

Análisis de descriptores locales para la detección de menores en imágenes

Modesto Castrillón Santana, Javier Lorenzo Navarro, David Freire Obregón
modesto.castrillon@ulpgc.es

SIANI. Universidad de Las Palmas de Gran Canaria. 35017, Las Palmas de Gran Canaria.

Abstract—Child abuse material (CAM) refers to photos or videos containing the sexual abuse of a child. This kind of material involves a criminal activity, whose detection may be assisted by automatic tools. Considering a two stages approach, that determines in its first level whether the content is pornographic or not. If the first situation occurs, the second stage is triggered to confirm the presence of any non adult. This work relies on the second stage. Our approach classifies using facial information, an individual as adult or not. For that purpose, different local descriptors are analyzed, in combination with areas of interest, and fusion techniques. The experimental evaluation is carried out on a dataset designed for this particular problem containing a balanced collection of male samples within the range from 12 till 21 years old.

I. INTRODUCCIÓN

En los últimos años, el tráfico y distribución de pornografía infantil [1] ha encontrado sofisticados mecanismos para facilitar su expansión en internet. Por este motivo, los cuerpos y fuerzas de seguridad de los distintos estados, dedican recursos a su persecución a través de este medio digital, principalmente mediante la vigilancia de redes de intercambio de archivos, en particular *peer to peer*, con el objetivo de identificar a los responsables de cometer delito con material pedófilo [2].

La creciente disponibilidad de cámaras en los diversos dispositivos de la vida cotidiana (móvil, tablet, cámaras de acción como la GoPro, etc.) ha contribuido directamente a un aumento de este tipo de material en la red. Por ello se hace necesario el uso de herramientas automáticas para su análisis, con el objetivo de determinar la presencia de contenido delictivo, así como la identificación del productor del mismo. Podemos considerar una arquitectura en dos etapas, donde la primera realiza la detección de material pornográfico, con técnicas similares a las utilizadas en filtros de control parental, basados desde hace años en la detección del color piel [3], o más recientemente en la combinación de distintas evidencias como color, textura y forma [4]. La segunda etapa debe determinar si en dicho material se encuentran menores, circunstancia que debería alertar a las unidades de vigilancia.

En este trabajo nos centramos en esta segunda etapa, asumiendo por tanto que previamente se ha confirmado que la imagen analizada posee contenido pornográfico. Básicamente nuestro estudio analiza el rendimiento de la clasificación, a partir de imágenes faciales, de individuos entre adultos y no adultos, para sujetos con edades próximas a los 18 años. Con dicho objetivo se parte de información somera de la localización facial del o los individuos presentes en la imagen, determinando si se confirma la presencia de un menor, considerando como menor a los individuos con una edad inferior a los 18 años.

La edad es ciertamente uno de los atributos asociados con la biometría blanda, que si bien no permiten diferenciar unívocamente dos individuos, son fácilmente estimados por los humanos en muy diversas situaciones de iluminación y resolución, permitiendo agrupar a individuos en categorías semánticas, o grupos demográficos [5].

Las propuestas recientes de la literatura para la estimación de la edad aparente se basan mayoritariamente en información facial, analizando resultados en bases de datos como FG-NET [6], MORPH [7] o Adience [8]. Sin embargo, estas bases de datos adolecen de contener una nutrida muestra de individuos con edades distribuidas en el rango del dominio de interés de nuestro problema, dado que han sido diseñadas para otros fines, como pueden ser la recogida de datos demográficos, principalmente con objetivos interactivos o comerciales. Mencionar el trabajo descrito en [9] centrado en la estimación de la edad en niños, destacando las dificultades al analizar exclusivamente el rostro. Dada la carencia de bases de datos adecuadas, los autores recopilan imágenes faciales, adoptando una estrategia que integra información contextual en el proceso de extracción de características.

Compendios recientes relacionados con descriptores de biometría blanda pueden encontrarse en [5], [10], con resultados para edad tanto en estimación, como asociación a grupos de edad. Un nutrido grupo de técnicas requiere de información precisa de múltiples puntos fiduciales sobre el rostro. Los trabajos basados en descriptores locales son más novedosos, si bien es poco frecuente encontrar un análisis que cubra un amplio espectro y variantes. Los buenos resultados recientemente obtenidos por las redes profundas en distintos problemas de Visión por Computador [11], no han sido ajenos al contexto de la biometría blanda, por lo que la literatura ya aplica redes neuronales convolutivas (CNN) [12] en la clasificación de sexo y estimación de la edad. En este sentido destacamos el reciente trabajo [13] con una red con tres capas convolutivas y dos capas totalmente conectadas, que obtiene tasas de acierto sobre 50%.

El presente estudio analiza para este problema distintos descriptores locales, considerando una propuesta basada en apariencia, zonas de interés, y estrategias de fusión, siguiendo resultados previos de los autores en clasificación de sexo. El análisis se lleva a cabo sobre una nueva base de datos, centrada en esta exploración inicial, en el sexo masculino y en el rango de edad desde los 12 a los 21 años, cuyos detalles se incluyen en la sección II.

II. CONJUNTO DE DATOS

Al contrario que otras bases de datos que incluyen anotaciones de la edad real o aparente en un amplio rango de edades,



Fig. 1. Imagen original, con su correspondiente imagen normalizada, de dimensiones 159×155 píxeles, fijando la posición de los ojos en las localizaciones (66,62) y (92,62), con una distancia interocular de 26 píxeles, tras realizar una rotación, escalado y traslación.

la base de datos Boys2Men, contiene unas 1000 imágenes obtenidas a través de buscadores en la red, repartidas de forma equilibrada para las edades comprendidas entre los 12 y los 21 años. En esta primera evaluación las imágenes se corresponden sólo con sujetos varones.

Para cada imagen, además de la información de la edad, se proporciona una posición, en ocasiones aproximada, de la posición de los ojos que ha sido obtenida haciendo uso de detectores automáticos de ojos [14], o de anotación manual en el caso de no obtenerse detección automática (principalmente caras rotadas).

En los experimentos presentados en la sección IV, se realiza una normalización simple del rostro en base a la localización de los ojos, realizando una rotación, escalado y traslación, obteniendo como ilustra la Fig. 1, la imagen normalizada a partir de su original. La imagen normalizada resultante presenta una distancia interocular de 26 píxeles, incluyendo también la zona del torso superior y la cabeza. Es el patrón que denominamos cabeza y hombros (HS por sus siglas en inglés).



Fig. 2. Imagen ejemplo de Boys2Men, ilustrando las regiones de interés utilizadas en los experimentos. De izquierda a derecha: cabeza y hombros (HS) (64×64 píxeles), cara (F) (59×65 píxeles), periocular (P) (49×19 píxeles), y boca (M) (37×31 píxeles).

III. REPRESENTACIÓN Y CLASIFICACIÓN

Inspirados en nuestros resultados recientes en clasificación de sexo [15], abordamos el problema de estimación de edad, analizando el rendimiento de diversos descriptores locales, regiones de interés (ver Fig. 2), y su fusión en una arquitectura en dos etapas, donde la segunda etapa fusiona a nivel de valor los resultados de los clasificadores que conforman la primera etapa. El objetivo es seleccionar, de entre todas las combinaciones posibles, un subconjunto de clasificadores para esa primera etapa, que permitan obtener una buena tasa de clasificación.

A. Descriptores

Describimos brevemente el conjunto de descriptores utilizados en el estudio, que se analizarán mediante una representación a través de histogramas, haciendo uso de una rejilla de celdas para reducir la pérdida de información espacial [16].

Cada clasificador de la primera etapa tiene distintos parámetros como son el operador a utilizar, la zona de interés sobre la que se aplica, y el tamaño de rejilla aplicado para el operador sobre la imagen normalizada de la zona de interés. La rejilla se define en base a un número de celdas horizontales y verticales, respectivamente cx y cy , donde $1 \leq cx \leq 5$ y $1 \leq cy \leq 5$. De este forma, el patrón analizado se subdivide en $cx \times cy$ celdas. La aplicación de un descriptor u operador en cada celda proporciona un histograma, h_i , donde cada valor representado indica el número de apariciones de un código en la celda. Un patrón se caracteriza por la concatenación de los $cx \times cy$ histogramas, así pues el vector de características de la imagen es $\mathbf{x} = \{h_1, h_2, \dots, h_{cx \times cy}\}$.

Como descriptores hemos considerado:

- Histograma de gradientes orientados (Histogram of Oriented Gradients HOG) [17]. Calcula y discretiza las orientaciones del gradiente en cada celda.
- Patrones binarios locales (Local Binary Patterns LBP) y su variante uniforme (LBP^{u2}) [16]. Conocido operador de textura, codifica cada píxel en base a su relación de magnitud con los vecinos, componiendo un código binario. LBP^{u2} reduce el diccionario a los códigos más habituales en imágenes de texturas.
- Patrones de gradiente local (Local Gradient Patterns LGP) [18]. Considera el gradiente con la vecindad para obtener el código del píxel.
- Patrones ternarios locales (Local Ternary Patterns LTP) [19]. Al contrario de los LBP, considera tres posibles relaciones con cada vecino, obteniendo un código ternario, que habitualmente se divide en parte inferior y superior, respectivamente LTP_{low} y LTP_{high} .
- Patrones salientes locales (Local Salient Patterns LSP) [20]. En lugar de analizar la diferencia con cada píxel de la vecindad, observa los vecinos con mayor y menor diferencia. Evaluamos cinco variantes: LSP_0 , LSP_1 , LSP_2 , LSP_{01} y LSP_{012} .
- Descriptor de Weber (Weber Local Descriptor WLD) [21]. Basado en la Ley de Weber, la cual establece que la percepción de un estímulo depende no sólo del estímulo sino también de su intensidad original.
- Cuantización de fase local (Local Phase Quantization LPQ) [22]. No afectada por cierto tipo de ruido, se calcula a partir de la transformada de Fourier breve (STFT) de los vecinos.
- Patrones binarios locales basados en intensidad (Intensity based Local Binary Patterns NILBP) [23]. Considera el valor de la media en lugar del valor del píxel central de la vecindad.
- Mejorador basado en información estadística de orientación local (Local Oriented Statistics Information Booster LOSIB) [24]. Basado en LBP, calcula la infracción estadística de orientación local en toda la celda.

B. Fusión

Como hemos mencionado anteriormente, se adopta una arquitectura en dos niveles, donde el primer nivel está conformado por clasificadores SVM con kernel RBF [25]. Cada clasificador produce como salida un valor, indicando la proximidad a la frontera entre las clases adulto y no adulto. El segundo nivel realiza la fusión a nivel de valoración (*score*),

TABLA I

RESUMEN DE RESULTADOS OBTENIDOS PARA CADA PATRÓN O ZONA DE INTERÉS Y DESCRIPTOR PAAR EL PRIMER SUBCONJUNTO. SE ANALIZAN REJILLAS DESDE 1×1 HASTA 5×5 , ES DECIR, 25 COMBINACIONES MOSTRANDO LA REJILLA QUE MEJOR RESULTADO HA PROPORCIONADO, INCLUYENDO EL NÚMERO DE CARACTERÍSTICAS QUE REQUIERE Y LA TASA DE ACIERTO ALCANZADA. DESTACAMOS LAS CONFIGURACIONES CON TASAS SOBRE 69%.

Descriptor	F		HS	
	# (grid)	Acc.	# (grid)	Acc.
HOG	225 (5 × 5)	68.2	180 (4 × 5)	64.7
LBP ^{u2}	531 (3 × 3)	68.7	531 (3 × 3)	69.7
LBP	256 (1 × 1)	61.6	768 (3 × 1)	68.2
LGP	6400 (5 × 5)	67.6	5120 (4 × 5)	64.1
LPQ	768 (1 × 3)	68.2	768 (3 × 1)	65.2
WLD	1280 (1 × 5)	66.7	768 (1 × 3)	63.6
LOSIB	160 (4 × 5)	69.2	128 (4 × 4)	59.6
NILBP	885 (3 × 5)	67.2	885 (3 × 5)	66.7
LSP ₀	684 (4 × 3)	69.2	684 (3 × 4)	67.7
LSP ₁	684 (4 × 3)	68.2	855 (3 × 5)	64.7
LSP ₂	855 (3 × 5)	65.2	684 (3 × 4)	65.2
LSP ₀₁	570 (1 × 5)	69.2	1026 (3 × 3)	65.7
LSP ₀₁₂	855 (3 × 5)	66.2	518 (3 × 1)	65.2
LTP _{high}	512 (2 × 1)	63.1	768 (1 × 3)	63.6
LTP _{low}	512 (1 × 2)	63.6	768 (3 × 1)	63.2

	P		M	
	# (grid)	Acc.	# (grid)	Acc.
HOG	180 (5 × 4)	69.7	225 (5 × 5)	62.6
LBP ^{u2}	885 (5 × 3)	67.7	590 (5 × 2)	66.7
LBP	1536 (3 × 2)	64.1	768 (3 × 1)	65.6
LGP	5120 (5 × 4)	69.2	3072 (4 × 3)	63.1
LPQ	512 (2 × 1)	66.2	512 (1 × 2)	65.7
WLD	1536 (2 × 3)	69.7	1536 (3 × 2)	68.2
LOSIB	120 (3 × 5)	67.7	160 (4 × 5)	61.1
NILBP	708 (4 × 3)	67.7	590 (2 × 5)	68.7
LSP ₀	513 (3 × 3)	67.2	228 (2 × 2)	67.2
LSP ₁	855 (5 × 3)	65.7	456 (4 × 2)	64.7
LSP ₂	855 (5 × 3)	69.2	456 (4 × 2)	63.7
LSP ₀₁	456 (2 × 2)	68.7	228 (2 × 1)	67.7
LSP ₀₁₂	1026 (2 × 3)	66.2	342 (1 × 2)	65.2
LTP _{high}	3072 (3 × 4)	68.7	1024 (1 × 4)	64.6
LTP _{low}	1024 (1 × 4)	66.2	1280 (1 × 5)	63.6

por tanto a partir de dichos valores. Esa opción ha sido preferida frente a la fusión de características, por permitir el cálculo paralelo, y reducir la dimensionalidad del problema.

IV. RESULTADOS

Para el conjunto de datos hemos definido un experimento de validación cruzada con 5 subconjuntos, con aproximadamente 800 muestras de entreno y 200 de test por subconjunto. Las siguientes subsecciones presentan en primer lugar resultados considerando un único descriptor y área de interés, posteriormente se aborda la fusión de hasta tres descriptores por área de interés, para finalizar con la fusión de varios descriptores, sin limitar el número de áreas de interés consideradas.

A. Descriptores individuales

La Tabla I muestra las mejores tasas de clasificación obtenidas para cada descriptor, 15 variantes, y zona de interés. Por simplicidad no se incluyen el resto de combinaciones, que suponen 25 por descriptor y área de interés.

Las mejores tasas de clasificación superan el 69% para todas las zonas de interés, excepto si se trabaja con la boca (M), donde se alcanzan valores próximos al 69% con NILBP. Si bien las diferencias no son excesivas, puede observarse que

TABLA II

RESUMEN DE RESULTADOS MOSTRANDO LOS MEJORES 3 RESULTADOS POR ÁREA DE INTERÉS.

Pattern	Acc.	Descriptors
F	73.7	LGP + LSP ₀ + LSP ₁
	73.2	LBP ^{u2} + NILBP + LSP ₀₁
	73.2	LBP ^{u2} + LPQ + LOSIB
HS	72.7	HOG + LSP ₂ + LTP _{high}
	72.2	WLD + LSP ₂ + LTP _{high}
	71.2	LSP ₀₁ + LSP ₀₁₂ + LTP _{high}
P	73.2	WLD + LSP ₀₁ + LTP _{high}
	73.2	LGP + WLD + LSP ₂
	72.2	WLD + LSP ₂ + LTP _{low}
M	72.7	LPQ + NILBP + LSP ₀₁₂
	72.7	LPQ + NILBP + LSP ₀
	71.7	LPQ + NILBP
F+HS+P+M	77.8	F-LSP ₀ + HS-LSP ₀ + P-LPQ
	77.2	P-WLD + P-LTP _{high} + M-LBP ^{u2}
	77.2	F-LSP ₀ + HS-LSP ₀ + P-LGP + P-LPQ + P-LTP _{high} + M-HOG + M-LBP ^{u2}
	77.2	F-LSP ₀ + HS-LSP ₀ + P-WLD + M-HOG + M-LBP ^{u2}

no son los mismos descriptores los que proporcionan mejores resultados de clasificación para todas las zonas. De esta forma para el patrón F destacan LOSIB, LSP₀ y LSP₀₁; para HS lo hace LBP^{u2}; y para P lideran HOG, LGP, WLD y LSP₂.

B. Fusión de múltiples descriptores por zona de interés

En un segundo experimento, hemos analizado la fusión de varios descriptores aplicados sobre un mismo patrón. De esta forma el primer nivel lo componen hasta un máximo de tres clasificadores, que se fusionan en el segundo nivel. Hemos limitado el número de descriptores a fusionar a un máximo de tres, para tener evidencias de mejoras rápidamente sin necesidad de evaluar las 2¹⁵ combinaciones posibles. La primera parte de la Tabla II resume los tres mejores resultados alcanzados para cada patrón, indicando los descriptores utilizados en la primera etapa. Se observa una mejora que ronda los 4 puntos porcentuales con respecto a los mejores resultados obtenidos con un único descriptor en la Tabla I. Lidera el patrón F con 73.7%, seguido de P con 73.2%, con HS y M empatados a 72.7%.

C. Fusión de múltiples descriptores y zonas de interés

El experimento final considera la combinación en la primera etapa de descriptores aplicados sobre distintos patrones. Dada la extensión del espacio de búsqueda, de nuevo se limita a considerar un máximo de tres descriptores por patrón, reduciendo las combinaciones a los descriptores que mejores resultados han proporcionado en las subsecciones precedentes. La segunda parte de la Tabla II resume los tres mejores resultados, donde se alcanza la tasa de acierto 77.8% combinando F-LSP₀ + HS-LSP₀ + P-LPQ + P-WLD + P-LTP_{high} + M-LBP^{u2}, lo cual supone un incremento de 8 puntos porcentuales comparado con la Tabla I.

D. Discusión

La distribución de errores por edad se muestra en la Fig. 3. En ella se observa un mayor número de errores entre los no adultos, es decir los menores de 18 años. En términos globales la combinación de todas las zonas de interés del rostro reduce

el número de errores, si bien no es la mejor para todas las edades. Como ilustración, se incluyen muestras de menores, de 12 y 13 años, erróneamente clasificadas como adultos en la Fig. 4. Ambas evidencias sugieren, incluso en el presente estudio reducido a varones, las dificultades aún existentes en el problema.

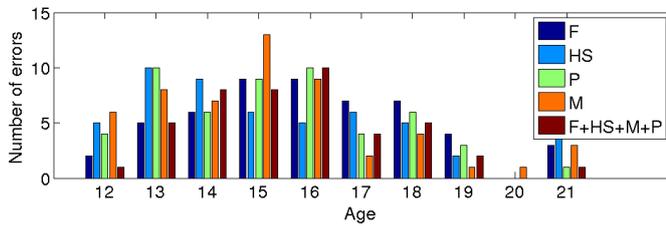


Fig. 3. Distribución de errores por edad para las distintas variantes fusionando varios descriptores y/o regiones del rostro.



Fig. 4. Muestras clasificadas como adultos por el clasificador basado en F+HS+P+M, con una edad anotada de 13 años.

V. CONCLUSIONES

El trabajo presenta resultados de clasificación entre adulto o no adulto a partir de información facial en una imagen. Surge como necesidad para asistir en la detección de presencia de menores en contenido pornográfico, sirviendo de evidencia para perseguir la comisión de delitos de abuso infantil.

Con este objetivo se ha recopilado una base de datos de varones comprendidos en el rango de edades entre 12 y 21 años, analizando distintos descriptores locales y zonas de interés, para observar su comportamiento y efectividad actuando de forma individual o fusionando evidencias. Los resultados iniciales obtenidos muestran una mejora de hasta 8 puntos porcentuales al realizar la fusión de múltiples descriptores y zonas de interés, lo cual evidencia que la información del rostro debe ser analizada a distintas resoluciones para obtener mejores sistemas de análisis y estimación de la edad. Sin embargo, estos resultados prometedores se han obtenido en una primera evaluación experimental con un conjunto de datos relativamente reducido que incluye sólo sujetos varones. El trabajo futuro requiere la recopilación de un conjunto de datos mayor sin restricciones de sexo.

AGRADECIMIENTOS

Trabajo financiado parcialmente por el proyecto TIN2015 64395-R del Ministerio de Economía y Competitividad, el Instituto Universitario de Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería (SIANI) y el Departamento de Informática y Sistemas de la ULPGC.

REFERENCIAS

[1] N. Kawale and S. Patil, "An approach to maintain the storage of contentious image in the form of descriptor," in *IEEE International Conference on Computational Intelligence and Computing Research (ICIC)*, 2014.

[2] G. Thompson, "Automatic detection of child pornography," in *Australian Digital Forensics Conference*, 2009.

[3] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," Cambridge Research Laboratory, Technical Report Series CRL 98/11, December 1998.

[4] S. H. Sengamedu, S. Sanyal, and S. Satish, "Detection of pornographic content in internet images," in *proceedings of the 19th ACM international conference on Multimedia*, 2011.

[5] M. Nixon, P. Correia, K. Nasrollahi, T. Moeslund, A. Hadid, and M. Tistarelli, "On soft biometrics," *Pattern Recognition Letters*, vol. 68, Part 2, pp. 218–230, 15 December 2015.

[6] F. project, "The FG-NET aging database," <http://www.fgnet.rsunit.com/>.

[7] K. J. Ricanek and T. Tesafaye, "MORPH: A longitudinal image database of normal adult age-progression," in *IEEE 7th International Conference on Automatic Face and Gesture Recognition (FG)*, Southampton, UK, April 2006, pp. 341–345.

[8] E. Eiding, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *Transactions on Information Forensics and Security, Special issue on Facial Biometrics in the Wild*, vol. 9, no. 12, pp. 2170–2179, December 2014.

[9] R. Satta, J. Galbally, and L. Beslay, "Children gender recognition under unconstrained conditions based on contextual information," in *22nd IEEE International Conference on Pattern Recognition (ICPR)*, Stockholm, Sweden, 2014.

[10] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometrics data reveal? A survey on soft biometrics," *IEEE Transactions on Information Forensics And Security*, vol. 11, pp. 441–467, 2016.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[12] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, 1998, pp. 2278 – 2324.

[13] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *IEEE Workshop on Analysis and Modeling of Faces and Gestures (AMFG), at the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, June 2015, pp. 34–42.

[14] M. Castrillón, O. Déniz, D. Hernández, and J. Lorenzo, "A comparison of face and facial feature detectors based on the violajones general object detection framework," *Machine Vision and Applications*, vol. 22, no. 3, pp. 481–494, 2011.

[15] M. Castrillón-Santana, J. Lorenzo-Navarro, and E. Ramón-Balmaseda, "Fusion of holistic and part based features for gender classification in the wild," in *New Trends in Image Analysis and Processing-ICIAP 2015 Workshops*. Springer International Publishing, 2015, pp. 43–50.

[16] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2044, December 2006.

[17] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *International Conference on Computer Vision & Pattern Recognition (CVPR)*, C. Schmid, S. Soatto, and C. Tomasi, Eds., vol. 2, June 2005, pp. 886–893.

[18] B. Jun and D. Kim., "Robust face detection using local gradient patterns and evidence accumulation," *Pattern Recognition*, vol. 45, no. 9, pp. 3304–3316, 2012.

[19] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1635 – 1650, 2010.

[20] Z. Chai, Z. Sun, T. Tan, and H. Mendez-Vazquez, "Local salient patterns - a novel local descriptor for face recognition," in *International Conference on Biometrics (ICB)*, 2013.

[21] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao, "WLD: A robust local image descriptor," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1705–1720, 2010.

[22] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *Image and Signal Processing, LNCS 5099*, A. Elmoataz, O. Lezoray, F. Nouboud, and D. Mammass, Eds. Springer, 2008, pp. 236–243.

[23] L. Liu, P. Fieguth, L. Zhao, Y. Long, and G. Kuang, "Extended local binary patterns for texture classification," *Image and Vision Computing*, vol. 30, no. 2, pp. 86–99, 2012.

[24] O. García-Olalla, E. Alegre, L. Fernández-Roble, and V. González-Castro, "Local oriented statistics information booster (LOSIB) for texture classification," in *International Conference on Pattern Recognition (ICPR)*, 2014.

[25] V. Vapnik, *The nature of statistical learning theory*. New York: Springer, 1995.