

Reconocimiento facial basado en Least-Squares Support Vector Machine y características LNDP

Pablo Augusto Campos-Zárate, Efrén Gorrostieta-Hurtado,
Jesús Carlos Pedraza-Ortega

Universidad Autónoma de Querétaro,
Facultad de Ingeniería,
México

{pacaza60, efrengorrostieta}@gmail.com,
caryoko@yahoo.com

Resumen. El reconocimiento facial en 3D se ha vuelto de gran interés para los investigadores en el campo de la visión artificial debido a las ventajas que posee contra el reconocimiento tradicional en 2D, en este trabajo se presenta un sistema de reconocimiento facial en 3D que utiliza características LNDP del rostro junto con un clasificador LSSVM que conserva las ventajas del SVM tradicional sin la necesidad de resolver un problema de programación cuadrático. Los experimentos realizados en la *Texas 3DFRD* muestran que este sistema es capaz de verificar e identificar rostros de la base de datos con valores de 99 %.

Palabras clave: Reconocimiento facial, clasificación multiclase, least-squares support vector machine, normales de superficie.

Facial Recognition Based on Least Squares Support Vector Machine and LNDP Features

Abstract. 3D face recognition has become very popular among computer vision researchers due to its advantages against traditional 2D face recognition, this paper presents a 3D face recognition LNDP+LSSVM-based system which keeps the benefits of a traditional SVM without having to solve a quadratic-programming problem but a linear-programming problem. Experiments conducted on the Texas 3DFR Database show that LNDP-LSSVM method can achieve verification and recognition rate over 99%.

Keywords: Face recognition, multiclass classification, least-squares support vector machine, surface normals.

1. Introducción

Un sistema de reconocimiento de rostros puede trabajar de dos formas, para autenticación o para identificación facial [1], la tarea del reconocimiento, que es

el tema de interés en este trabajo, consiste en dada una imagen de entrada que se compara contra todas las imágenes en una base de datos para determinar la identidad de la imagen de entrada. El reconocimiento facial ha atraído la atención de los investigadores debido a sus numerosas aplicaciones como seguridad y vigilancia, el reconocimiento facial utiliza información biométrica sin la necesidad de un contacto directo con el individuo que se está trabajando. Aunque existe mucho trabajo relacionado con el reconocimiento tradicional (en 2 dimensiones), este aún presenta retos cuando se enfrenta a variaciones de iluminación, pose y expresiones. El reconocimiento en 3D surgió como una solución a las dificultades inherentes a las variaciones de pose e iluminación debido a que aprovecha la información geométrica completa del rostro, aunque de igual manera sigue enfrentando retos cuando se trata de expresiones, variaciones de pose extremas y oclusiones. De acuerdo con [2], los algoritmos de reconocimiento pueden dividirse en 3 categorías: Métodos holísticos, métodos locales y métodos híbridos.

El enfoque holístico utiliza toda la información del rostro, los métodos basados en *Principal Component Analysis* (PCA) en [3,4] entran en dicho enfoque, aunque muchos trabajos se han realizado dentro de esta categoría con buenos resultados, aún se enfrentan dificultades frente a oclusiones y datos faltantes debido a que la mayoría de estos métodos requieren un modelo completo del rostro, por lo que los métodos basados en características locales han atraído la atención de investigadores debido a que ofrecen mayor robustez frente a oclusiones, expresiones y valores faltantes así como variaciones de pose, una revisión de los métodos del estado del arte que utilizan características locales es presentado en [5].

En [6] se presentó un método para el reconocimiento de expresiones faciales haciendo una combinación de características extraídas por PCA y LBP, logrando un híbrido de características globales y locales alimentadas después a una *Support Vector Machine* (SVM). Más trabajo con información biométrica en 3D del rostro es realizado en [7], el proceso se divide en 4, a) adquisición de datos, b) extracción de la cabeza, c) identificación del plano de la cabeza y d) clasificación de la forma de la cabeza. El proceso de la clasificación es también realizado mediante una SVM. Dentro de los avances en reconocimiento facial haciendo uso de características locales y SVM, se encuentra el propuesto en [8] que propuso una manera de extraer información facial llamada *Angular Radial Signature* (ARS), mismas características que fueron alimentadas a una SVM con un Kernel PCA obteniendo un alto desempeño tanto en reconocimiento como en eficiencia. Para lo mejor de nuestro conocimiento, no se ha combinado el descriptor LNDP con una LS-SVM para obtener un sistema de reconocimiento facial, por lo que este trabajo se enfocó en conocer las capacidades de dicho sistema.

2. Métodos

2.1. Descriptor utilizado

El descriptor *Local Derivative Pattern* (LDP) [9] está inspirado en el descriptor *Local Binary Pattern* [10] (LBP) que fue propuesto para describir texturas en

imágenes. El LDP aprovecha información direccional de cada píxel de la imagen, obteniendo así mejores características a diferencia del descriptor no direccional LBP. Posteriormente [11] propuso una variación del LDP que aprovecha la información de las normales de una superficie para extraer las características de objetos 3D.

Normales de superficie Para la estimación de los vectores normales a la superficie, dado un conjunto de puntos $P = \{p_1, p_2, \dots, p_n\}$ en una nube de puntos $p_i \in R^3$ donde $\mathbf{p}_i = [p_{ix}, p_{iy}, p_{iz}]^T$ representa las coordenadas 3D de cada punto y a cada punto le corresponde un vector normal $\mathbf{n}_i = [n_{ix}, n_{iy}, n_{iz}]^T$, se define un conjunto de k puntos vecinos para cada punto p_i , de manera que el conjunto de puntos vecinos a p_i es $Q_i = \{q_{i1}, q_{i2}, \dots, q_{ik}\}$. De esta forma la matriz aumentada de vecinos es:

$$\mathbf{Q}_i^+ = [p_i, q_{i1}, q_{i2}, \dots, q_{ik}]^T. \tag{1}$$

Para encontrar el vector n_i correspondiente a p_i , se ajusta un plano a la matriz (1) que minimice la distancia con cada punto, así el vector normal al plano, también será el vector normal al punto p_i [12]:

$$\min_{n_i} \|[\mathbf{Q}_i^+ - \bar{\mathbf{Q}}_i^+] \mathbf{n}_i\|, \tag{2}$$

donde $\bar{\mathbf{Q}}_i^+ = \mathbf{1}_k + \frac{1}{k+1}(\mathbf{p}_i + \sum_{j=1}^k \mathbf{q}_{ij})$, esto es equivalente a calcular el *Principal Component Analysis* en \mathbf{Q}_i^+ y tomar el componente principal con menor covarianza.

En [12] también se hace mención a los métodos para la obtención de los puntos vecinos a p_i .

Local Normal Derivative Pattern La teoría detrás de este descriptor es muy similar a la propuesta para el descriptor LDP, para un componente normal $N(P)$ el LNDP de primer orden en 4 direcciones distintas $\alpha = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ es [9,11] :

$$N'_{0^\circ}(p_i) = N(p_i) - N(Q_{i4}), \tag{3}$$

$$N'_{45^\circ}(p_i) = N(p_i) - N(Q_{i3}), \tag{4}$$

$$N'_{90^\circ}(p_i) = N(p_i) - N(Q_{i2}), \tag{5}$$

$$N'_{135^\circ}(p_i) = N(p_i) - N(Q_{i1}), \tag{6}$$

y el LNDP de orden n para un determinado α es como sigue:

$$LNDP_\alpha^n(p_i) = \{f(N_\alpha^{n-1}(p_i), N_\alpha^{n-1}(q_{i1})), f(N_\alpha^{n-1}(p_i), N_\alpha^{n-1}(q_{i2})), \dots, f(N_\alpha^{n-1}(p_i), N_\alpha^{n-1}(q_{ik}))\}, \tag{7}$$

donde k es el número de vecinos de p_i :

$$f(N_\alpha^{n-1}(p_i), N_\alpha^{n-1}(q_{ij})) = \begin{cases} 0, & \text{si } N_\alpha^{n-1}(q_{ij}) \cdot N_\alpha^{n-1}(p_i) > 0 \\ 1, & \text{si } N_\alpha^{n-1}(q_{ij}) \cdot N_\alpha^{n-1}(p_i) \leq 0 \end{cases} \quad j = 1, \dots, k. \quad (8)$$

Para cada píxel se concatenan los LNDP en las 4 direcciones y se forma un histograma espacial para cada componente normal:

$$LNDP^n(p_i) = \{LNDP_\alpha^n | \alpha = 0^\circ, 45^\circ, 90^\circ, 135^\circ\}. \quad (9)$$

El descriptor final se obtiene concatenando los histogramas de las tres componentes normales.

2.2. Least Squares Support Vector Machines

Las LS-SVM son una modificación de las SVM tradicionales, que fue presentada por [15], que simplifica la formulación de las SVM de Vapnik [13,14] sin perder las ventajas que estas ofrecen. De esta forma el problema de optimización se convierte en un problema de programación lineal (LP) en lugar de un problema de programación cuadrático (QP). El planteamiento del problema propuesto por Suykens consiste en:

$$\min_{w,b,e} J_p(w, e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^l e_k^2, \quad (10)$$

sujeito a

$$y_k [w^T \phi(x_k) + b] = 1 - e_k, \quad k = 1, \dots, l.$$

El Lagrangiano del problema 10 es:

$$L(w, b, e; \alpha) = J_p(w, e) - \sum_{k=1}^l \alpha_k \{y_k [w^T \phi(x_k) + b] - 1 + e_k\}, \quad (11)$$

después de obtener las derivadas parciales de 11, el clasificador se puede encontrar resolviendo el sistema de ecuaciones:

$$\begin{bmatrix} 0 & y^T \\ y & \Omega + I/\gamma, \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1_v \end{bmatrix} \quad (12)$$

donde $Z^T = [\phi(x_1)^T y_1; \dots; \phi(x_l)^T y_l]$, $y = [y_1; \dots; y_l]$, $1_v = [1; \dots; 1]$ y $\Omega = Z^T Z$. El *kernel trick* puede realizarse dentro de la matriz Ω . Así el clasificador es:

$$y(x) = \text{sign} \left[\sum_{k=1}^l \alpha_k y_k K(x, x_k) + b \right]. \quad (13)$$

Para el problema de la clasificación multiclase se utilizó en enfoque *OnevsAll* debido a que requiere menor trabajo computacional, en este se entrenan tantos

clasificadores binarios como clases distintas, si se tiene k individuos deben ser entrenados k clasificadores $y = \mathbf{w} \cdot \mathbf{x} + b$. Dada una imagen de entrada \mathbf{x} se obtienen $y_q(\mathbf{x})(q = 1, \dots, k)$, de esta forma para verificación se establece un umbral t y la muestra se acepta si $t - y_q(\mathbf{x}) \geq 0$ de lo contrario la muestra se rechaza. Para identificación la etiqueta de \mathbf{x} se asigna como:

$$etiqueta(\mathbf{x}) = \underset{1 \leq q \leq k}{argmax}(y_q(\mathbf{x})). \quad (14)$$

2.3. Metodología y experimentación

El enfoque general propuesto para un sistema basado en una LSSVM consiste en lo siguiente, primero la base de datos es dividida en un conjunto de entrenamiento y un conjunto de prueba, para cada imagen en el conjunto de entrenamiento se hace un re-escalado al 10 %, las imágenes reducidas se utilizan como entrada para el algoritmo LNDP, mismo que retorna los vectores de características LNDP para cada imagen, estos vectores son utilizados para entrenar una LSSVM, las imágenes en el conjunto de prueba pasan por el mismo proceso de re-escalación y extracción de características antes de poder ser alimentadas a la LSSVM entrenada para la predicción de las nuevas etiquetas, cada uno de los pasos aquí mencionados será detallado más adelante en esta sección.

La base de datos utilizada es la Texas 3DFRD [16,17], cuenta con modelos 2D y 3D de 118 individuos en un rango de edad de 22 a 75 años aproximadamente de distintos grupos étnicos, hay modelos para rostros con expresión neutral y expresiones faciales, dado que la base de datos presenta los modelos normalizados, este trabajo se enfocó principalmente en el trabajo del método propuesto. Las imágenes de rango originales son de 751x501 pixeles donde el píxel con mayor intensidad representa el punto más cercano entre el rostro y el dispositivo de captura. Para la experimentación cada imagen en la base de datos se re-escaló a un tamaño de 75x50 pixeles por medio de una interpolación bilineal, esto es necesario debido a la cantidad de operaciones que se requieren para el cálculo de las normales de superficie, la Figura 1 muestra el comportamiento del tiempo cuando se varía el tamaño de la imagen con escalas entre el 10 % y el 100 %, como es claro el comportamiento se asemeja a un comportamiento de un algoritmo $O(n^3)$, que es precisamente lo que se espera de este algoritmo de acuerdo con [12], por lo que se consideró un tamaño de imagen al 10 % como adecuado para poder competir con otros algoritmos en términos de tiempo de cómputo. Para las normales de superficie, se utilizó una vecindad de los 25 puntos más cercanos utilizando las implementaciones de *k-Nearest Neighbor* y PCA del software Matlab, de acuerdo con lo experimentos de [12] utilizar más de 25 puntos en la vecindad para la estimación no mejora de manera significativa la calidad de las normales.

Para la extracción de las características LNDP corresponda a características locales, cada imagen se dividió en parches de 5x5 y a cada parche se aplicó el descriptor LNDP, los histogramas de todos los parches se concatenaron para formar el vector final de características LNDP, el grado del descriptor utilizado

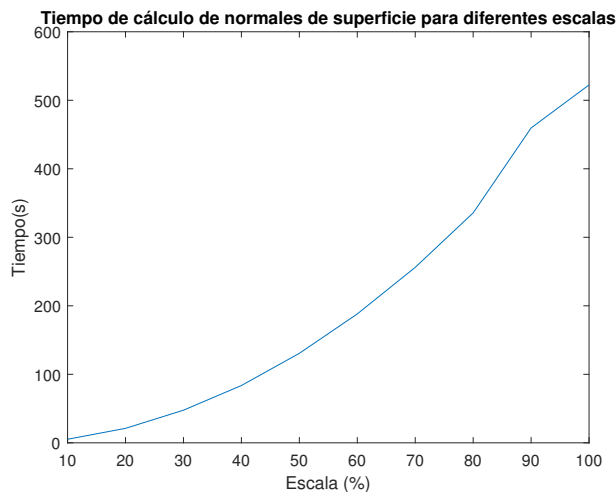


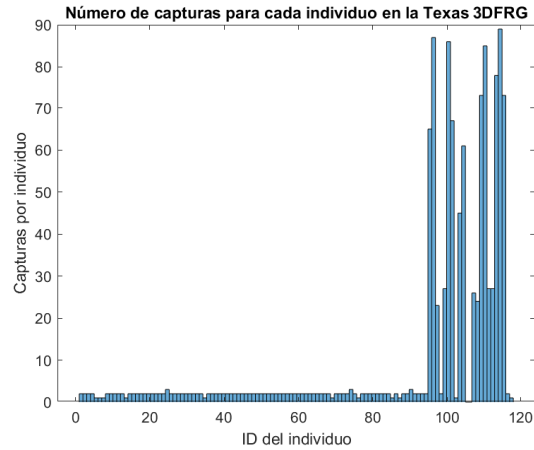
Fig. 1. Tiempo de cómputo de características LNDP para una imagen a diferentes escalas.

es $n = 2$. Por propósitos prácticos, se extrajeron las características de todas las imágenes en la base de datos y se almacenaron como en un archivo .csv mismo del que se tomaron los vectores para el entrenamiento y pruebas.

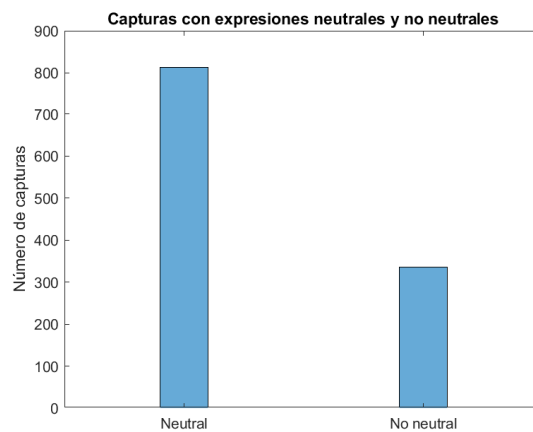
La Figura 2 muestra la cantidad de imágenes que hay por cada sujeto y la cantidad de imágenes que corresponden a expresiones faciales neutrales y no neutrales en la base de datos, como se observa en 2.a, los individuos con más de 5 capturas son pocos en comparación con los que tienen menos de 5, siendo que la mayoría tiene 2 capturas, también existe un desbalance entre las imágenes con expresiones neutrales y no neutrales, para que las comparaciones con otros algoritmos sean justas se utilizó la partición de datos propuesta en [16], el conjunto de entrenamiento consiste de 30 capturas aleatorias de 12 individuos distintos, el conjunto de prueba se dividió en 2: *gallery* y *probe*, la galería consiste de 105 imágenes neutrales de 103 individuos, la prueba contiene imágenes neutrales y no neutrales de 95 individuos distintos, el conjunto de entrenamiento se utilizó tal como está propuesto, para el conjunto de *testing* se utilizaron las imágenes de *probe* de personas que existen en el entrenamiento y por lo tanto son conocidas por el sistema. El entrenamiento de la LSSVM se realizó utilizando el paquete de LSSVMLab [18] con un kernel polinomial y una codificación 1vA.

2.4. Evaluación comparativa

Como base para comparar el desempeño del algoritmo se tomaron en cuenta los algoritmos [8] y [20], para una justa comparación, los tres algoritmos se ejecutaron en el mismo computador y con los mismos datos de entrenamiento y



(a)



(b)

Fig. 2. Histogramas de las capturas en la base de datos Texas 3DFRD. a) Número de capturas por cada individuo. b) Número de capturas que tienen expresiones neutrales y no neutrales

prueba, no se agregó ningún paso extra de normalización ni correcciones a las imágenes.

Lo primero que se verificó fue el tiempo que toma la extracción de características para cada algoritmo, la Tabla 1 muestra el tiempo que toma la extracción de características para los 3 métodos en una sola imagen, como era de esperarse el método de extracción LNDP toma el mayor tiempo dentro de los 3 debido a la cantidad de operaciones que deben realizarse para la extracción.

Tabla 1. Tiempo de extracción de características de los métodos evaluados.

Método	Tiempo (s)	Std (s)
LBP	0.0477	0.0017
ARS	0.0000536	0.00014
LNDP	5.92	0.1836

Los tiempos de entrenamiento para cada algoritmo también fueron computados, en la Tabla 2 se observa que el tiempo de entrenamiento para el algoritmo propuesto es ligeramente menor que el tiempo de [20], el tiempo para [8] aumenta debido a la transformación de los datos previa al entrenamiento.

Tabla 2. Tiempos de entrenamiento para los 3 algoritmos evaluados.

Método	Tiempo(s)
SVM+ARS [8]	1.62
SVM+LBP [20]	0.0897
Propuesto	0.0414

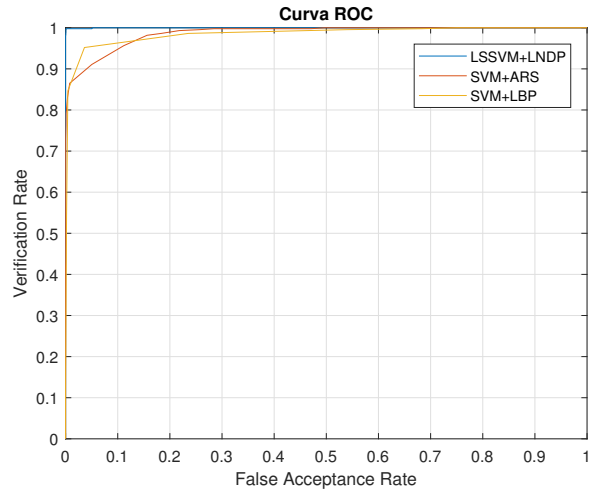
Finalmente el desempeño de los algoritmos se midió para las tareas de verificación e identificación utilizando las curvas ROC y CMC respectivamente, la Figura 3 presenta las curvas para los tres algoritmos evaluados, se observa que el método propuesto es superior, en lo que a *Verification rate* y *Recognition rate* respecta, la Tabla 3 compara los valores de VR y FAR para los 3 distintos métodos, en donde el método propuesto destaca en ambas métricas.

Tabla 3. Verification rate y Recognition rate para puntos fijos en 0.1 FAR y Rank-1

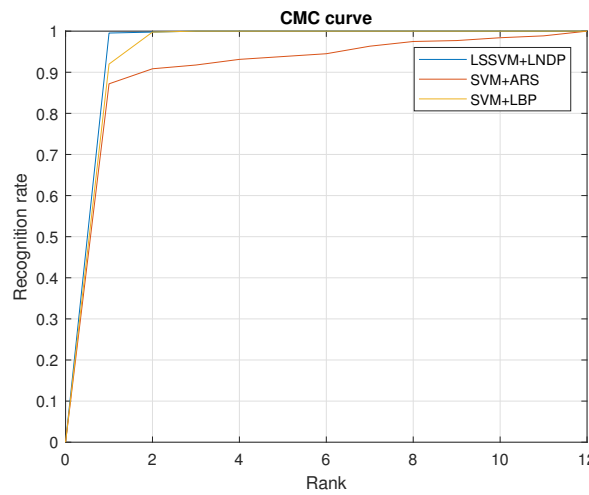
Método	VR en 0.1 FAR (%)	RR en Rank-1 (%)
SVM+ARS[8]	95.65	87.19
SVM+LBP[20]	95.8	91.99
Propuesto	99.54	99.77

3. Discusión

Con los datos utilizados para entrenamiento y pruebas con imágenes de la *Texas 3DFRD* [16,17], el método LSSVM+LNDP ofrece un alto desempeño en las tareas de identificación y verificación en comparación los otros métodos revisados, el apartado en el que este método es superado es en el tiempo de la extracción de características, donde los otros dos métodos resultaron ser altamente eficientes en comparación. El método propuesto tiene un alto desempeño cuando la base de datos contiene información con poca variabilidad en cuanto a



(a)



(b)

Fig. 3. Curvas ROC y CMC para los algoritmos evaluados. Los valores de Recognition rate en Rank-1 y Verification rate en 0.1 FAR son 99.54% y 99.77% respectivamente.

expresiones e información faltante respecta y con pocos individuos distintos a discriminar, más experimentos con diferentes datos deben realizarse para verificar el desempeño del método con datos con mayor variabilidad. En LSSVM+LNDP se pueden obtener mejores resultados en verificación e identificación a un costo computacional mayor aunque aceptable aún.

4. Conclusiones

Se ha presentado un sistema para reconocimiento facial que desempeña las tareas de verificación e identificación con desempeño del 99 % en la base de datos *Texas 3DFRD*, permitiendo discriminar entre individuos con imágenes con expresiones neutrales y no neutrales, el uso del descriptor LNDP permite obtener mejor información para discriminación que los descriptores ARS y LBP aunque a un mayor costo computacional, en los experimentos también se observó que el uso de una LSSVM requiere ligeramente menor trabajo computacional que una SVM tradicional. Se espera que este trabajo de pie a mayor investigación en algoritmos de reconocimiento facial basados en características locales y SVM's.

Acknowledgment. Este trabajo se realizó con el apoyo del Consejo Nacional de Ciencia y Tecnología.

Referencias

1. Li, S. Z., Jain, A. K.: Handbook of Face Recognition. 2nd edn. Springer-Verlag, London (2011)
2. Zhao, W., Chellappa, R., Phillips, P. J., Rosenfeld, A.: Face Recognition: A literature survey. *ACM Computing Surveys*, 35(4), 399–459 (2003)
3. Gao, Y., Leung, M. K. H.: Face recognition using line edge map. *IEEE Trasaction on Pattern Analysis and Machine Intelligence*, 24(6), 764–779 (2016)
4. Heseltine, T., Pears, N., Austin, J.: Three-dimensional face recognition: an eigensurface approach. In: 2004 International Conference on Image Processing, pp. 1421–1424, IEEE, Singapore (2004)
5. Soltanpour, S., Boufama, B., Jonathan-Wu, Q. M.: A survey of local feature methods for 3D face recognition. *Pattern Recognition*, 72, 391–406 (2017)
6. Luo, Y., Wu, C. M., Zhang, Y.: Facial expression recognition based on fusion feature of PCA and LBP with SVM. *Optik-International Journal for Light and Electron Optics*, 124(17), 2767–2770 (2013)
7. Sarakon, P., Charoenpong, T., Charoensiriwath, S. Face shape classification from 3D human data by using SVM. In: The 7th 2014 Biomedical Engineering International Conference, pp. 1–5, IEEE (2014)
8. Lei, Y., Bennamoun, M., Hayat, M., Guo, Y.: An efficient 3D face recognition approach using local geometrical signatures. *Pattern Recognition*, 47(2), 509–524 (2014)
9. Zhang, B., Gao, Y., Zhao, S., Liu, J.: Local Derivative Pattern Versus Local Binary Pattern: Face Recognition with High-Order Local Pattern Descriptor. *IEEE Transactions on Image Processing*, 19(2) (2010)
10. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and MACHine Intelligence*, 24(7), pp. 971–987 (2002)
11. Soltanpour, S., Jonatha-Wu, Q.M.: High-order local normal derivative pattern (LNDP) for 3D face recognition. In: 2017 IEEE International Conference on Image Processing, pp. 2811–2815, IEEE, Beijing, China (2017)

12. Klasing, K., Althoff, D., Wollherr, D., Buss, M.: Comparison of Surface Normal Estimation Methods for Range Sensing Applications. In: IEEE International Conference on Robotics and Automation, pp. 3206–3211, Kobe, Japan (2009)
13. Vapnik, V.: Estimation of dependencies based on empirical data. *Addendum 1* (1982)
14. Cortes, C., Vapnik, V.: Support Vector Networks. *Machine Learning*, 20, 273–297 (1995)
15. Suykens, J. A., Vandewalle, J.: Least squares support vector machine classifiers. *Neural processing letters*, 9(3), 293–300 (1999)
16. Gupta, S., Castleman, K. R., Markey, M. K., Bovik, A. C.: Texas 3D face recognition database. In: IEEE Southwest symposium on image analysis & interpretation, pp. 97–100 (2010)
17. Gupta, S., Markey, M. K., Bovik, A. C.: Anthropometric 3d face recognition. *International journal of computer vision*, 90(3), 331–349 (2010)
18. Brabanter, K., Karsmakers, P., Ojeda, F., Alzate, C., Brabanter, J., Pelkams, K., Moor, B., Vandewalle, J., Suykens, J. A. K.: LSSVMLab. URL: <https://www.esat.kuleuven.be/sista/lssvmlab/> (2011)
19. Fawcett, T.: An introduction to ROC analysis. *Pattern recognition letters*, 27(8), 861–874 (2006)
20. Shi, L., Wang, X., Shen, Y.: Research on 3D Face Recognition Method Based on LBP and SVM. *Optik*, 165157 (2020)