Rastreamento facial e refinamento de pontos fiduciais 3D baseado na região do nariz em ambientes não controlados

Luan P. e Silva, Luciano Silva, Olga R. P. Bellon
Departamento de Informática,
Universidade Federal do Paraná (UFPR)
Av. Francisco H. dos Santos, 100, 81531-980, Curitiba, PR, Brazil
Email: {luan.porfirio,luciano,olga}@ufpr.br

***Resumo*—Neste trabalho[1] foi explorado o uso da região do nariz para realizar duas etapas essenciais para a análise de faces em ambientes sem restrição: (1) o rastreamento de faces em vídeos, e (2) o alinhamento de imagens de faces. O rastreamento da face utiliza a informação temporal para inferir a posição da mesma em cada *frame*. Quando o rastreamento é aplicado em cenários não controlados, os métodos de detecção da face podem falhar. As abordagens atuais são em geral baseadas em pontos fiduciais, porém encontram dificuldades quando aplicadas em cenários não triviais. Para lidar com esta dificuldade, o presente trabalho explora de forma original o rastreamento da região do nariz, inicializando-o no *frame* de vídeo com melhor qualidade da face. A região do nariz, ao invés da face inteira, foi escolhida devido sua menor probabilidade de estar oclusa, ser invariante a expressões faciais e visível em grande variações de pose. Foram realizados experimentos na base de dados 300 *Videos in the Wild* e *Point and Shoot Challenge*, em comparação ao rastreamento da face, mostrando que o rastreamento pela região do nariz possui melhores resultados quando imagens mais complexas são usadas. Já o alinhamento de faces em ambientes sem restrições consiste em localizar pontos fiduciais com precisão, auxiliando em tarefas como reconstrução 3D e análise de expressões faciais. Em cenários com variações de poses, a aparência da face difere da face frontal, dificultando a tarefa de alinhamento. Para contornar esta situação, este trabalho propõe refinar a localização dos pontos fiduciais com regressões em cascata e vetores de suporte (SVR), auxiliado pela pose estimada com a região do nariz. Experimentos foram realizados na base de dados 3D *Face Alignment in the Wild*, demonstrando resultados superiores ao alinhamento baseado na pose da cabeça, e comparáveis ao estado-da-arte.**

I. Introdução

Segundo Jain *et al.* [1], o reconhecimento facial é um dos problemas amplamente estudados no campo de visão computacional. Com esta finalidade, as abordagens existentes assumem que o primeiro estágio é a detecção da face. Em consideração a vídeos, a detecção da face pode ser associada à informação temporal, de forma que a localização da região de interesse nos *frames* subsequentes sejam estimados através do rastreamento.

Embora as principais abordagens de rastreamento da face utilizem *landmarks* (pontos fiduciais) [2], [3], tais métodos encontram dificuldades em lidar com ambientes sem restrições, os quais apresentam oclusões parciais, variações de escala, problemas de iluminação, resolução e diferentes orientações de pose da cabeça.

O rastreamento visual genérico é uma alternativa sem *landmarks* que tem sido aplicado com sucesso para estimar a localização de diferentes regiões, incluindo faces [4], [5]. Neste sentido, Nam e Han [4] propõem a MDNet (*Multi-Domain Network*), uma rede neural convolucional (Convolutional Neural Network - CNN) que possibilita treinar um conjunto de vídeos de objetos diversos com respectivas anotações, alcançando resultados estado-da-arte nos desafios *Visual Object Tracking* [6] e *Object Tracking Benchmark* [7]. No presente trabalho foi utilizado o rastreamento visual MDNet [4] para localizar a região do nariz em uma dada sequência de vídeo, possibilitando aumentar a confiabilidade do resultado.

O nariz é um componente facial que já foi comprovado ser eficiente para a biometria em Chang *et al.* [8], Zavan [9] e Zehngut *et al.* [10], sendo este um elemento visível em *frames* de perfil e praticamente invariante a deformações provocadas por expressões faciais, além de ser robusto quanto a oclusões mediante acessórios ou pelos faciais.

Outro fator que prejudica o rastreamento de face é a informação adquirida no *frame* inicial, o qual delimita a região de interesse a ser estimada nos *frames* seguintes, podendo apresentar baixa iluminação, oclusão ou orientação da cabeça distante da frontal. Neste sentido, foi empregado no presente trabalho a escolha automática do *frame* de melhor qualidade [9], [11] para rastreamento do nariz, evitando que o resultado obtido seja prejudicado em casos onde o primeiro *frame* contenha tais dificuldades.

Já o desafio do alinhamento facial pode ser formulado como, a partir de uma imagem de face, localizar com precisão regiões discriminantes da face, tais como olhos, nariz, boca, sobrancelha e contorno, traçando, desta forma, a geometria da face. Por meio do alinhamento facial é possível extrair informações que auxiliam no reconhecimento facial, reconstrução 3D e análise de expressões faciais.

[1]Este artigo é baseado em uma Dissertação de Mestrado

Embora o alinhamento facial tenha alcançado resultados com grande precisão, tal como constatado por Xiao *et al.* [3] e Zhu *et al.* [12] nas principais bases de dados em alinhamento 2D [13], [14], tais bases de dados contém imagens cujo sujeito apresenta a orientação da cabeça próxima ao frontal, sem a existência de imagens com faces em poses extremas (por exemplo, faces em perfil).

Para contornar tais limitações, a utilização da informação de profundidade da face na tarefa de alinhamento facial em imagens 2D [15]–[19] permite inferir a localização de *landmarks* em poses distantes da frontal, possibilitando avaliar com maior confiabilidade o desempenho dos métodos de alinhamento em imagens com poses extremas, tal como descrito por Jeni et al. [20], denominando tais abordagens como alinhamento 3D.

Em [15], [16] o alinhamento 3D é subdividido em duas etapas através de regressão com redes neurais convolucionais. Gou *et al.* [18] aproxima a localização das *landmarks* 2D mediante regressão da geometria da face, enquanto Li *et al.* [19] propõe um algoritmo de força bruta para classificação em 2D, e ambos recuperam a informação de profundidade (3D) utilizando modelos deformáveis 3D. Em Zavan *et al.* [17] foi desenvolvido um método de alinhamento 3D que utiliza como principal informação a pose extraída a partir da região do nariz, desprezando as características específicas da face durante o alinhamento.

No presente trabalho são combinadas a classificação de orientação da cabeça de Zavan *et al.* [17] com a regressão 2D em cascata de Xiong e De la Torre [21]. Adicionalmente, regressão com vetores de suporte (SVR) foi empregado, viabilizando a obtenção dos pontos faciais 3D com precisão.

Desta forma, este trabalho contribui para duas etapas necessárias à análise da biometria facial em ambientes sem restrição: o rastreamento em vídeos pela região do nariz e o alinhamento de faces em imagens.

II. Rastreamento de nariz em ambientes não controlados

Um dos objetivos do presente trabalho é contribuir para o rastreamento em ambientes sem restrições através da utilização da região do nariz como alvo para aumentar a confiabilidade do resultado, uma vez que o nariz é, em relação à face inteira e outros componentes, menos suscetível a variações de expressões faciais e oclusão. Para tanto, foi realizada a combinação da escolha automática do melhor *frame* de Zavan [9] com o rastreamento visual estado-da-arte MDNet de Nam e Han [4].

A análise de qualidade de *frames* de faces proposta em Zavan [9] possui as seguintes etapas: (1) região da face é inicialmente detectada mediante o emprego do classificador Faster-RCNN de Ren *et al.* [22]; (2) a qualidade desta é estimada pela média geométrica, com base em Abaza *et al.* [23], estimando parâmetros de contraste, brilho, foco, nitidez e iluminação; e (3) é detectado o nariz por intermédio da Faster-RCNN [22]; (4) um classificador SVM (Support Vector Machines - máquinas de vetores de suporte) utiliza a região do nariz para estimar a pose da cabeça.

O rastreamento MDNet [4] empregado para localizar o nariz nos *frames* seguintes do vídeo consiste de uma rede neural convolucional em duas etapas. Na primeira, estão as camadas compartilhadas, representadas por três camadas convolucionais e duas camadas totalmente conectadas. Para a segunda, há uma camada totalmente conectada adicional, denominada camada de domínio específico, com K ramificações, onde K é representado pela quantidade de vídeos de treinamento, de forma que cada K ramificação faz uma classificação binária de região de interesse e fundo.

Os *frames* cuja a região da face ou nariz não foram detectados são desconsiderados pela avaliação de qualidade e classificação de pose. O método de rastreamento do nariz é representado no diagrama da Figura 1, possibilitando identificar a integração da análise de qualidade [9] com o rastreamento [4]. Após escolha de melhor *frame* de inicialização, o vídeo é desmembrado em duas partes e o rastreamento é realizado, para ambos os casos, a partir do *frame* de melhor qualidade. O resultado é posteriormente reordenado para a sequência original.

A. Resultados experimentais

Foram utilizadas as bases de dados 300 Videos in the Wild (300VW) [14] e Point and Shoot Challenge (PaSC) [24], comparando o rastreamento do nariz desenvolvido neste trabalho e o rastreamento da região da face, possibilitando identificar as situações aonde o rastreamento do nariz, combinado à escolha de melhor *frame*, sejam superiores à face. Note que no rastreamento da face foi empregado o método MDNet [4] iniciado à partir do primeiro *frame* de vídeo com a região de interesse manualmente anotada.

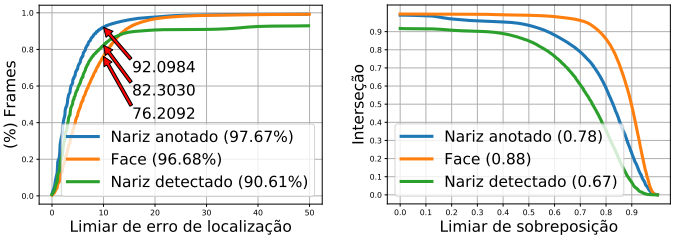
O desempenho do rastreamento visual é avaliado *frame* a *frame* utilizando duas métricas, o coeficiente de interseção [25], denominado também como taxa de sucesso [6], [7] e a precisão [6], que identifica a taxa de acerto em relação à distância da região estimada e a respectivo *ground-truth*.

1) Base de dados 300VW: A base de dados 300VW [14] dispõe de 50 vídeos para treino, 64 para testes e 68 pontos faciais anotados em cada *frame*, possibilitando extrair a região do nariz e da face para quantificar os resultados. Os vídeos de teste são categorizados em três níveis crescentes de dificuldade, consistindo de 31, 19 e 14 vídeos.

O rastreamento do nariz foi treinado com a região do nariz anotada nos 50 vídeos do conjunto de treino da base 300VW [14]. Foram realizados duas avaliações para o rastreamento do nariz: iniciando-o pela detecção automática e pelo nariz manualmente anotado, em ambos os casos a partir do *frame* de melhor qualidade.

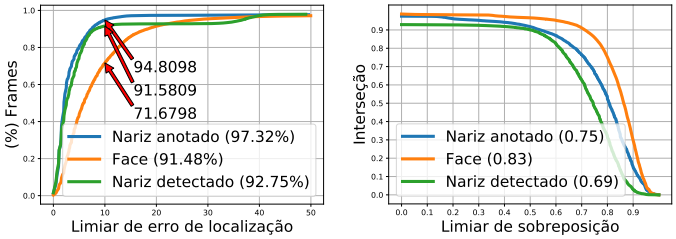
Os resultados obtidos nos 64 vídeos de testes demonstram que o rastreamento do nariz apresenta precisão superior à face, conforme Figura 2a, alcançando precisão de traslação em 97,67% quando iniciado a partir do anotação manual e 90,61% quando iniciado pela detecção automática do nariz. O rastreamento da face alcançou a precisão de 96,68%. Para todos os casos foi levado em consideração o limiar de 20 pixels

Fig. 1. Diagrama de qualidade da face e rastreamento do nariz. Em vermelho são as detecções e em azul o resultado do rastreamento [11]



(a) Gráfico de precisão. (b) Gráfico de interseção.

Fig. 2. Resultados do conjunto de testes da base 300VW [14]. Figura 2a: Entre parênteses porcentagem de *frames* cujo limiar de acerto é de 20 pixels de distância. As setas demarcam o limiar de 10 pixels. Figura 2b: A área sob a curva está disposta entre parênteses.



(a) Gráfico de precisão. (b) Gráfico de interseção.

Fig. 3. Resultados na categoria 3 do conjunto de testes da base 300VW [14]. Figura 3a: Entre parênteses porcentagem de *frames* cujo limiar de acerto é de 20 pixels de distância. As setas demarcam o limiar de 10 pixels. Figura 3b: A área sob a curva está disposta entre parênteses.

de distância, conforme adotado pela avaliação de rastreamento visual em [6].

Em uma avaliação de precisão mais restrita, reduzindo o erro para o limiar de 10 pixels de distância, o rastreamento do nariz alcança a taxa de acerto de 82,30% e 92,09%, iniciando-o pela detecção automática e anotação manual, respectivamente. Nesta margem de erro o rastreamento da região da face obtém acerto de 76,20%, comprovando melhor desempenho do rastreamento do nariz em localizar a região esperada.

Embora descrito que o rastreamento pelo nariz seja eficiente, tal região não obtém o resultado desejado quando comparado à face na avaliação do coeficiente de interseção (figuras 2b e 3b), cálculo este que indica que a sobreposição da região estimada em relação ao *ground-truth*, para todos os *frames* avaliados no conjunto de testes da base 300VW [14].

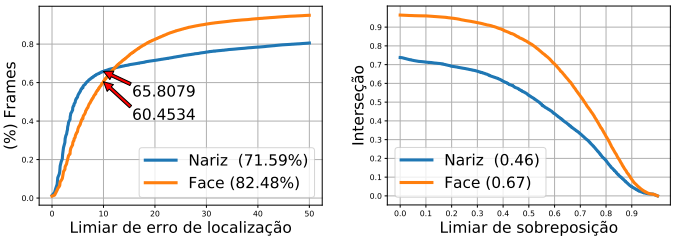
A análise visual nos resultados obtidos indica que tal evento é ocasionado devido os seguintes fatores: O resultado do rastreamento do nariz é ligeiramente maior do que o respectivo *ground-truth*, devido a dificuldade em delimitar com precisão a região do nariz da face. Já no rastreamento da face, a região de fundo a ser destacada do alvo de rastreamento apresenta notável diferença visual, o que favorece o controle de escala em se manter dentro da respectiva região de interesse.

Na avaliação individual das categorias de teste presentes na base 300VW [14] o rastreamento da região do nariz e da face obtiveram precisão similar nas categorias 1 e 2, as quais apresentam variação de iluminação e expressão facial.

Já na categoria 3, que consiste de vídeos completamente sem restrições (i.e. maior incidência de oclusão, mudança de iluminação, grande variação de pose e expressões faciais), a abordagem proposta de rastreamento pela região do nariz é superior à face, atingindo a precisão em 92,75% quando o rastreamento do nariz é iniciado pela detecção automática e precisão de 97,32% para o rastreamento do nariz iniciando com a região manualmente anotada, enquanto a face atinge 91,47% de precisão, conforme disposto na Figura 3a, implicando em maior eficiência em rastrear o nariz em detrimento à face nos cenários mais difíceis.

2) Base de dados PaSC: A base de dados PaSC [24] consiste de imagens e vídeos sem restrições com alto grau de dificuldade, no entanto, não contém anotação da região da face e nariz para avaliação. Para tanto, foram selecionados de forma aleatória 100 vídeos e anotados manualmente a região da face e do nariz.

O rastreamento da face na base de dados PaSC [24] apresentou resultado superior ao nariz, devido a baixa resolução dos *frames* e, principalmente, em situações onde existem grandes variações de escala, de forma a reduzir drasticamente a região do nariz, prejudicando seu rastreamento.



(a) Gráfico de precisão. (b) Gráfico de interseção.

Fig. 4. Resultados em 100 vídeos da base PaSC [24]. Figura 4a: Entre parênteses porcentagem de *frames* cujo limiar de acerto é de 20 pixels de distância. As setas demarcam o limiar de 10 pixels. Figure 4b: A área sob a curva está disposta entre parênteses.

Embora a precisão aferida para o rastreamento do nariz seja superior ao rastreamento da face, levando-se em consideração o limiar de 10 pixels como margem de erro (precisão de 65,30% para o nariz e 60,45% para a face), tal proporção de acerto não é mantida no decorrer da avaliação de precisão dos *frames*. No que se refere ao limiar de 20 pixels de distância, o rastreamento da região da face é superior ao nariz, atingindo a precisão de 82,48% e 71,59%, respectivamente, conforme demonstrado na Figura 4.

III. Alinhamento 3D em ambientes não controlados

Visto que o alinhamento 3D possibilita inferir a posição das *landmarks* em faces com maior precisão para imagens de faces em poses extremas [20], e que a informação da pose auxilia como etapa inicial no alinhamento [17], [26], neste trabalho é proposto o alinhamento de *landmarks* 3D em imagens faciais 2D. Este alinhamento consiste de duas principais etapas: (1) o refinamento 2D a partir de *landmarks* previamente posicionados, conforme classificação de orientação do nariz através de Zavan *et al.* [17], e (2) a regressão do eixo Z, que consiste em estimar a profundidade de cada um dos pontos 2D, conforme exemplificado na Figura 5.

Para o refinamento de pontos 2D foi utilizada a regressão em cascata de Xiong e De la Torre [21], porém substituindo o uso da face média frontal na etapa de inicialização para a face neutra rotacionada conforme a pose estimada pelo nariz, aumentando a precisão do alinhamento 2D.

Durante o treinamento do refinamento o *bounding box* da face é aumentado em 8%, em seguida a face é recortada e redimensionada para 250x250 pixels, normalizando desta forma, variações de escala e translação existentes nas imagens de treino. Para cada imagem de treino, a face neutra das *landmarks* 3D é encaixada mediante translação, escala e rotação, através da informação de pose 3D. Desta forma, o algoritmo de regressão em cascata aprende a minimizar o erro entre as *landmarks* da face neutra rotacionadas e os pontos manualmente anotados.

O alinhamento do eixo Z, correspondente à profundidade da face, foi obtido através da aplicação de regressão com

Fig. 5. Etapas do alinhamento desenvolvido: refinamento 2D dos pontos fiduciais previamente estimados pela pose e obtenção da informação 3D com base no 2D detectado.

vetores de suporte, estimando de forma independente cada *landmark*. Para tanto, foi necessário normalizar os valores das coordenadas X,Y e Z previamente, tornando-os invariantes à escala e translação, conforme realizado por Zhao *et al.* [15].

A. Resultados experimentais

O alinhamento 3D proposto foi avaliado na base de dados *3D Face Alignment in the Wild* (3DFAW) [20], composta por 23.606 imagens 2D e 66 landmarks 3D, subdividida em 13.969 imagens para treino, 4.725 imagens para validação e 4.912 imagens para para teste. A base 3DFAW [20] apresenta cenários controlados e ambientes sem restrição, grande variação de orientação, diferentes condições de iluminação, resolução da imagem e expressões faciais.

A precisão do alinhamento foi aferida através de duas métricas: o *Ground Truth Error* (GTE), que calcula distância Euclideana entre o resultado obtido no alinhamento e o ground truth normalizado pela distância intra-ocular; *Cross View Ground Truth Consistency Error* (CVGTCE), que avalia a consistência dos pontos 3D estimados em diferentes vistas para o mesmo sujeito.

O método de alinhamento foi avaliado em dois momentos: no subconjunto de validação, utilizando 13.969 imagens para treino; avaliação no conjunto de testes, realizando o treinamento com as imagens de treino e validação, totalizando 18.694 imagens. Em ambos os casos, no treinamento da estimativa de pose as imagens foram classificadas de acordo com a orientação do sujeito, variando entre -60 e 60 graus, a cada 7,5 graus, em torno dos eixos vertical e lateral, conforme relatado em Zavan *et al.* [17].

O conjunto de imagens de treino do refinamento de pontos 2D foi aumentado em 10 vezes mediante variação uniforme

TABELA I
RESULTADOS NO CONJUNTO DE VALIDAÇÃO DA BASE DE DADOS 3DFAW [20] PARA OS EIXOS 2D (XY), 3D (XYZ) E EIXOS X, Y E Z INDEPENDENTES. 1º E 2º MELHORES RESULTADOS EM AZUL E VERMELHO. O MÉTODO [21] NÃO REALIZA ALINHAMENTO 3D.

	Resultados - GTE (%)				
Eixos	XY	XYZ	X	Y	Z
Bulat & Tzimiropoulos [16]	3.626	4.940	2.12	2.48	2.77
Zavan *et al.* [17]	7.787	10.442	4.97	4.94	5.75
Xiong & De La Torre [21]	4.736	-	3.37	3.36	-
Refinamento	3.526	5.613	2.19	2.28	3.72

TABELA II
RESULTADOS OBTIDOS NO SUBCONJUNTO DE TESTES DA BASE DE DADOS 3DFAW [20] PARA GTE E CVGTCE, EM ORDEM CRESCENTE.

Métodos	Resultados	
	% CVGTCE	% GTE
Bulat & Tzimiropoulos [16]	3,4767	4,5623
Zhao *et al.* [15]	3,9700	5,8835
Refinamento	**4,035**	**6,317**
Li *et al.* [19]	4,891	7,589
Gou *et al.* [18]	4,9488	6,2071
Zavan *et al.* [17]	5,9093	10,8001

de translação (-5%, 5%), rotação (-π/4, π/4) e escala (-10%, 10%), possibilitando que o refinamento seja mais robusto a diferentes inicializações durante a etapa de testes, uma vez que a estimativa inicial das *landmarks* geradas através da aplicação do método utilizado [17] é impreciso no alinhamento local. Este acréscimo de treinamento refletiu um ganho de 15% de precisão durante os testes realizados no conjunto de validação da base de dados 3DFAW [20].

Dada a disponibilidade das landmarks 3D das faces do conjunto de validação da base de dados 3DFAW [20], foi realizada inicialmente a avaliação do desempenho individual de cada eixo (X, Y e Z), em seguida a avaliação em 2D (XY) e 3D (XYZ), comparando os resultados obtidos com: o método de alinhamento baseado na pose utilizado como base [17], o alinhamento 2D a partir da face média [21] (não utiliza a informação da pose) e o estado-da-arte [16], observando a métrica GTE, conforme destacado na Tabela I.

Na segunda coluna da Tabela I são relacionados os resultados para o alinhamento 2D (XY), sendo possível identificar que o método de alinhamento desenvolvido supera o estado-da-arte em 2,75% nesta categoria e é 25% mais preciso do que o alinhamento pela face média de [21], enfatizando o ganho em utilizar a informação precedente de pose para o alinhamento. Em relação ao método tomado como base [17], a precisão do alinhamento é superior em 54%.

Na terceira coluna da Tabela I é descrito o resultado do alinhamento 3D (XYZ), no qual o desempenho é superior ao método base [17] em mais de 46%, confirmando a consistência do resultado também no conjunto de *landmarks* 3D. Em consideração ao estado-da-arte [16], o resultado obtido é inferior devido a menor precisão alcançada ao estimar a profundidade das *landmarks* (correspondentes à linha Z), tal como destacado o desempenho individual do eixo Z na última linha da Tabela. O método de alinhamento [21] gera apenas *landmarks* 2D, não sendo possível o comparativo nesta categoria.

As três últimas colunas da Tabela I correspondem ao erro de translação de cada *landmark* estimada em relação ao *ground-truth* para cada um dos eixos X, Y e Z, independentes, sendo possível identificar a proximidade dos valores obtidos entre o refinamento e o método estado-da-arte [16].

Na Tabela II são relacionados os resultados do conjunto de testes da base 3DFAW [20] para o alinhamento 3D dos métodos estados-da-arte [15]–[19] e o refinamento proposto, comparados através das métricas GTE e CVGTCE, em ordem crescente de desempenho, onde é possível identificar que o método de refinamento aumenta a precisão do alinhamento 3D em 31% na métrica de avaliação CVGTCE, e um ganho de 41% se considerado o GTE, ambos em relação ao método base [17] que fora utilizado como inicialização.

Em relação ao resultado descrito pelo método estado-da-arte [16], o alinhamento 3D desenvolvido possui erro superior devido a dois fatores: menor precisão em estimar a profundidade (eixo Z), conforme relatado no conjunto de validação; erros obtidos na etapa inicial em detectar a região do nariz ou estimar a pose para algumas imagens do conjunto de teste, haja visto que em situações onde a pose estimada é completamente incoerente à pose da imagem, o modelo de inicialização de *landmarks* da face média encaixado na face não consegue convergir para o esperado na etapa de refinamento, gerando um resultado de alinhamento 3D inconsistente com a face.

IV. CONSIDERAÇÕES FINAIS

Este trabalho abrangeu duas etapas intermediárias utilizadas na análise biométrica da face, o rastreamento facial em vídeos e o alinhamento 3D de face em imagens 2D, ambos os casos consistindo de imagens em ambientes sem restrições.

Em relação ao rastreamento, foi demonstrado na base de dados 300VW [14] que o rastreamento do nariz supera amplamente a face (97,32% e 91,47%) em translação no cenário mais desafiador. Contudo, não é trivial delimitar o tamanho da região do nariz com precisão durante o rastreamento, dado a semelhança dos pixels da região de interesse com a face, perdendo precisão em relação à escala.

Testes realizados em 100 vídeos manualmente anotados da base de dados PaSC [24] comprovam a dificuldade encontrada pelo rastreamento do nariz em situações em que existem variações de escala, devido a redução da região de interesse a ser rastreada. Neste sentido é preferível utilizar a região da face e detrimento ao nariz para o rastreamento em tais casos.

Em um segundo momento foi elaborado um método de alinhamento 3D com a finalidade de encontrar *landmarks* em imagens sem restrição. Para tal, foi exposto uma abordagem que consiste em refinar com precisão a localização das *landmarks* da face estimadas conforme a classificação de orientação obtida pela região do nariz.

Foram realizados experimentos nos subconjuntos de validação e teste da base de dados sem restrição 3DFAW [20]. No primeiro caso, o alinhamento desenvolvido superou

o método utilizado como alinhamento inicial de Zavan *et al.* [17] em mais de 54%, o alinhamento sem auxílio da pose de Xiong e De la Torre [21] em 25% e o estado-da-arte de Bulat & Tzimiropoulos [16] em 2,75%, no que se refere ao alinhamento 2D (XY).

Em relação ao 3D (XYZ), o método de alinhamento supera o método base [17] no subconjunto de validação em 46%, porém apresenta resultado inferior ao estado-da-arte [16], devido a menor precisão em estimar a profundidade das *landmarks* (eixo Z).

O experimento realizado no subconjunto de testes da base 3DFAW [20] avaliou o método apresentado em comparação aos trabalhos mais relevantes no alinhamento facial 3D, alcançando resultados competitivos com o estado-da-arte [16]. Em relação ao método utilizado como base [17] foi constatado maior precisão em 41% na métrica de avaliação GTE e 31% em relação à avaliação aferida pelo CVGTCE.

Em alguns casos onde o nariz foi detectado incorretamente ou a pose do sujeito foi estimada de forma incoerente com o esperado, ocorreram falhas no alinhamento, não convergindo para o resultado esperado. Em trabalhos futuros o refinamento de *landmarks* 3D pode ser estendido com a informação temporal, auxiliando no rastreamento de faces em ambientes sem restrições.

AGRADECIMENTOS

Os autores agradecem ao apoio da Capes e CNPq.

V. PUBLICAÇÕES

- de Bittencourt Zavan, F. H., Nascimento, A. C., **e Silva, L. P.**, Bellon, O. R., & Silva, L. (2016, October). 3D face alignment in the wild: A landmark-free, nose-based approach. In European Conference on Computer Vision (pp. 581-589). Springer, Cham.
- **Silva, L. P.**, Zavan, F. H. D. B., Bellon, O. R., & Silva, L. (2016). Follow that nose: tracking faces based on the nose region and image quality feedback. In Conf. on Graphics, Patterns and Images-W. Face Processing.
- de Bittencourt Zavan, F. H., Gasparin, N., Batista, J. C., **e Silva, L. P.**, Albiero, V., Bellon, O. R. P., & Silva, L. (2017, October). Face Analysis in the Wild. In 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T) (pp. 9-16). IEEE.

REFERENCIAS

[1] A. Jain, P. Flynn, and A. A. Ross, *Handbook of biometrics*. Springer Science & Business Media, 2007.

[2] E. Sánchez-Lozano, B. Martinez, G. Tzimiropoulos, and M. Valstar, "Cascaded continuous regression for real-time incremental face tracking," in *European Conference on Computer Vision (ECCV)*. Springer, 2016.

[3] S. Xiao, S. Yan, and A. A. Kassim, "Facial landmark detection via progressive initialization," in *Proceedings of the International Conference on Computer Vision Workshops (ICCV Workshops)*. IEEE, 2015.

[4] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2016.

[5] S. Hong, T. You, S. Kwak, and B. Han, "Online tracking by learning discriminative saliency map with convolutional neural network," in *International Conference on Machine Learning (ICML)*, 2015.

[6] M. Kristan, J. Matas, A. Leonardis, M. Felsberg, L. Cehovin, G. Fernandez, T. Vojir, G. Hager, G. Nebehay, and R. Pflugfelder, "The visual object tracking vot2015 challenge results," in *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 2015.

[7] Y. Wu, J. Lim, and M.-H. Yang, "Object tracking benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2015.

[8] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "Multiple nose region matching for 3d face recognition under varying facial expression," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2006.

[9] F. H. B. Zavan, "Nose pose estimation in the wild and its applications on nose tracking and 3d face alignment," in *Dissertação de Mestrado, Universidade Federal do Paraná - UFPR*, 2016.

[10] N. Zehngut, F. Juefei-Xu, R. Bardia, D. K. Pal, C. Bhagavatula, and M. Savvides, "Investigating the feasibility of image-based nose biometrics," in *International Conference on Image Processing (ICIP)*. IEEE, 2015.

[11] L. Silva, F. Zavan, L. Silva, and O. Bellon, "Follow that nose: tracking faces based on the nose region and image quality feedback," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2016.

[12] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li, "Face alignment across large poses: A 3d solution," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016.

[13] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: The first facial landmark localization challenge," in *Proceedings of the International Conference on Computer Vision Workshops*. IEEE, 2013.

[14] J. Shen, S. Zafeiriou, G. G. Chrysos, J. Kossaifi, G. Tzimiropoulos, and M. Pantic, "The first facial landmark tracking in-the-wild challenge: Benchmark and results," in *International Conference on Computer Vision Workshop (ICCV Workshops)*. IEEE, 2015.

[15] R. Zhao, Y. Wang, C. F. Benitez-Quiroz, Y. Liu, and A. M. Martinez, "Fast and precise face alignment and 3d shape reconstruction from a single 2d image," in *European Conference on Computer Vision (ECCV)*. Springer, 2016.

[16] A. Bulat and G. Tzimiropoulos, "Two-stage convolutional part heatmap regression for the 1st 3d face alignment in the wild (3dfaw) challenge," in *European Conference on Computer Vision (ECCV)*. Springer, 2016.

[17] F. H. de Bittencourt Zavan, A. C. Nascimento, L. P. e Silva, O. R. Bellon, and L. Silva, "3d face alignment in the wild: A landmark-free, nose-based approach," in *European Conference on Computer Vision (ECCV)*. Springer, 2016.

[18] C. Gou, Y. Wu, F.-Y. Wang, and Q. Ji, "Shape augmented regression for 3d face alignment," in *European Conference on Computer Vision (ECCV)*. Springer, 2016.

[19] M. Li, L. Jeni, and D. Ramanan, "Brute-force facial landmark analysis with a 140,000-way classifier," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.

[20] L. A. Jeni, S. Tulyakov, L. Yin, N. Sebe, and J. F. Cohn, "The first 3d face alignment in the wild (3dfaw) challenge," in *European Conference on Computer Vision (ECCV)*. Springer, 2016.

[21] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2013.

[22] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems (NIPS)*, 2015.

[23] A. Abaza, M. A. Harrison, T. Bourlai, and A. Ross, "Design and evaluation of photometric image quality measures for effective face recognition," *IET Biometrics*, 2014.

[24] J. R. Beveridge, P. J. Phillips, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, K. W. Bowyer *et al.*, "The challenge of face recognition from digital point-and-shoot cameras," in *IEEE BTAS*, 2013.

[25] A. Hoover, G. Jean-Baptiste, X. Jiang, P. J. Flynn, H. Bunke, D. B. Goldgof, K. Bowyer, D. W. Eggert, A. Fitzgibbon, and R. B. Fisher, "An experimental comparison of range image segmentation algorithms," *IEEE ansactions on Pattern Analysis and Machine Intelligence (TPAMI)*.

[26] H. Yang, W. Mou, Y. Zhang, I. Patras, H. Gunes, and P. Robinson, "Face alignment assisted by head pose estimation," in *Proceedings of the British Machine Vision Conference (BMVC)*. BMVA, 2015.