

Effect of Nonrigid Alignment Using Free-Form Deformation in Frontal Face Images

Igor R. R. Xavier
Centro Universitário da FEI
São Bernardo do Campo, Brasil.
iradames@fei.edu.br

Carlos E. Thomaz
Centro Universitário da FEI
São Bernardo do Campo, Brasil.
cet@fei.edu.br

Abstract—Face alignment is an important step in any task related to face recognition. Basically, this step is performed applying a transformation that maps points from a face image to the same points in the reference (or target) one. These transformations can be rigid or affine, but due to some specific and local deformations caused by facial expressions, for instance, the application of different alignment techniques must be necessary. In this paper we have used Free-Form Deformation based on B-Splines to align frontal face images with support of a number of basic facial landmarks extended here automatically. Our results have shown that with such approach is possible to create very realistic human-like mean images using only few samples of frontal faces.

Keywords—Nonrigid alignment, Free-Form Deformation, microexpressions, expressions, face recognition.

I. INTRODUÇÃO

No contexto de reconhecimento de faces e expressões faciais, uma importante etapa é o alinhamento das imagens com relação a uma face de referência para posterior extração de características e comparações com uma base de conhecimento pré-existente [1], [2]. Em outras palavras, antes de comparar a informação existente em duas faces distintas, deve-se alinhar espacialmente a localização das características de uma imagem em relação às mesmas características da outra.

Na verdade, mesmo duas imagens de face da mesma pessoa podem apresentar variações de pose, escala e expressão, fazendo com que a tarefa de identificação se torne muito mais problemática sem o alinhamento correto dessas amostras. No caso de imagens da mesma pessoa, um alinhamento rígido pode corrigir a variação de pose e um alinhamento afim pode corrigir a variação de escala. Porém, quando trabalhamos com imagens de pessoas distintas ou quando lidamos com expressões faciais, pode haver variações de natureza não-rígida.

Em algumas áreas de pesquisa, a etapa de alinhamento pode ser ainda mais determinante. Por exemplo, na tentativa de detecção e reconhecimento de micro-expressões faciais, o alinhamento precisa ser feito entre muitas imagens, pois uma micro-expressão é uma variação espacial e temporal da mesma face [3] usualmente manifestada sutilmente. Em [4], cada face em uma sequência de quadros de um vídeo foi alinhada rigidamente às demais utilizando as coordenadas dos centros dos olhos. Em [3], [5] e [6], nos quais experimentos foram conduzidos no intuito de detectar e reconhecer micro-expressões faciais de origem espontânea, as imagens de face foram marcadas com 68 pontos usando o método *Active Shape*

Model [7], amplamente conhecido, para depois alinhá-las a uma face de referência aplicando uma função de transformação não-rígida descrita em [8]. Esta última abordagem reforça a necessidade de alinharmos as imagens de forma não-rígida para levar em consideração deformações globais e locais simultaneamente.

Neste trabalho, utilizou-se além de transformações rígidas e afins, uma deformação livre de forma baseada em *B-Splines* e comumente denominada em inglês de *Free-Form Deformation* (FFD) [9]. Esta técnica vem sendo amplamente aplicada com sucesso no contexto de imagens médicas. Rueckert *et al.* [10] utilizaram FFD para alinhar imagens de ressonância magnética de mamas e consequentemente melhorar o contraste e detecção de tumores. Neste caso, a descoberta de tumores pode ser prejudicada pela movimentação do paciente no momento do exame, entre outros fatores. Chandrashekar *et al.* [11] utilizaram *Multilevel* FFD em imagens de ressonância magnética marcadas para construção de modelos 4D de funcionamento do coração. Papatheodorou [12] usou FFD para aprimorar técnicas de reconhecimento facial através do alinhamento e representação de faces em 3D. Mais recentemente, Sandbach *et al.* [13] utilizaram FFD em um sistema de reconhecimento de expressões faciais em 3 dimensões.

Neste artigo, estudou-se o efeito do alinhamento de faces baseado em pontos fiduciais usando FFD e o impacto do aumento do número de pontos na qualidade do alinhamento para poucas amostras disponíveis. A principal contribuição deste estudo está na melhoria da qualidade do registro de imagens frontais de face, sem o uso intensivo de segmentação de imagens, necessária em abordagens holísticas de reconhecimento de padrões [14], [2], [15] e [16]. Portanto, acredita-se que qualquer abordagem que trate uma imagem de face como um vetor de atributos e aplique métodos estatísticos para redução de dimensionalidade e extração de características possa tirar proveito de um alinhamento espacial melhor entre as imagens devido à redução de variabilidade das amostras obtida pelo algoritmo descrito aqui.

O restante do trabalho está dividido da seguinte forma. A seção II descreve sucintamente os métodos de alinhamentos existentes e apresenta o algoritmo proposto. A seção III descreve as imagens utilizadas neste trabalho. A seção IV discute os resultados obtidos e, por fim, a seção V finaliza o artigo com conclusões e possibilidades de trabalhos futuros.

II. MÉTODOS DE ALINHAMENTO

A. Conceitos Fundamentais

Geralmente, o alinhamento de imagens é realizado aplicando-se uma transformação T que mapeia o ponto a de uma face para o correspondente ponto b de outra face, tal que:

$$\mathbf{b} = T(\mathbf{a}), \quad (1)$$

onde $\mathbf{a} = (a_x, a_y)$ e $\mathbf{b} = (b_x, b_y)$ [12]. As transformações podem ser classificadas em dois tipos: aquelas que preservam as proporções geométricas das coordenadas iniciais, portanto, não deformam o objeto transformado, e aquelas que não preservam essas proporções e, por sua vez, deformam o objeto. Transformações rígidas e afins são do tipo que preservam as proporções geométricas. Uma transformação rígida é uma transformação geométrica que quando aplicada a um objeto, por exemplo, uma imagem de face, mantém todas as distâncias entre os pontos do objeto e também seus ângulos internos, sendo uma combinação linear de rotação e translação, tal que $\mathbf{b} = \mathbf{R}_{AB}\mathbf{a} + \mathbf{t}_{AB}$, cuja forma matricial é descrita por:

$$T_{rígida}(a_x, a_y) = \begin{pmatrix} b_x \\ b_y \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix} \begin{pmatrix} a_x \\ a_y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}, \quad (2)$$

onde $\mathbf{R}_{AB} = \{r_{ij}\}, i, j \in \{1, 2\}$, é a matriz de rotação da transformação. Um exemplo de transformação rígida aplicada em uma imagem de face pode ser observado na Figura 1.

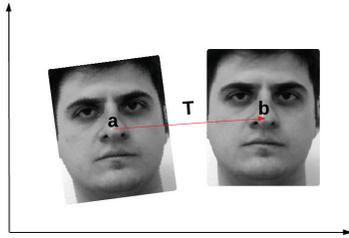


Figura 1: Transformação rígida aplicada a uma imagem do banco de dados de faces da FEI. Percebe-se que houve rotação e translação da face exemplificada pelo deslocamento do ponto a para o ponto b .

Uma classe mais geral de transformações lineares é a transformação afim. Esta além de combinar rotação com translação também promove mudança de escala e recortes no objeto transformado. Sua forma matricial é:

$$T_{afim}(a_x, a_y) = \begin{pmatrix} b_x \\ b_y \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} a_x \\ a_y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}, \quad (3)$$

onde $\mathbf{A}_{AB} = \{\alpha_{ij}\}, i, j \in \{1, 2\}$, é a matriz que descreve os componentes de escala, recorte e rotação da transformação. Um exemplo de transformação afim pode ser visto na Figura 2.

Porém, muitas vezes, o mapeamento entre dois objetos pode não ser linear para todas as coordenadas, ou seja, uma transformação global não é suficiente para capturar a deformação necessária para alinhar um objeto ao outro, como no caso de imagens de face onde as deformações podem

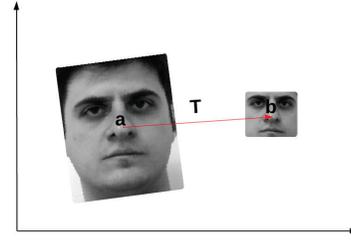


Figura 2: Transformação afim aplicada a uma imagem do banco de dados de faces da FEI. Percebe-se que além de rotação e translação da face, houve também mudança de escala e recorte da região central da face.

ocorrer global e localmente. Tais transformações são ditas não-afins, não-lineares ou não-rígidas. Na prática, esse tipo de transformação é definida por pontos de controle ou marcações no objeto e a deformação é suavemente interpolada por pontos intermediários [17]. Transformações baseadas em *Splines* atuam nos níveis local e global. Estas foram originalmente concebidas para modelagem computacional, projeto de aviões, carros ou em atividades onde os projetistas usavam tiras de metal ou madeira com ajuda de pesos para deformação [18]. Um esquema visual de transformação não-rígida pode ser observado na Figura 3.

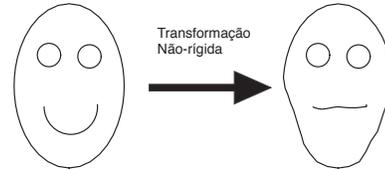


Figura 3: Transformação não-rígida aplicada a uma face pode deformar uma face global e localmente.

B. Algoritmo Proposto

O diagrama de blocos do algoritmo proposto neste trabalho pode ser visto na Figura 4. O algoritmo consiste dos seguintes passos:

- 1) Estimar a transformação rígida necessária para alinhar os pontos da imagem de face k aos pontos da imagem padrão com base apenas em um conjunto de pontos fiduciais previamente computados;
- 2) Estimar a transformação afim necessária para alinhar os pontos da imagem de face k aos pontos da imagem padrão com base apenas em um conjunto de pontos fiduciais previamente computados e o resultado do alinhamento rígido;
- 3) Estimar a transformação não-rígida necessária para alinhar os pontos de face k aos pontos da imagem padrão com base apenas em um conjunto de pontos fiduciais e o resultado do alinhamento afim;

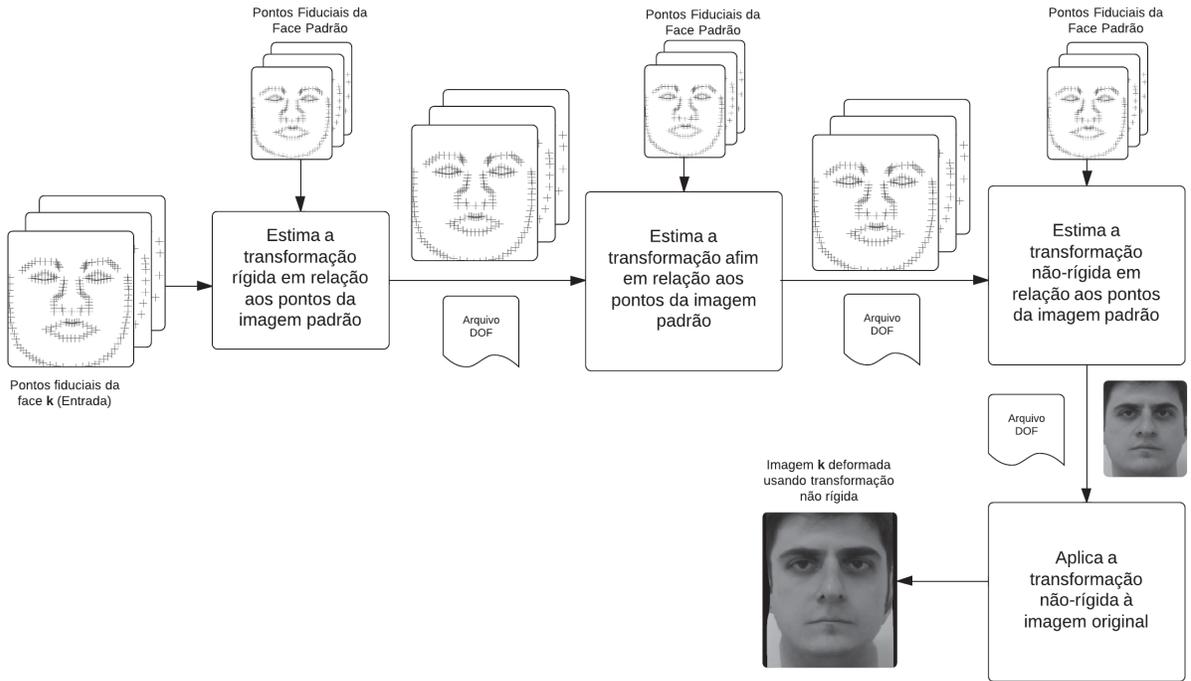


Figura 4: Diagrama de blocos do algoritmo proposto.

- 4) Computar a imagem de face k aplicando a transformação não-rígida com base na face k original e o resultado do alinhamento não-rígido.

O arcabouço computacional descrito na Figura 4 é aplicado para diversas combinações de pontos fiduciais. A partir de um conjunto inicial de m pontos, utilizou-se o algoritmo 1 para computar um ponto médio entre cada par de pontos sequenciais, gerando como saída um novo conjunto de n pontos, onde $n \geq m$ e ε é a máxima distância que dois pontos podem ter para serem interpolados, determinada empiricamente neste trabalho.

Algorithm 1 InterporlarPontos(m)

```

 $k \leftarrow 1$ 
for  $i = 1$  to  $\text{comprimento}[m] - 1$  do
  if  $0 < \text{dist}(m(i), m(i + 1)) < \varepsilon$  then
     $n(k) \leftarrow m(i)$ 
     $n(k + 1) \leftarrow [m(i) + m(i + 1)]/2$ 
     $n(k + 2) \leftarrow m(i + 1)$ 
     $k \leftarrow k + 2$ 
  end if
end for
return  $n$ 

```

III. MATERIAL

Para este estudo, escolheu-se aleatoriamente 15 imagens de face de 250x300 pixels, neutras e frontais, do banco de dados

de faces da FEI [19], conforme Figura 5. Além disso, foram utilizados como pontos fiduciais iniciais as $m = 46$ marcações feitas no trabalho [20] no qual foi utilizado o mesmo banco de dados. A face média das imagens neutras do banco de dados da FEI também foi anotada com os mesmos 46 pontos fiduciais com intuito de ser utilizada como face padrão (referência), conforme Figura 6.

IV. EXPERIMENTO E RESULTADOS

Para efeito de comparação, calculou-se a média das 15 imagens selecionadas, as quais já estavam alinhadas no banco de dados em relação somente às coordenadas dos centros dos olhos de cada face. O resultado pode ser visto na Figura 7.

O experimento foi dividido em três etapas, cada uma com um determinado número de pontos fiduciais anotados. Aproveitou-se a marcação feita previamente no banco de dados [20] e a partir dela aumentou-se o número de pontos para as etapas seguintes. Em cada etapa, aplicou-se o algoritmo da Figura 4 nas 15 imagens selecionadas e calculou-se a média das faces para comparação.

Utilizando as imagens anotadas, aplicou-se a transformação não-rígida. Essa transformação deforma as imagens originais na tentativa de alinhar os pontos fiduciais de cada imagem aos pontos da imagem de referência. Um resultado desta transformação pode ser visto na Figura 8 usando 46 pontos fiduciais. A fim de melhorar o resultado do alinhamento, interpolaram-se os 46 pontos iniciais utilizando o algoritmo



Figura 5: As 15 imagens de faces selecionadas aleatoriamente do banco de dados da FEI.



Figura 6: Faces originais do banco de dados da FEI marcadas com 46 pontos fiduciais. A imagem da esquerda mostra a face padrão (face média de todos os indivíduos da base) marcada com os 46 pontos e a imagem da direita mostra uma das faces do banco de dados marcada com os 46 pontos correspondentes aos da imagem padrão.

1 proposto. Os resultados da primeira e segunda rodadas de inclusão de novos pontos podem ser vistos na Figura 9.

O processo automático de inclusão de novos pontos criou duas novas máscaras com 86 e 166 pontos. Aplicou-se a estas novas configurações o algoritmo de alinhamento não-rígido e o resultado pode ser observado na Figura 10.

Por fim, calculou-se a imagem média dos resultados do alinhamento não-rígido das 3 configurações de pontos: 46, 86 e 166 pontos, conforme a Figura 11.

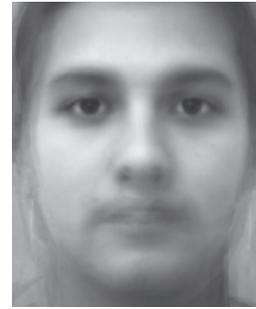


Figura 7: Imagem média resultante do alinhamento rígido das faces selecionadas aleatoriamente do banco de dados da FEI.



Figura 8: Resultado de uma transformação não-rígida usando 46 pontos fiduciais como entrada para o algoritmo proposto.

Pode-se observar comparando, especificamente, as Figuras 8 e 10 que o aumento na quantidade de pontos fiduciais de entrada para o algoritmo de alinhamento não-rígido fez com que a rugosidade do resultado final diminuísse consideravelmente. Isto pode ser observado tanto nas imagens de exemplo quanto na média da amostra, onde a rugosidade das amostras permanecia clara nas faces médias correspondentes. Percebe-se também, pela Figura 11, que o uso do alinhamento não-rígido proporcionou uma melhora significativa na visualização das características médias das amostras. Além dos olhos, a boca, o nariz e as sobrancelhas puderam ser alinhados de forma bastante realista em comparação com a face média do alinhamento rígido (Figura 7).

Como medida da qualidade do alinhamento, calculou-se o desvio padrão médio de cada pixel em cada conjunto de imagens, ou seja, para o conjunto original sem a aplicação do algoritmo proposto neste trabalho e para os conjuntos de imagens transformados não-rigidamente com os acréscimos automáticos das quantidades de pontos fiduciais. Conforme pode-se observar na Figura 12, o alinhamento não-rígido diminuiu a variância em regiões da face onde os estudos, principalmente sobre expressões e micro-expressões faciais, mais se concentram, por exemplo.

Como medida quantitativa, calculou-se a decomposição espectral da matriz de covariância de cada conjunto de imagens alinhadas. A Figura 13 mostra a magnitude desses autovalores. Esta análise de componentes principais foi aplicada à região de interesse definida pela envoltória convexa dos pontos fiduciais

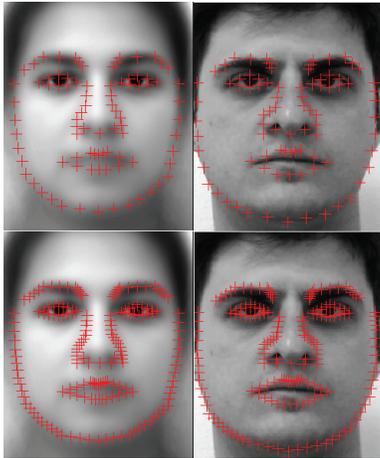


Figura 9: Resultado da inclusão automática de novos pontos na marcação inicial de 46 pontos. Acima e à esquerda: a face média do banco de dados da FEI marcada com 86 pontos fiduciais, acima e à direita: uma das faces selecionadas marcada com 86 pontos, abaixo e à esquerda: a face média do banco de dados da FEI marcada com 166 pontos fiduciais e abaixo e à direita: uma das faces selecionadas marcada com 166 pontos.



Figura 10: Resultado do alinhamento não-rígido utilizando as máscaras com 86 pontos (esquerda) e 166 pontos (direita) calculados automaticamente pelo algoritmo proposto.

anotados em cada grupo de imagens.

Calculou-se também, o coeficiente de correlação médio de cada grupo de imagens definido pela região de interesse, conforme tabela I.

Tipo de Alinhamento	CC
Rígido	0.83
Não-rígido 46 pontos	0.96
Não-rígido 86 pontos	0.97
Não-rígido 166 pontos	0.96

Tabela I: Coeficiente de correlação médio obtido entre todas as imagens de cada grupo.



Figura 11: Médias de imagens de face pós alinhamento não-rígido com 46 (esquerda), 86 (centro) e 166 (direita) pontos respectivamente.

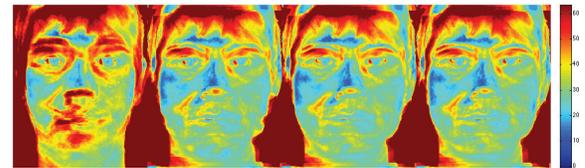


Figura 12: Desvio padrão médio de cada pixel em cada grupo de imagens gerado. Da esquerda para direita: faces alinhadas pela transformação afim usando os centros dos olhos, faces alinhadas pelo método não-rígido usando 46, 86 e 166 pontos fiduciais, respectivamente.

V. CONCLUSÃO

Neste trabalho, explorou-se uma técnica de alinhamento de faces utilizando um algoritmo para alinhamento não-rígido com o objetivo de diminuir os efeitos de problemas de escala e deformação, e até mesmo de morfologia, entre as amostras de um banco de dados de face disponível publicamente.

Os resultados observados no cálculo das faces médias indicam que o aumento automático proposto de pontos fiduciais melhora o alinhamento das características da face e que este aumento pode ser conseguido partindo-se de uma marcação básica e incrementando o número de pontos por meio de uma simples interpolação de pares sequenciais com seus respectivos pontos médios. Adicionalmente, nota-se que o alinhamento espacial foi feito utilizando somente 15 imagens de face, ou seja, mesmo com poucas amostras a média das faces transformadas apresenta características de uma face humana extremamente realista. Este resultado pode ser comparado ao de outras técnicas de síntese e transformação de faces como em [21], porém com a vantagem de utilizar, por meio do algoritmo proposto, poucos pontos base fiduciais.

Como trabalhos futuros, acredita-se que seria interessante verificar a combinação desta técnica com métodos automáticos de marcação de pontos fiduciais como em [22] e extrapolar essa marcação inicial utilizando, possivelmente como informação *a priori*, medidas antropométricas da face.

AGRADECIMENTOS

Os autores gostariam de agradecer o suporte financeiro provido pelo programa PROSUP/CAPES e também ao Dr. Daniel Rueckert por disponibilizar as ferramentas utilizadas neste trabalho para o alinhamento das faces. Estas ferramentas se encontram disponíveis em <http://www.doc.ic.ac.uk/~dr/software/>.

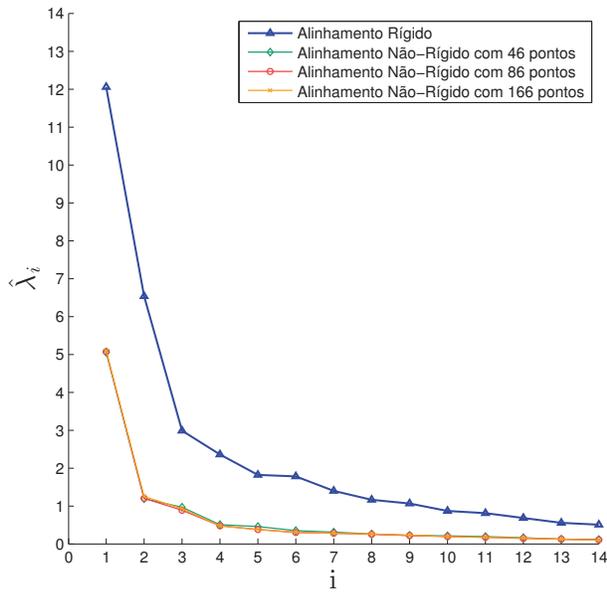


Figura 13: Os n autovalores não nulos das matrizes de covariância, obtidos através da decomposição espectral de cada conjunto de imagens.

REFERÊNCIAS

- [1] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Computer vision-ecv 2004*. Springer, 2004, pp. 469–481.
- [2] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 1, pp. 103–108, 1990.
- [3] T. Pfister, X. Li, G. Zhao, and M. Pietikainen, "Recognising spontaneous facial micro-expressions," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1449–1456.
- [4] M. Shreve, S. Godavarthy, D. Goldgof, and S. Sarkar, "Macro-and micro-expression spotting in long videos using spatio-temporal strain," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 2011, pp. 51–56.
- [5] J. A. Ruiz-Hernandez and M. Pietikainen, "Encoding local binary patterns using the re-parametrization of the second order gaussian jet," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 2013, pp. 1–6.
- [6] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, "Casmie ii: An improved spontaneous micro-expression database and the baseline evaluation," *PloS one*, vol. 9, no. 1, p. e86041, 2014.
- [7] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.
- [8] A. Goshtasby, "Image registration by local approximation methods," *Image and Vision Computing*, vol. 6, no. 4, pp. 255–261, 1988.
- [9] T. W. Sederberg and S. R. Parry, "Free-form deformation of solid geometric models," in *ACM Siggraph Computer Graphics*, vol. 20, no. 4. ACM, 1986, pp. 151–160.
- [10] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images," *Medical Imaging, IEEE Transactions on*, vol. 18, no. 8, pp. 712–721, 1999.
- [11] R. Chandrashekar, R. H. Mohiaddin, and D. Rueckert, "Analysis of 3-D myocardial motion in tagged MR images using nonrigid image registration," *Medical Imaging, IEEE Transactions on*, vol. 23, no. 10, pp. 1245–1250, 2004.
- [12] T. Papatheodorou, "3D Face Recognition Using Rigid and Non-Rigid Surface Registration," Ph.D. dissertation, University of London, 2006.
- [13] G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert, "Recognition of 3D facial expression dynamics," *Image and Vision Computing*, vol. 30, no. 10, pp. 762–773, 2012.
- [14] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [15] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 696–710, 1997.
- [16] W. Zhao, A. Krishnaswamy, R. Chellappa, D. L. Swets, and J. Weng, "Discriminant analysis of principal components for face recognition," in *Face Recognition*. Springer, 1998, pp. 73–85.
- [17] S. Lee, G. Wolberg, and S. Y. Shin, "Scattered data interpolation with multilevel B-splines," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 3, no. 3, pp. 228–244, 1997.
- [18] R. H. Bartels, J. C. Beatty, and B. A. Barsky, *An introduction to splines for use in computer graphics and geometric modeling*. Morgan Kaufmann, 1987.
- [19] C. E. Thomaz and G. A. Giralaldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image and Vision Computing*, vol. 28, no. 6, pp. 902–913, 2010.
- [20] E. Z. Tenorio and C. E. Thomaz, "Análise Multilinear Discriminante de Formas Frontais de Imagens 2D de Face," *Proceedings of the X Simposio Brasileiro de Automacao Inteligente SBAI 2011*, pp. 266–271, September 2011.
- [21] B. Tiddeman, M. Stirrat, and D. I. Perrett, "Towards realism in facial image transformation: Results of a wavelet mrf method," in *Computer Graphics Forum*, vol. 24, no. 3. Wiley Online Library, 2005, pp. 449–456.
- [22] M. Valstar, B. Martinez, X. Binefa, and M. Pantic, "Facial point detection using boosted regression and graph models," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2729–2736.