em duas bases de dados, a *BioID* e a *Color FERET*. Os experimentos indicam que o método proposto supera os demais em precisão e taxa de acerto. Como o sistema proposto possui uma complexidade computacional maior que os demais métodos, podemos utilizá-lo em aplicações off-line.

Palavras-Chave—Reconhecimento de Padrões, *landmarks* faciais, Máquina de Vetor Suporte.

Abstract—This paper address the problem of locating facial landmarks. In this context, we present a system to detect landmarks based on SVM (*Support Vectors Machine*). The proposed method is compared with others found in literature. To evaluate the methods, we use two databases, the *BioID* and *Color FERET*. The experiments indicate that the proposed method outperforms the others in precision and hit rate. On the other hand, due to SVM, the proposed system has a computational cost greater than the one of other methods. This limits its use in real time applications.

Keywords—Pattern Recognition, Facial Landmarks, Support Vector Machines.

I. INTRODUÇÃO

Reconhecimento de Padrões é uma área de grande abrangência e com aplicações em vários campos de pesquisa. Alguns exemplos podem ser encontrados em [1]. Embora diversos algoritmos eficientes tenham sido desenvolvidos, verificamos que a detecção/reconhecimento de padrões é um problema em aberto.

Recentemente, notamos um crescente interesse da comunidade científica por técnicas que proporcionam a detecção/reconhecimento e o rastreamento de *landmarks* em faces humanas. Não existe uma definição universal para *landmark* em faces, porém pode-se dizer que eles determinam características salientes como os cantos dos olhos, ponta do nariz ou cantos da boca. Os algoritmos desenvolvidos podem ser utilizados em diversos problemas como o reconhecimento de faces [2] e expressões [3], estimação de pose [2], [4], modelagem de faces em 3D [5] e rastreamento de faces [6].

Os métodos desenvolvidos para detecção/reconhecimento de *landmarks* faciais podem ser classificados em dois grupos principais [7]. O primeiro grupo, consistindo de métodos globais,

utiliza informações globais, como contornos ou textura da face para localizar os *landmarks*. Estes métodos são capazes de localizar vários pontos simultaneamente com robustez. Muitas técnicas dessa categoria, como o ASM [7] (*Active Shape Model*) e o AAM [8] (*Active Appearance Models*) utilizam *templates* deformáveis. O outro grupo é o dos métodos locais, cujos algoritmos detectam cada *landmark* sem utilizar informações de outras partes da face e por isso são mais rápidos. Exemplos de métodos locais podem ser encontrados em [2], [9] e [10].

Neste artigo, nós apresentamos um sistema completo de detecção de *landmarks* em faces humanas. Trata-se de um método local que utiliza, como ferramenta principal, a bem conhecida SVM (*Support Vector Machine*). O método proposto é dividido em três partes, o pré-processamento, a detecção e o pós-processamento. No pré-processamento temos uma combinação de correção de iluminação [11], detecção de faces com o algoritmo Viola-Jones [12] e limitação da região de interesse através de um modelo Gaussiano à priori [9]. Na detecção utilizamos a SVM com kernel gaussiano, que também é conhecido como RBF (*Radial Basis Function*). A saída desta etapa é uma lista de candidatos a *landmark*. O pós-processamento escolhe um único ponto como saída. Além do sistema de detecção, outra contribuição deste artigo é o modelo utilizado na escolha dos parâmetros e treinamento dos classificadores SVM.

Os experimentos foram realizados utilizando-se duas bases de dados, a *BioID* [13] e a *Color FERET* [14], [15]. A referência foi obtida através da anotação manual de um conjunto de onze *landmarks* em todas as imagens utilizadas. Estes pontos estão representados a figura 3. O método proposto foi comparado com outros encontrados na literatura e os resultados dos experimentos demonstram que o nosso sistema supera os demais.

O restante deste artigo está organizado da seguinte maneira. Na próxima seção é feita uma revisão da literatura acerca de trabalhos relacionados. Na Seção III, é feita uma breve descrição da SVM. O método proposto é apresentado em detalhes na seção IV, com ênfase no modelo de seleção de parâmetros e treinamento dos classificadores. Na Seção V, apresentamos a metodologia utilizada, o resultados obtidos e fazemos uma discussão sobre estes resultados. Por fim, o artigo é concluído na seção VI.

II. TRABALHOS RELACIONADOS

Conforme descrito na Seção I, os métodos para detecção de *landmarks* podem ser divididos em duas categorias principais, denominadas por métodos globais e locais. Atualmente, podemos encontrar alguns sistemas locais e globais (ou públicos ou proprietários) que proporcionam soluções para o problema

Gabriel M. Araújo[†], Waldir S. S. Júnior[§], Eduardo A. B. da Silva[†], Siome K. Goldenstein[‡], [†]PEE/COPPE/DEL/POLI, Universidade Federal do Rio de Janeiro, Cx. P. 68504, Rio de Janeiro - RJ, 21945-970, BRASIL [§]DET/PPGEE/CETELI, Universidade Federal do Amazonas, Av. Gen. Rodrigo Octávio Jordão Ramos, 3000, Coroado I, Manaus - AM, 69077-000, BRASIL, [‡] IC, Universidade Estadual de Campinas, Cx. P. 6176, Campinas, SP, 13084-971, BRASIL. E-mails: gmatos@lps.ufrj.br, waldirjr@ufam.edu.br, eduardo@lps.ufrj.br, siome@ic.unicamp.br. Este trabalho foi financiado com recursos do CNPq.

da detecção de *landmarks* em faces humanas. Em particular, o sistema proposto neste artigo é categorizado como método local. Nesta seção, faremos uma breve revisão da literatura acerca de métodos locais para detecção de *landmarks* faciais.

Em [2], [16] a detecção é feita pela combinação de um modelo de posicionamento com um modelo de aparência das *landmarks*. No modelo de posição, temos que a posição conjunta das *landmarks* é modelada por uma mistura de árvores gaussianas. No segundo modelo, a aparência de cada *landmark* é considerado independente e uma adaptação do Viola-Jones ([12]) é utilizado para localizar os *landmarks*.

O sistema apresentado em [9] é construído utilizando-se uma versão de um filtro bidimensional, denominado Filtro Discriminativo. Trata-se de uma versão robusta, que incorpora em seu projeto a teoria de Análise de Componentes Principais (PCA). É importante comentar que este método possui uma interessante interpretação como um problema de restauração do impulso. A detecção é feita através da filtragem de um candidato a *landmark* \mathbf{X} por um filtro discriminativo Θ . Em outras palavras, calcula-se a convolução circular entre \mathbf{X} e Θ . Se o resultado da filtragem for um impulso, então \mathbf{X} será categorizado como um *landmark*. Note que \mathbf{X} e Θ são matrizes.

Já no sistema proposto em [10], a detecção é feita por um conjunto de detectores baseados em Filtros de Correlação, denominados IPD (*Inner Product Detector*). Estes detectores estão organizados em uma estrutura de cascata e cada estágio possui um detector \mathbf{h} e um limiar θ para cada *landmark*. Em cada estágio a detecção é feita calculando-se o produto interno entre o detector e um candidato a *landmark* \mathbf{x} . Se o resultado for maior que o limiar θ então \mathbf{x} será categorizado como ponto fiducial e passará para próximo estágio. Caso contrário o candidato será eliminado da cascata.

Por fim, em [17], [18] e [19] temos três sistemas proprietários que proporcionam soluções para o problema da detecção de pontos fiduciais. Nestes casos, encontramos um software de demonstração (demo) que pode ser utilizado para detectar pontos fiduciais em imagens fornecidas pelo usuário.

III. SVM (MÁQUINA DE VETORES SUPORTE)

A SVM, (*Support Vector Machine*) [20] é uma técnica de aprendizado de máquina que tem sido usada para solucionar problemas de classificação e de regressão. Na SVM busca-se um hiperplano de separação que possui a maior distância possível entre as amostras mais próximas das duas classes (essa distância é conhecida como margem). Em outras palavras, na SVM busca-se um hiperplano de separação que possui a maior margem possível. Em um caso de classificação binária este hiperplano H pode ser escrito da seguinte maneira:

$$f(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + b, \quad (1)$$

na qual \mathbf{w} é a direção normal à H e $|b|$ é a distância do hiperplano à origem ($\|\mathbf{w}\| = 1$). Os hiperplanos H_- e H_+ delimitam a margem, que pode ser escrita de maneira compacta da seguinte forma:

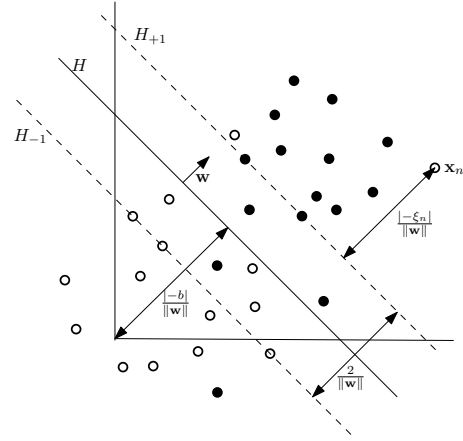


Fig. 1. SVM linear com margem suave (*soft margin*).

$$y_n(\mathbf{w}^t \mathbf{x}_n + b) - 1 + \xi_n \geq 0, \quad (2)$$

na qual \mathbf{x}_n é uma amostra do conjunto de treinamento, y_n é o rótulo de classe e ξ_n é uma variável que introduz uma tolerância, pois permite que algumas amostras estejam na região da margem ($0 \leq \xi_n \leq 1$), ou estejam na região correspondente à outra classe ($\xi_n > 1$). Por causa desta tolerância, esta abordagem é conhecida como *soft margin*. Estes conceitos estão ilustrados na Figura 1.

O hiperplano ótimo pode ser obtido resolvendo-se o seguinte problema de otimização:

$$\min_{\mathbf{w}, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \left(\sum_{n=1}^N \xi_n \right), \quad (3)$$

com uma restrição dada por

$$y_n(\mathbf{w}^t \mathbf{x}_n + b) - 1 + \xi_n \geq 0, \quad (4)$$

no qual C é um parâmetro livre que define uma penalidade sobre o erro.

As equações (3) e (4) representam um problema de otimização quadrática com restrições lineares. Adotando uma formulação lagrangeana e aplicando o truque do *kernel*, pode ser mostrado que o treinamento da SVM para o problema de reconhecimento de padrões é dado por [20]:

$$L_d = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j), \quad (5)$$

com as seguintes restrições:

$$0 \leq \alpha_n \leq C, \quad (6)$$

$$\sum_{n=1}^N \alpha_n y_n = 0, \quad (7)$$

no qual α_i é um multiplicador de Lagrange e $k(\mathbf{x}_i, \mathbf{x}_j)$ denota um *kernel* avaliado sobre as amostras \mathbf{x}_i e \mathbf{x}_j . Uma função *kernel* retorna o produto interno entre duas amostras em um espaço de dimensão elevada (possivelmente infinita), no qual as classes possivelmente são linearmente separáveis. Na Tabela III, apresentamos os tipos de *kernel* mais utilizados:

TABELA I
FUNÇÕES *kernel* MAIS UTILIZADAS

	Função $k(\mathbf{x}, \mathbf{y})$
linear	$\mathbf{x}^t \mathbf{y}$
polinomial	$(\mathbf{ax}^t \mathbf{y} + b)^d$
sigmoidal	$\tanh(\mathbf{ax}^t \mathbf{y} + b)$
gaussiano	$\exp(-\gamma \mathbf{x} - \mathbf{y} ^2)$

Usando uma notação vetorial, podemos reescrever o problema da seguinte forma:

$$L_d = \alpha^t \mathbf{1} - \frac{1}{2} \alpha^t \mathbf{Q} \alpha \quad (8)$$

com as seguintes restrições

$$0 \leq \alpha \leq C \mathbf{1}, \quad (9)$$

$$\alpha^t \mathbf{y} \quad (10)$$

nas quais α é um vetor com N entradas α_i e \mathbf{Q} é uma matriz $N \times N$, cujas entradas, são dadas por $\mathbf{Q}_{ij} = y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$.

Como a dimensão da matriz \mathbf{Q} depende do número de amostras N , o problema de otimização pode se tornar inviável quando temos um conjunto de dados suficientemente grande. Em [21] é proposta uma forma de otimizar problemas deste tipo e em [22] temos uma ferramenta denominada *SVMLight*, que implementa esta proposta. Uma descrição detalhada da SVM está fora do escopo deste artigo. O leitor interessado deve encontrar detalhes em [20] e [21].

IV. DETECÇÃO DE *Landmarks* USANDO SVM

A contribuição principal deste trabalho é a proposição de um sistema de detecção de *landmarks* usando SVM. Este sistema é constituído por três partes principais, o pré-processamento, a detecção e o pós-processamento. Na Figura 2, temos um diagrama em blocos do sistema proposto. Note que este esquema é similar ao utilizado em [10], com a diferença de que temos a SVM na etapa de classificação. Nas subseções seguintes cada bloco deste sistema é descrito em detalhes.

A. Pré-processamento

A etapa de Pré-processamento consiste em três passos, a localização da face a correção de iluminação e a seleção da região de interesse. A linha tracejada na Figura 2 indica a etapa de pré-processamento. A localização da face é feita usando o algoritmo Viola-Jones, disponível na biblioteca OpenCV [23].

Após a obtenção da localização da face é aplicada uma correção de iluminação. Neste trabalho foi aplicado o esquema de correção de iluminação proposto em [24]. A correção de iluminação consiste na aplicação de uma correção gamma, seguida por uma filtragem por diferença de gaussianas e por uma normalização de contraste. Os valores dos parâmetros utilizados nesta etapa são os mesmo recomendados por [24].

A última etapa do pré-processamento, consiste no reescalamando das imagens para uma dimensão 200×200 e na seleção da região de interesse ou ROI (*Region of Interest*). Temos 11 regiões de interesse, uma para cada *landmark*.

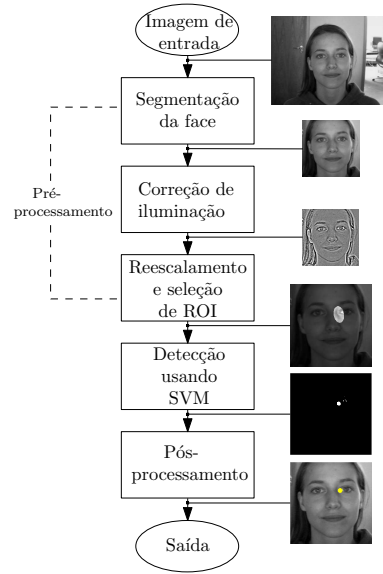


Fig. 2. Sistema proposto.

Cada região de interesse corresponde a uma região elíptica na imagem que possui uma grande probabilidade de conter o *landmark*. Estas regiões são determinadas da seguinte forma. Para cada *landmark*, nós adotamos um modelo gaussiano para a distribuição das anotações manuais. As anotações do conjunto de treinamento são então utilizadas para estimar os parâmetros do modelo: a matriz de covariâncias $\Sigma_{\mathbf{X}}$ e a média $\mu_{\mathbf{X}}$. Em seguida é selecionada a anotação manual \mathbf{x}_{\max} que maximiza a distância de Mahalanobis d :

$$d = \sqrt{(\mathbf{x} - \mu_{\mathbf{X}})^t \Sigma_{\mathbf{X}}^{-1} (\mathbf{x} - \mu_{\mathbf{X}})}. \quad (11)$$

A ROI é determinada aplicando-se uma tolerância de 5% à distância de Mahalanobis máxima (d_{\max}). Para avaliar se um candidato \mathbf{x}_c pertence à região de interesse basta verificar se a condição da Equação (12) é satisfeita.

$$(\mathbf{x}_c - \mu_{\mathbf{X}})^t \Sigma_{\mathbf{X}}^{-1} (\mathbf{x}_c - \mu_{\mathbf{X}}) \leq (1.05 d_{\max})^2. \quad (12)$$

B. Detecção

A etapa de detecção consiste na classificação das amostras dentro das regiões de interesse. Cada amostra consiste em um bloco de 13×13 pixels centrado em um ponto dentro destas regiões. A classificação é feita por um conjunto de classificadores SVM usando *kernel* gaussiano, que também é conhecido como RBF (*Radial Basis Function*). O treinamento dos classificadores foi dividido em duas partes: a determinação dos parâmetros e o aprendizado.

A SVM com *kernel* RBF possui dois parâmetros, o C e o γ (para mais detalhes, veja a Seção III, Equação (9) e Tabela III). Para determinar os valores deste parâmetros, é feito uma "grid search", que consiste em uma busca exaustiva no espaço de parâmetros. Esta busca foi feita sobre o conjunto de treinamento, no qual as amostras são selecionadas da seguinte maneira: as amostras positivas são os blocos centrados nas coordenadas das anotações manuais e nos 8 pixels vizinhos, enquanto as amostras negativas são 9 amostras centradas em

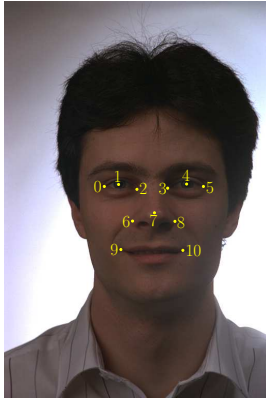


Fig. 3. Disposição dos 11 landmarks adotados.

pontos aleatórios dentro da região de interesse. Assim, dado um landmark, temos 18 amostras por imagem (9 positivas e 9 negativas). Uma vez que as amostras foram selecionadas, foi feita uma validação cruzada por k -fold com 7 folds. Com isso as amostras foram divididas em 7 partes. Para cada combinação de valores dos parâmetros, 6 partes foram utilizadas para o treinamento e 1 para a validação. Este procedimento foi repetido 7 vezes de modo que todas as partes tenham sido usadas como validação e em seguida foi calculada a média (sobre os folds) da taxa de acerto do conjunto de validação. Por fim, foram selecionadas as combinações de valores dos parâmetros que apresentaram as maiores taxas médias de acerto.

Na etapa de aprendizado, de posse dos melhores valores para C e γ , o conjunto de treinamento foi utilizado para treinar os classificadores a serem utilizados no conjunto de teste.

C. Pós-processamento

De uma maneira geral a saída da etapa de detecção não é um único ponto. Por outro lado os pontos de saída tendem a se agrupar em uma pequena região em torno do landmark desejado. Com o intuito de fornecer apenas um ponto como saída do sistema, nós aplicamos uma mediana vetorial nos pontos de saída da etapa de detecção. A mediana vetorial de um conjunto de pontos pode ser definida como sendo o ponto que minimiza a soma das distâncias euclidianas para os demais.

V. EXPERIMENTOS E DISCUSSÃO

Para avaliar o desempenho do sistema proposto, o treino e o teste foram feitos usando validação cruzada por k -fold, com 7 folds. Também foram utilizadas duas bases de dados, a *BioID* [13] e a *Color FERET* [15], ambas descritas com mais detalhes na Seção V-B. A referência das duas bases foi obtida através da anotação manual dos 11 landmarks escolhidos, como ilustrado na Figura 3.

Para a comparação foram utilizadas, além da SVM RBF, outras duas versões com SVM. Uma delas utilizando o kernel linear em uma estrutura de cascata, como descrito em [9], [10] (denotada por SVM-LIN) e a outra utilizando o kernel polinomial, como descrito em [9] (denotada por SVM-POL).

Além disso foram comparados três outros métodos descritos na Seção II. O primeiro deles utiliza filtro discriminativo com PCA [9] (denotado por FD-PCA). O segundo, descrito em [10], utiliza detectores por produto interno (IPD) em uma estrutura em cascata (denotado por IPD). O último método é o descrito em [2] (denotado por NFACE). Note que a comparação com este último método deve ser apenas qualitativa, uma vez que apenas executamos o algoritmo disponível em [25] nas nossas imagens de teste, sem efetuarmos treinamento algum.

A. Métrica de Desempenho

A métrica utilizada para avaliar o desempenho do sistema proposto é a diferença entre os rótulos manuais e os automáticos. Para padronizar esta medida, esta diferença é expressa como um percentual da distância inter-ocular. Considerando \mathbf{p}_l e \mathbf{p}_r os rótulos manuais das pupilas esquerda e direita, respectivamente, a medida e de erro é dada por:

$$e = \frac{\|\mathbf{b}_m - \mathbf{b}_a\|}{\|\mathbf{p}_r - \mathbf{p}_l\|}, \quad (13)$$

na qual \mathbf{b}_m é um rótulo manual e \mathbf{b}_a é um rótulo automático. Nesta seção, os resultados são apresentados tomando-se como base as taxas de acerto para 5% e 10% da distância inter-ocular.

B. Datasets

Em nossos experimentos foram utilizadas duas bases de dados. A primeira delas foi a *BioID* [13]. Esta base é composta por 1521 imagens em níveis de cinza de 23 indivíduos em pose frontal com resolução 384×286 pixels. Foi utilizado um subconjunto de 503 imagens, nas quais os indivíduos não possuem barba, bigode e não estão usando óculos. Os resultados referentes à *BioID* estão na Tabela II. Uma vez que landmarks simétricos apresentam resultados similares e devido à restrição de espaço, foram apresentadas as taxas de acerto para somente 6 pontos. Note que a numeração adotada para cada landmark está representada na Figura 3.

TABELA II
RESULTADOS DA *BioID*.

landmark	$e \leq 0,05$	SVM-RBF	SVM-POL	SVM-LIN	FD-PCA	IPD	NFACE
0	$e \leq 0,05$	97,62	73,73	45,51	85,10	69,78	46,03
	$e \leq 0,10$	99,40	91,04	66,59	93,64	87,28	96,95
1	$e \leq 0,05$	99,80	97,02	95,02	82,89	93,84	-
	$e \leq 0,10$	99,80	98,02	98,60	93,23	98,80	-
2	$e \leq 0,05$	99,40	85,89	43,93	70,57	22,04	20,16
	$e \leq 0,10$	99,60	93,44	79,71	92,65	79,33	91,65
6	$e \leq 0,05$	98,61	85,30	65,14	69,98	68,58	78,82
	$e \leq 0,10$	99,60	95,63	87,65	86,48	81,70	98,37
7	$e \leq 0,05$	89,27	68,81	34,80	62,04	37,79	01,22
	$e \leq 0,10$	96,03	78,94	43,15	85,69	54,30	40,33
9	$e \leq 0,05$	90,06	74,34	43,54	65,80	32,22	32,79
	$e \leq 0,10$	97,81	89,86	71,39	84,29	57,36	63,75

A segunda base de dados utilizada foi a *Color FERET* [15]. A *FERET* é composta de 11338 imagens coloridas de 994 indivíduos em várias poses e com resolução de 512×786 pixels. Para os nossos experimentos, foi selecionado um subconjunto de 2003 imagens com pose frontal, cujos indivíduos não possuem barba, bigode e não estão usando óculos. Os resultados para a *FERET*, estão na Tabela III. Novamente, foram apresentados os resultados somente para 6 landmarks.

TABELA III
 RESULTADOS DA FERET,

landmark	$e \leq 0,05$ $e \leq 0,10$	SVM-RBF	SVM-POL	SVM-LIN	FD-PCA	IPD	NFACE
0	82, 65 97, 23	65, 09 86, 26	36, 07 74, 10	54, 18 85, 41	- 90, 59	15, 26 72, 83	
1	96, 13 97, 43	86, 87 96, 88	44, 92 73, 70	78, 02 90, 39	- 93, 00	- -	
2	93, 31 97, 74	83, 80 96, 38	55, 31 85, 03	66, 96 93, 06	- 87, 77	33, 12 94, 17	
6	91, 25 97, 68	75, 73 88, 30	32, 15 61, 92	66, 45 91, 60	- 87, 62	79, 49 97, 23	
7	78, 12 95, 98	31, 73 58, 75	27, 11 61, 42	45, 72 78, 87	- 90, 31	01, 00 17, 21	
9	73, 44 94, 57	64, 94 82, 99	28, 92 53, 67	54, 58 81, 39	- 84, 65	35, 24 83, 32	

Dos resultados apresentados nas Tabelas II e III, pode-se observar que o método proposto supera os demais métodos em todos os *landmarks*. Embora as taxas de acerto sejam superiores, a complexidade computacional dos sistemas que utilizam SVM são maiores que as dos outros sistemas. Para exemplificar, considere a execução do procedimento de teste em um *fold* da base *BioID* (utilizamos 72 imagens de teste por *fold*). Neste conjunto, o método proposto teve um tempo de execução de 252 segundos, enquanto o método que utiliza o IPD em cascata (o mais rápido, pois usa produtos internos) levou 73 segundos na mesma máquina¹. Daí conclui-se que o método apresentado é perfeitamente utilizável em sistemas que possam operar *off-line*, onde o usuário pode esperar alguns segundos, enquanto os demais métodos são mais adequados para aplicações em tempo real.

VI. CONCLUSÕES

Neste artigo propomos um sistema de detecção de *landmarks* faciais baseado em SVM. O sistema proposto pode ser dividido em três partes. A primeira delas, o pré-processamento, consistem localização da face, correção de iluminação e na seleção das regiões de interesse. A segunda etapa, o núcleo do sistema, consiste em um conjunto de classificadores SVM com *kernel* RBF. Por fim, a etapa de pós-processamento consiste em aplicar uma mediana vetorial nos pontos de saída da etapa de classificação. Além do sistema, outra contribuição deste artigo é o modelo utilizado na seleção dos parâmetros da SVM. Embora a *grid search* e a validação cruzada sejam práticas comuns em Aprendizado de Máquina, a combinação destas técnicas e a escolha das amostras dos conjuntos de treinamento (incluindo aí a etapa de validação) e de teste não são tarefas triviais.

Para avaliar o sistema proposto, foram utilizadas duas bases de dados, denominadas *BioID* e *Color FERET*. Nosso método foi comparado com 5 outros. Dois deles são baseados em SVM, um deles utiliza o *kernel* linear em uma cascata e o outro utiliza *kernel* polinomial. Os outros três são métodos encontrados na literatura.

Dos resultados apresentados, pode-se concluir que as taxas de acerto do método proposto superam as taxas dos métodos apresentados para todos os *landmarks*. Embora, a complexidade computacional do método proposto possa tornar inviável a sua aplicação em sistemas de tempo real, pode-se utilizá-lo de maneira eficaz em aplicações *off-line*.

¹Utilizamos um PC com um processador Intel quad core de 3 GHz e uma memória RAM de 8 GB. Todos os sistemas foram implementados em MATLAB.

REFERÊNCIAS

- [1] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," vol. 22, pp. 4–37, 2000.
- [2] J. Sivic, M. Everingham, and A. Zisserman, "Who are you? - learning person specific classifiers from video," in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'09)*, Florida, US, jun 2009, pp. 1145–1152.
- [3] Y. li Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 97–115, 2001.
- [4] S. Zhao and Y. Gao, "Automated face pose estimation using elastic energy models," in *Proceedings of the International Conference on Pattern Recognition*, Washington, DC, USA, 2006, pp. 618–621.
- [5] P. Nair and A. Cavallaro, "3d face detection, landmark localization, and registration using a point distribution model," *IEEE Transactions on Multimedia*, vol. 11, no. 4, pp. 611–623, 2009.
- [6] J. F. Cohn, A. J. Zlochower, J. J. Lien, and T. Kanade, "Feature-point tracking by optical flow discriminates subtle differences in facial expression," in *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, Los Alamitos, CA, USA, 1998, pp. 396–401.
- [7] C. Du, Q. Wu, J. Yang, and Z. Wu, "SVM based ASM for facial landmarks location," in *Proc. IEEE International Conference on Computer and Information Technology (CIT'08)*, nov 2008, pp. 321–326.
- [8] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *Proc. of the European Conference on Computer Vision (ECCV'98)*, Freiburg, DE, jun 1998, pp. 484–498.
- [9] W. S. da Silva Júnior, G. M. Araujo, E. A. B. da Silva, and S. K. Goldenstein, "Facial fiducial points detection using discriminative filtering on principal components," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, sept. 2010, pp. 2681–2684.
- [10] G. M. Araujo, W. S. da Silva Júnior, E. A. B. da Silva, and S. K. Goldenstein, "Facial landmarks detection based on correlation filter," in *Proceedings of the IEEE International Telecommunication Symposium*, Manaus, AM, Brazil, October 2010.
- [11] D. Cristinacce and T. Cootes, "Automatic feature localisation with constrained local models," *Journal of Pattern Recognition*, no. 10, pp. 3054–3067, 2008.
- [12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, 2001, pp. I-511 – I-518 vol.1.
- [13] "Bioid database," 2010, [Último acesso em Abril de 2011]. [Online]. Available: <http://www.bioid.com/support/downloads/software/bioid-face-database.html>
- [14] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The feret database and evaluation procedure for face recognition algorithms," *Journal of Image and Vision Computing*, vol. 16, no. 5, pp. 295–306, 1998.
- [15] "The color feret database," 2008, [Último acesso em Abril de 2011]. [Online]. Available: <http://face.nist.gov/colorferet/>
- [16] M. Everingham, J. Sivic, and A. Zisserman, "Hello! my name is... buffy. automatic naming of characters in TV video," in *Proceedings of the British Machine Vision Conference*, 2006.
- [17] Disponível em: <http://www.pittpatt.com>. Acesso em: Abril de 2011.
- [18] Disponível em: <http://www.luxand.com>. Acesso em: Abril de 2011.
- [19] Disponível em: <http://developers.face.com/new-detector-demo/>. Acesso em: Abril de 2011.
- [20] V. N. Vapnik, *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995.
- [21] T. Joachims, *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA: MIT press, 1999, ch. Making Large-Scale Support Vector Machine Learning Practical, pp. 169–184.
- [22] —, "Svmlight," 2010, [Último acesso em Abril de 2011]. [Online]. Available: <http://svmlight.joachims.org/>
- [23] "Open computer vision library - opencv," 2010, [Último acesso em Abril de 2011]. [Online]. Available: <http://sourceforge.net/projects/opencvlibrary/>
- [24] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'07)*, Rio de Janeiro, BR, oct 2007, pp. 168–182.
- [25] J. S. Mark Everingham and A. Zisserman, "Automatic naming of characters in tv video," 2011, [Último acesso em Abril de 2011]. [Online]. Available: <http://www.robots.ox.ac.uk/vgg/research/nface/index.html>