Reconhecimento Facial 2D para Autenticação Contínua

Adeilson Antônio da Silva, Maurício Pamplona Segundo
Intelligent Vision Research Lab, Departamento de Ciência da Computação
Universidade Federal da Bahia, Brasil
E-mail: {adeilsonsilva,mauricio}@dcc.ufba.br

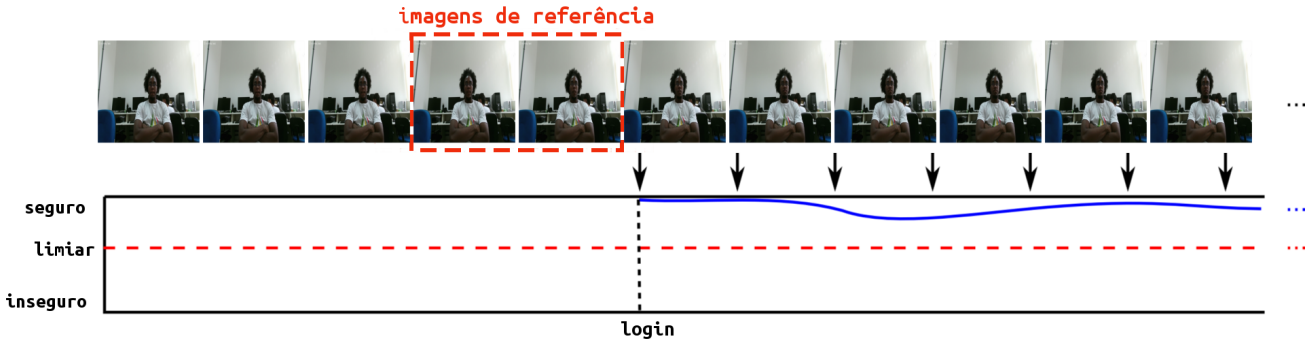

Fig. 1. Propósito do sistema: verificação contínua da identidade do usuário.

***Resumo*—Este artigo apresenta um sistema de autenticação facial 2D contínua que tem por intuito impedir o acesso não autorizado de pessoas a uma aplicação. A utilização de imagens 2D torna o sistema mais acessível, dada a facilidade de encontrar câmeras do tipo nos computadores atuais. Devido aos fatores limitantes na utilização de imagens 2D, como variações de pose e oclusão da face, um processo de normalização é utilizado para diminuir os efeitos desses problemas. Com a pontuação obtida pelo reconhecimento, a probabilidade do sistema estar seguro é calculada. O sistema foi testado utilizando vídeos de 10 minutos de duração com variações na expressão facial, angulação e oclusão, e atingiu uma taxa de erro de aproximadamente 8%.**

***Keywords*-reconhecimento facial 2D; autenticação contínua; biometria; visão computacional;**

***Abstract*—This paper presents a continuous 2D face authentication system with the purpose of forbid unauthorized people to access an application. The use of 2D images makes the system more accessible, given the ease of finding cameras of this type in current computers. Because of the limiting factors on the use of 2D images, like pose variation and face occlusion, a normalization process is used to reduce the effects of those problems. With the score given by the recognition stage, the probability of the system being safe is calculated. The system was evaluated through 10 minutes long videos with variations in facial expression, pose and occlusion, and a equal error rate of approximately 8% was achieved.**

***Keywords*-2D face recognition; continuous authentication; biometrics; computer vision;**

I. Introdução

A utilização de dados biométricos de usuários num sistema computacional fundamenta-se na ideia de que os métodos clássicos de autenticação não oferecem um alto nível de segurança. Isso acontece porque a maioria dos métodos existentes baseia-se estritamente em nomes de usuário e senhas, o que não garante a veracidade da identidade de quem realiza o acesso ao sistema [1], [2].

Um outro problema dos sistemas comuns de autenticação é o fato de que a verificação da identidade é realizada apenas uma vez. Isso também diminui a segurança da aplicação, uma vez que não há garantia de que a pessoa utilizando-a é a mesma verificada no início do acesso. A autenticação contínua resolve esse problema, uma vez que a verificação da identidade é realizada constantemente após o acesso inicial [1], [2]. Para tornar constante a validação da identidade, algumas soluções podem ser tomadas. Solicitar novamente a senha, solicitar a confirmação de algum dado pessoal, solicitar a leitura de uma determinada frase em caso de sistemas que utilizam o reconhecimento de voz, dentre outras. Entretanto, estes controles requerem a participação ativa do usuário, o que além de ser inconveniente em alguns casos, pode tornar inviável a utilização de um sistema desse tipo no dia a dia.

O processo de reconhecimento facial, ao contrário do exposto acima, pode ser feito de maneira despercebida, uma vez que não necessita da participação ativa do usuário para captura e reconhecimento. Existem diversos tipos de informações faciais que podem ser utilizadas com este intuito. Alguns trabalhos utilizam imagens 3D [1], [3], mas as pesquisas que utilizam imagens 2D são feitas a mais tempo e as técnicas utilizadas para tal estão bastante consolidadas [4],

[5], [6].

Dentre as vantagens da utilização de imagens 2D, além da existência dos mais variados algoritmos utilizados para detecção e reconhecimento facial [5], [6], [7], [8], estão a disponibilidade de uso da cor para a separação entre regiões onde há ou não há pele [9], para a detecção de regiões salientes nas imagens [10], para a localização de pontos característicos na imagem [8], [11], dentre outras finalidades. Outro ponto a ser observado é que os sensores capazes de capturar imagens desse tipo são amplamente difundidos, sendo encontrados em grande parte dos computadores pessoais atualmente. Dentre as desvantagens da utilização de imagens 2D, é possível citar o fato de elas serem afetadas de maneira significativa por variações de pose, iluminação e expressões faciais [1], [5], [8]. Apesar das dificuldades citadas, o trabalho desenvolvido tem o intuito de minimizar e/ou eliminar algumas dessas desvantagens através das técnicas para pré-processamento de imagens e normalização da pose e da iluminação. Com isso é possível obter melhores resultados nas fases de reconhecimento e de cálculo da probabilidade do sistema estar seguro.

Este artigo é dividido da seguinte maneira. Na Seção II as principais técnicas utilizadas no sistema são apresentadas. Na Seção III os resultados obtidos utilizando três vídeos capturados por um Microsoft Kinect v2 são apresentados. Por fim, na Seção IV as conclusões são apresentadas seguidas pelos agradecimentos e as referências.

II. Materiais e Métodos

Inicialmente, o sistema captura cinco imagens que servirão de referência para a fase de reconhecimento, como é possível observar na Fig. 1. Após isto, as cinco etapas do sistema a seguir são realizadas constantemente, como visto na Fig. 2 : (1) aquisição de imagens 2D, responsável pela captura das imagens utilizando o Microsoft Kinect v2; (2) detecção e rastreamento facial, que realiza o rastreamento não rígido da face em cada quadro do vídeo para fornecer à fase seguinte a localização dos pontos-chave da face, como canto dos olhos, nariz, cantos da boca, dentre outros; (3) normalização de iluminação e pose, onde as características faciais encontradas pelo rastreador são utilizadas para alinhar a face em uma posição padrão, de forma que as variações de expressão facial e da posição da cabeça tenham o menor impacto possível no resultado do reconhecimento, e então as variações de brilho e contraste da imagem são normalizadas utilizando equalização de histograma; (4) reconhecimento da face, onde padrões binários locais são utilizados para a representação da face, que é então comparada às imagens de referência para gerar uma pontuação de similaridade; e por fim a (5) fusão de pontuações, etapa onde as pontuações ao longo do tempo são combinadas para se obter a probabilidade de o sistema estar seguro em um determinado momento. Mais detalhes sobre cada etapa são dados nas subseções seguintes.

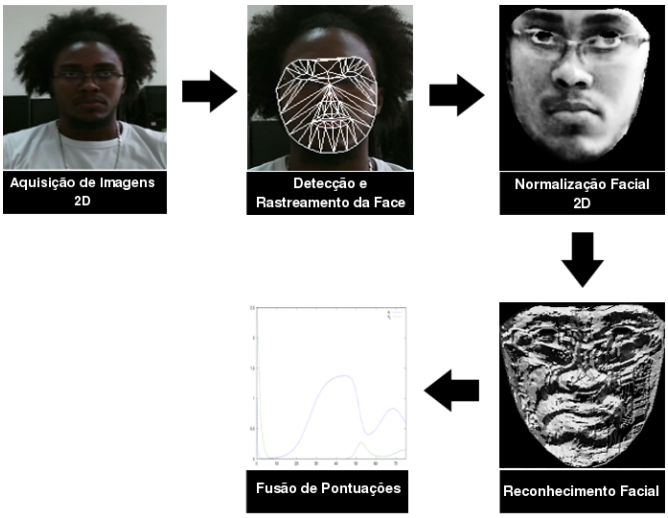

Fig. 2. Ilustração de algumas etapas do sistema de autenticação contínua.

A. Aquisição de imagens 2D

Num primeiro momento, utiliza-se o Microsoft Kinect v2 para a captura das imagens. Este dispositivo contém um sensor de captura 2D eficiente e de boa qualidade, capturando imagens com 1920×1080 pixels de resolução a uma freqüência de 30Hz (*i.e.* 15hz em baixas condições de iluminação). Apesar disso, qualquer câmera RGB comum pode ser utilizada. Após a captura, as imagens são redimensionadas para 980×540 pixels para melhorar a performance do sistema.

B. Detecção e Rastreamento facial

Após a captura das imagens, o próximo passo é a detecção da face. Para esta finalidade, a aplicação de um classificador em cascata em uma imagem utilizando janelas de busca com tamanho variável é suficiente para classificar as regiões da imagem entre face e não-face [7], [8]. Após isso, nas imagens com as faces detectadas, utiliza-se um rastreador facial não-rígido. O rastreador tem o objetivo de detectar pontos característicos da face como os cantos dos olhos, a ponta do nariz, os cantos da boca, dentre outros. A quantidade e a localização dos pontos costumam variar entre as diversas aplicações existentes, porém estes fatores podem impactar no desempenho do sistema [8]. Uma quantidade maior de pontos torna o sistema mais preciso em relação ao rastreamento da face, mas o custo computacional do processamento desses pontos é muito alto. Sistemas com uma quantidade baixa de pontos podem ser muito rápidos, porém pouco precisos em relação ao seu objetivo. Neste sistema, 65 pontos característicos são utilizados. A opção por uma técnica de rastreamento não-rígido em detrimento de uma técnica de detecção da localização da face, onde a configuração das características faciais não importa, se dá para que na etapa seguinte do sistema possa ser realizado o alinhamento dessas características, tornando o sistema robusto a expressões faciais, por exemplo.

O rastreamento funciona da seguinte maneira: utilizando descritores de características faciais, são realizadas sucessivas comparações entre cada descritor e diferentes regiões da imagem de entrada. A partir disso, são encontradas as posições que possuem uma alta similaridade com uma determinada característica facial e as outras regiões que possuem um baixo nível de similaridade com o descritor são desprezadas. Depois disso, um modelo canônico de face é levado em conta para garantir que os pontos encontrados respeitem a estrutura que uma face deve possuir [8], [11].

Como visto na Fig. 3, o rastreamento é bastante robusto às mudanças na angulação e posição da face. Oclusões mínimas na face ainda permitem o rastreamento de alguns pontos característicos. Oclusão dos olhos ou boca fazem com que a face seja totalmente perdida na imagem. Isso ocorre porque o detector facial utilizado não é robusto a oclusão [7], [8].

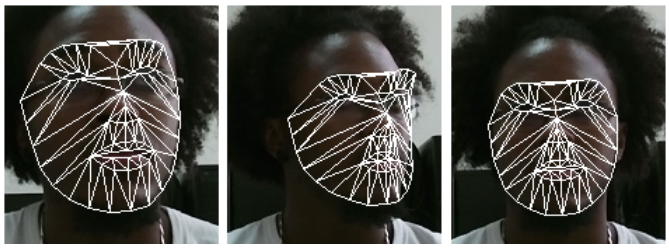

Fig. 3. Rastreamento dos pontos característicos na face e formação dos triângulos para Normalização

C. Normalização de Iluminação e Pose

Com a posição dos pontos característicos da face já determinados, agrupamos esses pontos em triângulos com tamanhos e posições diferentes ao longo da imagem, como observado na Fig. 3. A intenção por trás da formação desses triângulos é fazer o mapeamento da posição de cada ponto característico encontrado na imagem capturada pela câmera para a posição do mesmo ponto num modelo padrão de face. A justificativa para esse processo é o fato de expressões faciais, movimentação da posição da cabeça, movimentação de componentes da face, dentre outros aspectos, serem capazes de modificar completamente o resultado do reconhecimento facial. É muito importante que a face capturada no vídeo esteja o mais próximo possível das faces de referência.

Para cada triângulo formado na imagem, é calculada uma sequência de transformações lineares que colocarão este triângulo na mesma posição em que ele está localizado na face padrão. Uma vez encontrada a transformação linear necessária, basta aplicá-la, rotacionando, transladando e modificando o tamanho do triângulo para que os pontos posicionem-se como na face padrão, como visto na Fig. 4. Realiza-se ainda nesta etapa uma equalização de histograma, que tem por objetivo reduzir a variação da iluminação na imagem, como observado na Fig. 5. Esse passo se faz necessário pois, tratando-se de imagens 2D, o nível de brilho e contraste pode influenciar o resultado do reconhecimento [4], [5], [9].

Fig. 4. Normalização de pose: faces detectadas e rastreadas (cima) faces normalizadas (baixo)

Fig. 5. Equalização de Histograma: diferença entre imagens sem equalização (cima) e imagens equalizadas (baixo)

D. Reconhecimento da face

Após a normalização da face realiza-se o reconhecimento e, para tal, utilizamos Padrões Binários Locais (LBP, *Local Binary Patterns*) [4]. A escolha pelo LBP se dá porque este método alcança uma taxa maior de reconhecimento quando a variação na intensidade da iluminação é levada em conta, o

Fig. 6. Exemplos de imagens do Georgia Tech face database [12]

que tem grande impacto no resultado para imagens 2D [4], [5]. Além disso, ele obteve melhores resultados no reconhecimento do que outros métodos testados, como Fisherfaces [5] e Eigenfaces [6].

As imagens capturadas no momento do login são utilizadas para o treino do algoritmo do LBP. No LBP, cada cada pixel da imagem é comparado com seus vizinhos. Caso o valor do pixel vizinho seja maior ou igual ao valor do pixel central, o valor 1 é atribuído a esse pixel, caso contrário, o valor 0 é atribuído. Como é possível observar na Fig. 7, para cada pixel um número em formato binário é obtido através da concatenação dos vizinhos. Após obter esses números, gera-se um histograma que será utilizado para o cálculo de similaridade e consequentemente no reconhecimento.

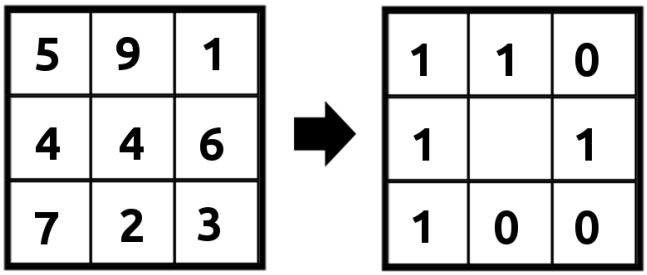

Fig. 7. Ilustração de funcionamento do LBP

E. Fusão de pontuações

A classificação entre sistema seguro e inseguro é realizada utilizando o valor da similaridade obtido através de histogramas de LBP. Utilizando a técnica de fusão de pontuações apresentada por Pamplona Segundo *et al.* [1], é possível obter a probabilidade do sistema estar seguro num determinado intervalo de tempo a partir do histórico das observações do reconhecimento. Esta probabilidade será referenciada aqui como P_{seguro}. Quanto maior o número retornado pela técnica, maior a confiança do sistema na identidade da face reconhecida. Quando esse número começa a diminuir a confiança do sistema cai, o que significa que o sistema pode estar inseguro. A partir disso, é possível estabelecer um limiar e qualquer valor abaixo deste faz com que o usuário perca imediatamente o acesso à aplicação.

Este método possui quatro parâmetros; μ_{seguro}, $\mu_{inseguro}$, σ_{seguro} e $\sigma_{inseguro}$. Estes valores representam a média e o desvio padrão da pontuação de similaridade do LBP para observações seguras e inseguras, respectivamente. Para a obtenção desses valores foram utilizadas 750 imagens normalizadas, retiradas da base de imagens Georgia Tech [12]. Os valores encontrados foram 20.2725, 43.4733, 16.1414 e 7.3204.

III. Resultados Experimentais

Para a validação do sistema desenvolvido, alguns experimentos foram realizados. Para os experimentos foram utilizados vídeos gravados por três usuários diferentes, referenciados aqui como usuário 1, usuário 2 e usuário 3, cada um com cerca de 10 minutos de duração (*i.e.* aproximadamente 3000 quadros por vídeo).

Os vídeos de dois dos usuários contém modificações constantes na posição da face, bem como oclusão da mesma. Um dos vídeos utilizados possui uma variação menor, pois o usuário está olhando para frente a maior parte do tempo. Todos os vídeos possuem variações na expressão facial.

Fig. 8. Exemplos de imagens retiradas dos vídeos utilizados para experimentos

A. Resultados da Normalização Facial

A utilização do Microsoft Kinect v2 para a captura das imagens mostrou-se eficaz devido a qualidade dos seus sensores. A resolução de 1080p das imagens capturadas é alta, o que pode impactar na velocidade das fases seguintes. Entretanto, é possível redimensionar a imagem para que as fases seguintes tenham um desempenho melhor.

Posteriormente, realiza-se o rastreamento e a extração das características faciais utilizando a biblioteca Facetracker [11], [8]. Nesta fase, modificações da posição da face são suportadas até certo ponto. Como observado na Fig. 9, variações extremas de posição fazem com que o rastreador perca a localização da face na imagem. Entretanto, não há muito o que ser feito pois essa é uma limitação natural quando trabalha-se com imagens 2D [8].

Um dos objetivos do rastreamento de pontos é garantir informações suficientes para a realização da normalização das faces encontradas (reposicionamento das faces na imagem e remoção de expressões faciais, como observado na Figura Fig. 4). Como observado na Fig. 10, a técnica de normalização aqui discutida não é robusta a variações intensas na angulação da face (*i.e.* olhar para os lados, para cima ou para baixo). Essas variações causam uma deformação considerável na face normalizada.

Ainda na etapa de normalização, utilizando equalização de histogramas, a variação de briho e contraste na imagem de entrada é diminuida. A técnica de equalização de histogramas utilizada nesta etapa mostrou-se eficaz na maior parte dos

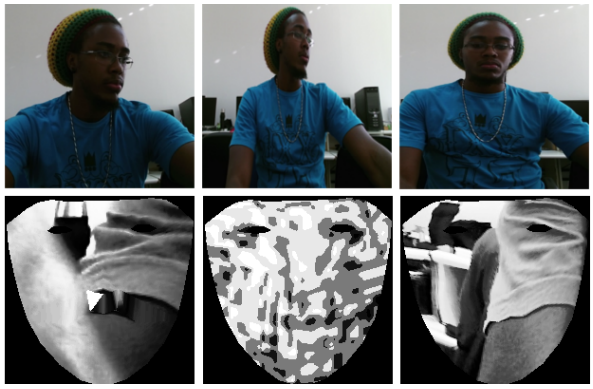

Fig. 9. Localização da face não encontrada pelo rastreador

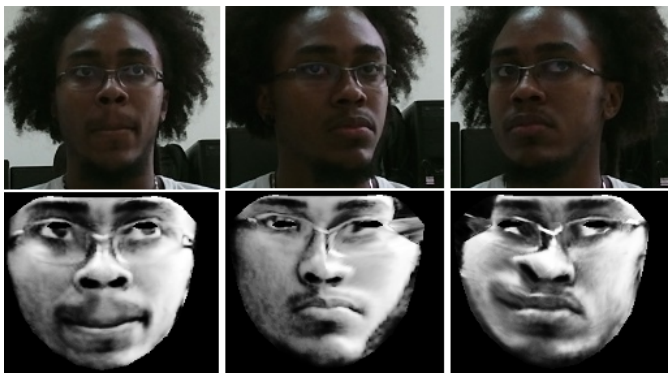

Fig. 10. Deformação da face quando há variação na angulação da face

casos, reduzindo bastante a variação da iluminação nas imagens, como observado anteriormente na Fig. 5.

B. Resultados da Autenticação Contínua

Para a validação desta etapa, os vídeos foram concatenados dois a dois com a intenção de simular uma invasão ao sistema. Como observado na Fig. 8, variações na posição da face são levadas em consideração, bem como variações de iluminação. Os experimentos foram divididos da seguinte forma:

Primeiro experimento: Usuário 1 utiliza o sistema e os usuários 2 e 3 são os invasores.

Segundo experimento: Usuário 2 utiliza o sistema e os usuários 1 e 3 são os invasores.

Terceiro experimento: Usuário 3 utiliza o sistema e os usuários 1 e 2 são os invasores.

Nesta última etapa do sistema acontece o reconhecimento. Dadas as imagens capturadas no início da aplicação, que foram utilizadas para treinar o algoritmo utilizado, e as imagens capturadas posteriormente, realiza-se a comparação através do LBP. Quando a etapa de normalização ocorre sem problemas, a taxa de acerto do reconhecimento é alta, com altos níveis de confiança. Quanto mais deformada a imagem resultante da normalização, menor a confiança dada pelo algoritmo em relação à similaridade da face. Percebe-se a partir disso que a etapa de normalização é a que tem maior impacto no resultado do reconhecimento e deve ser bastante precisa.

Como observado nas Figuras 11, 12 e 13, os melhores resultados foram obtidos no terceiro experimento. Isso ocorreu pois o vídeo do usuário 3 é o que possui menor quantidade de variação da posição da face na imagem.

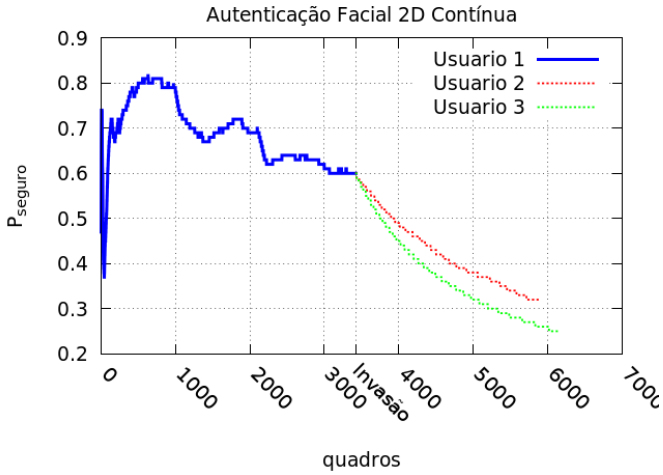

Fig. 11. Resultados do primeiro experimento.

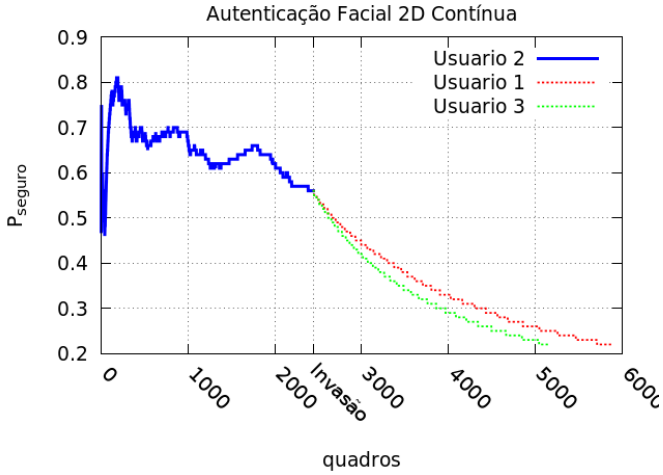

Fig. 12. Resultados do segundo experimento.

Em todos os experimentos o resultado da fusão de pontuações foi maior durante a permanência do usuário do que dos invasores. Os valores das probabilidades foram altos nos momentos em que a normalização foi correta e a face normalizada estava bastante próxima das imagens de referência. Os resultados podem ser vistos de uma outra forma na curva de características operacionais do receptor (ROC, *Receiver Operating Characteristics*) para os valores de P_{seguro} mostrados na Fig. 14, atingindo uma taxa de erro de aproximadamente 8%.

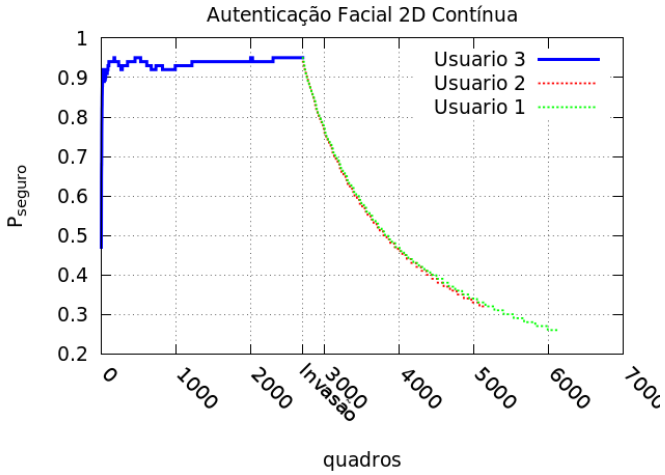

Fig. 13. Resultados do terceiro experimento.

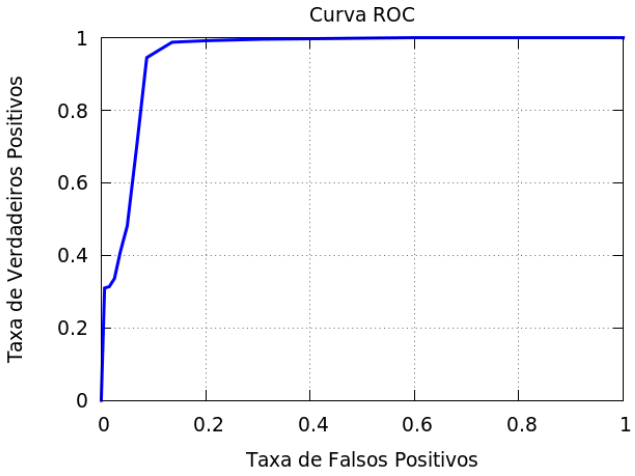

Fig. 14. Curva de características operacionais do receptor.

IV. Conclusões

Neste artigo demonstramos a implementação de um sistema robusto de reconhecimento facial para utilização na autenticação contínua. O sistema proposto tem um bom desempenho e demonstrou ser adequado para uso nas mais diversas situações.

O sistema proposto captura imagens 2D através do Microsoft Kinect. As faces são detectadas utilizando um classificador em cascata e utilizando um rastreador facial não-rígido características da face são extraídas para realizar a normalização da pose e a equalização do brilho e contraste, e então realizar a comparação com as imagens do banco de dados. A partir dos experimentos realizados é possível chegar às seguintes conclusões:

1) As técnicas utilizadas para a detecção e para o rastreamento das faces nas imagens mostraram-se eficazes mesmo com as limitações naturais das imagens 2D.
2) A técnica utilizada para normalização da pose não comportou-se da maneira esperada quando ocorrem variações intensas na angulação da face, modificando o resultado do reconhecimento e consequentemente da autenticação contínua.
3) A técnica utilizada para reconhecimento mostrou-se eficaz, apesar das limitações impostas pela etapa de normalização.
4) O cálculo da probabilidade de segurança do sistema mostrou-se eficaz durante os experimentos. Apesar disso, algumas melhorias podem ser feitas para que a detecção de invasores ocorra num tempo menor.

Para trabalhos futuros, melhorias na técnica de normalização serão implementadas para que não ocorram deformações tão grandes nessa etapa. Além disso, o método de cálculo da fusão de pontuações será atualizado para que as invasões sejam detectadas mais rapidamente.

Agradecimentos

Os autores gostariam de agradecer à Universidade Federal da Bahia (UFBA) que tornou esse projeto possível através do Programa Institucional de Bolsas de Iniciação Científica (PIBIC) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro.

References

[1] M. Pamplona Segundo, S. Sarkar, D. Goldgof, L. Silva, and O. Bellon, "Continuous 3d face authentication using rgb-d cameras," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2013, pp. 64–69.

[2] S. Zhang, R. Janakiraman, T. Sim, and S. Kumar, "Continuous verification using multimodal biometrics." in *Pattern Analysis and Machine Intelligence, IEEE Transactions on (Volume:29 , Issue: 4)*., 2007, pp. 687–700.

[3] F. Al-Osaimi, M. Bennamoun, and A. Mian, "An expression deformation aproach to non-rigid 3d face-recognition," in *International Journal of Computer Vision (IJCV)*, 2009, pp. 302–316.

[4] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns." in *European Conference on Computer Vision.*, 2004, p. 469–481.

[5] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," in *IEEE Transactions on pattern analysis and Machine Intelligence, Vol.19, No. 7*, 1997, pp. 711–720.

[6] M. Turk and A. Pentlan, "Eigenfaces for recognition," in *Journal of Cognitive Neuroscience*, 1991, pp. 71–86.

[7] P. Viola and M. J. Jones, "Robust real-time face detection," in *International Journal of Computer Vision (IJCV)*, 2004, p. 137–154.

[8] D. L. Baggio, S. Emami, D. M. Escrivá, K. Ievgen, N. Mahmood, J. M. Saragih, and R. Shilkrot, "Mastering opencv with practical computer vision projects," 2012.

[9] F. Y. Shih, S. Cheng, C. Chuang, and P. S. P. Wang, "Extracting faces and facial features from color images," in *International Journal of Pattern Recognition and Artificial Intelligence (IJPR)*, 2008, pp. 515–534.

[10] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007, p. 1–8.

[11] J. M. Saragih, S. Lucey, and J. F. Cohn, "Face alignment through subspace constrained mean-shifts," in *International Conference of computer Vision (ICCV)*, 2009.

[12] "Georgia tech face database," 1999. [Online]. Available: http://www.anefian.com/research/face_reco.htm