

APPENDIX C

Google v Cellular South

Appendix C. Materials Considered 5/8/2025 version

Exhibit	Document
1006	Kouzani, A. Z., He, F., & Sammut, K. (1997, October). Fractal face representation and recognition. In <i>1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation</i> (Vol. 2, pp. 1609-1613). (“Kouzani”)
1006A	Copy of <i>Kouzani</i> , obtained from IEEE.
1006B	Bibliographic and MARC records for <i>IEEE International Conference on Systems, Man, and Cybernetics, Conference Proceedings</i> , available at https://lccn.loc.gov/93657490 from the online catalog of the Library of Congress.
1006C	LOCIS: Library of Congress Information System, archived by Internet Archive on May 24, 1997, available from https://web.archive.org/web/19970524160715/http://library.usask.ca/hytelnet/us3/us373.html ; and Profile of the Library of Congress, available at https://librarytechnology.org/library/2263 from libraries.org, a directory of libraries throughout the world.
1006D	Bibliographic and MARC records for <i>1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation</i> that contains <i>Kouzani</i> , available at https://catalog.princeton.edu/catalog/9911918883506421 from the online catalog of the Princeton University Library.
1006E	Profile of Princeton University Library, available at https://librarytechnology.org/library/2168 From libraries.org, a directory of libraries throughout the world.
1006F	Metadata record for <i>Kouzani</i> , available at https://ieeexplore.ieee.org/abstract/document/638231 from IEEE Xplore.
1006G	Citations to <i>Kouzani</i> , obtained from Google Scholar.

1007	Yang, M. H., Kriegman, D. J., & Ahuja, N. (2002, January). Detecting faces in images: A survey. <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> , 24(1), 34-58. (“Yang”)
1007A	Copy of Yang, obtained from IEEE.
1007B	Bibliographic and MARC records for <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> that contains Yang, available at https://lccn.loc.gov/80640328 from the online catalog of the Library of Congress.
1007C	Profile of the Library of Congress, available at https://librarytechnology.org/library/2263 from libraries.org, a directory of libraries throughout the world.
1007D	Webpage of <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> , January 2002 (Vol. 24, No. 1), archived by Internet Archive on March 2, 2002, and available at https://web.archive.org/web/20020302204214/http://computer.org/tpami/tp2002/i1toc.htm
1007E	Copyright registration record for <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> that contains Yang, available at https://cocatalog.loc.gov/cgi-bin/Pwebrecon.cgi?v1=15&ti=1,15&SEQ=20250506120806&Search%5FArg=IEEE%20Transactions%20on%20Pattern%20Analysis%20and%20Machine%20Intelligence&Search%5FCode=TALL&CNT=25&PID=diZ12SP1GThHhc1ZR2rMXqpA-vkBn&SID=2 from the public catalog of the US Copyright Office.
1007F	Citations to Yang, obtained from Google Scholar.
1008	Romdhani, S., Blanz, V., & Vetter, T. (2002, May). Face identification by fitting a 3d morphable model using linear shape and texture error functions. In <i>Computer Vision—ECCV 2002: 7th European Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part IV</i> , pp. 3-19. Springer Berlin Heidelberg. (“Romdhani”)

1008A	Copy of <i>Romdhani</i> , obtained from the publisher, available at https://link.springer.com/content/pdf/10.1007/3-540-47979-1_1.pdf
1008B	Bibliographic and MARC records for <i>Computer Vision-ECCV 2001</i> that contains <i>Romdhani</i> , available at https://lccn.loc.gov/2002070542 from the online catalog of the Library of Congress.
1008C	Profile of the Library of Congress, available at https://librarytechnology.org/library/2263 from libraries.org, a directory of libraries throughout the world.
1008D	Citations to <i>Romdhani</i> , obtained from Google Scholar.
1011	Lubbers, P., Albers, B., & Salim, F. (2010). <i>Pro HTML5 Programming: Powerful APIs for Richer Internet Application Development</i> . New York: Apress ; Springer. (" <i>Lubbers</i> ")
1011A	Sample pages from a used copy of <i>Lubbers</i> .
1011B	Bibliographic and MARC records for <i>Lubbers</i> , available at https://uno.ent.sirsi.net/client/en_US/uno/search/detailnonmodal/ent:\$002f\$002fSD_UNO\$002f0\$002fSD_UNO:2320192/one from the online catalog of the University of New Orleans Library.
1011C	Profile of the University of New Orleans Library, available at https://librarytechnology.org/library/2240 from libraries.org, a directory of libraries throughout the world.
1011D	Apress webpage of Web Development titles, archived by Internet Archive on September 17, 2010, available at https://web.archive.org/web/20100917083011/http://apress.com/book/catalog?category=72
1011E	Apress website for <i>Lubbers</i> , archived by Internet Archive on September 17, 2010, available at https://web.archive.org/web/20100917083959/http://apress.com/book/view/1430227907

1011F	Amazon webpage for <i>Lubbers</i> , archived by Internet Archive on September 26, 2010, available at https://web.archive.org/web/20100926221239/http://www.amazon.com/Pro-HTML5-Programming-Application-Development/dp/1430227907
1011G	Citations to <i>Lubbers</i> , obtained from Google Scholar.
1013	Excerpts of Van Vliet, J., & Paganelli, F. (2011). <i>Programming Amazon EC2</i> . 1st ed. Sebastopol, Calif. : O'Reilly. (" <i>Van Vliet</i> ")
1013A	Sample pages from a used copy of <i>Van Vliet</i> .
1013B	Bibliographic and MARC records for <i>Van Vliet</i> , available at https://lccn.loc.gov/2011377697 from the online catalog of the Library of Congress.
1013C	Profile of the Library of Congress, available at https://librarytechnology.org/library/2263 from libraries.org, a directory of libraries throughout the world.
1013D	Copyright registration record for <i>Van Vliet</i> , available at https://cocatalog.loc.gov/cgi-bin/Pwebrecon.cgi?v1=2&ti=1,2&Search%5FArg=Programming%20Amazon%20EC2%2E&Search%5FCode=TALL&CNT=25&PID=ofVjkhDe2ydIqZNceFtUi5BI8r6J9&SEQ=20250506130023&SID=1 from the public catalog of the US Copyright Office.
1013E	Publisher webpage for <i>Van Vliet</i> , archived by Internet Archive on March 21, 2011, available at https://web.archive.org/web/20110321181132/http://oreilly.com/catalog/9781449393687/#product-details
1013F	Citations to <i>Van Vliet</i> , obtained from Google Scholar.
1014	Tang, X., & Qu, C. (2010). Facial image recognition based on fractal image encoding. <i>Bell Labs Technical Journal</i> , 15(1), 209-214. (" <i>Tang</i> ")

1014A	Copy of <i>Tang</i> , obtained from IEEE.
1014B	Bibliographic and MARC records for <i>Bell Labs Technical Journal</i> , available at https://catalog.lindahall.org/permalink/01LINDAHALL_INST/19lda7s/alma992781963405961 from the online catalog of the Linda Hall Library.
1014C	Profile of the Linda Hall Library, available at https://librarytechnology.org/library/2294 from libraries.org, a directory of libraries throughout the world.
1014D	Copyright registration record for <i>Bell Labs Technical Journal</i> , vol. 15, no. 1, available at https://cocatalog.loc.gov/cgi-bin/Pwebrecon.cgi?v1=20&ti=1,20&Search%5FArg=bell%20labs%20technical%20journal&Search%5FCode=TALL&CNT=25&PID=WDMJ5HdJVHMTWjCmLG3t7ppCoxtCL&SEQ=20250506130827&SID=4 from the public catalog of the US Copyright Office.
1014E	Webpage for <i>Bell Labs Technical Journal</i> , vol. 15, Issue 1 (June 2010), archived by Internet Archive on November 27, 2010, available at https://web.archive.org/web/20101127224126/http://onlinelibrary.wiley.com/doi/10.1002/bltj.v15:1/issuetoc
1014F	Citations to <i>Tang</i> , obtained from Google Scholar.
Appendix A	Hsieh-Yee CV
Appendix B	Parts VII to X of <i>Understanding MARC Bibliographic</i> , available at https://www.loc.gov/marc/umb/um07to10.html from the Library of Congress.

EXHIBIT 1006A

Fractal Face Representation and Recognition

A.Z. Kouzani, F. He, and K. Sammut
School of Engineering, Flinders University of SA
GPO Box 2100, Adelaide
SA 5001, Australia
e-mail: egazk, egfh, egks@eng.flinders.edu.au

Abstract— This paper presents a face representation and recognition scheme based on the theory of fractals. Each face image is represented by its fractal model which is a small collection of transformation parameters. The transformation is carried out once for known face images. For recognition, the input face image is transformed and its fractal model is then compared against the database of fractal models of known faces. Feedforward neural networks are utilised to implement the compression and recognition parts. Some experimental results are presented. The maximum compression ratio obtained for the successful recognition of known faces was observed to be 89:1 (for a compression threshold of 0.002).

I. INTRODUCTION

A fractal is a fragmented geometric shape that can be subdivided in parts, each of which is a reduced-size copy of the whole. Fractals model many real-world objects that do not correspond to simple geometric shapes, e.g. clouds, mountains, turbulence, and coastlines [9]. Fractals are generally self-similar and independent of scale.

Iterated function systems (IFSs) are linear fractal models that can describe complex objects using self-reference. An IFS consists of a set of maps from $\mathbb{R}^n$ to itself. If the maps of an IFS are contractive, then the IFS converges to a unique fixed point called an *attractor*. *Partitioned iterated function systems* (PIFSs) are generalisations of IFSs. In PIFSs, the domains of the maps are restricted. The domain of a PIFS is a subset of the transformation space, not the whole space.

Fractal theory has recently found applications in image processing. Particularly, a part of the fractal approach, which is based on the theory of PIFSs, has received a great deal of attention in image compression [1][2][4][5]. This method assumes that image redundancy can be efficiently exploited through self-similarity on a blockwise basis. The basic idea is to approximate an image by the attractor of a set of PIFSs. The original image is therefore represented by copies of properly transformed parts of itself.

Important advances in face representation and recognition have employed forms of principal com-

ponent analysis (PCA) which is an expansion of the Karhunen-Loève representation [3], [6], [10]-[11], [12]-[13]. In PCA, the optimal basis is given by the eigenvectors of the correlation matrix. However, the algorithm to calculate the eigenvectors of a correlation matrix has cubic complexity. Even in a simplified case [12], if more known faces are added to the database, then (i) the calculation of the basis becomes computationally expensive, (ii) the compression ratio declines, (iii) the basis must be recalculated, and (iv) all faces within the database must be re-compressed using the new basis.

This paper presents a face representation and recognition model based on the theory of fractals. Each input face image is represented by its fractal model which is a small collection of contractive transformation parameters. An input face image is partitioned into non-overlapping smaller blocks (*range* blocks) and overlapping larger blocks (*domain* blocks). For each range block, a search is done through the domain block pool to find a domain block whose contractive affine transformation best approximates the range block. Therefore, for each range block the parameters of the related transformations are stored. The union of these parameters represents the compressed image of a person's face. The input to the face recognition system are centred face images detected from an input image by a facial detection system [7]-[8]. The transformation is carried out once for all the face images in the database. For recognition, the input face image is transformed and its fractal model is then compared against the database of fractal models of known faces. Feedforward neural networks are employed to implement the proposed face representation and recognition system.

In Section II, a computational model of face recognition is described. The realisation method is then explained in Section III. In Section IV, simulation results are presented. The performance of the system is then discussed and compared with the performance obtained for PCA. Finally, concluding remarks are given in Section V.

II. REPRESENTATION AND RECOGNITION OF HUMAN FACE IMAGES

Fractal compression performs a vector quantisation of image blocks. The vector codebook is constructed from locally averaged and subsampled larger blocks of the image. The codebook is used for coding constant regions and edges of the image. The compression is done through contractive maps from the plane to itself. The coded image is the fixed point of the maps. The images are stored by saving the parameters of the maps and are decoded by iterating the maps to find the fixed point. In this section, the computational model of the face representation and recognition is explained. The model is based on the theory of IFSs.

A. Image Representation

An image can be represented as a collection of discrete pixels. Each pixel takes a discrete value (typically between 0 to 255) representing a gray level. A function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ can be thought of as an image of infinite resolution. Because real images are finite in extent, and for simplicity, the domain of f is assumed to be the square $(x, y) \in \{(u, v) : -1 \leq u \leq 1, -1 \leq v \leq 1\}$ and the range to be $f(x, y) \in [0, 1]$. The function $f(x, y)$ gives the gray level z at each point (x, y) .

B. Contractive Transformations

A transform w of the form

$$w \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} a_i & b_i & 0 \\ c_i & d_i & 0 \\ 0 & 0 & s_i \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} e_i \\ f_i \\ o_i \end{bmatrix} \quad (1)$$

takes an input image and produces an output image. The transform w can scale, stretch, skew, and rotate. Combinations of these transformations can map in a variety of ways. A transformation is *contractive* if it brings points closer together. Figure 1 illustrates a contractive transformation of a face image in the plane.

C. Iterated Function Systems

An IFS consists of a set of maps $\{w_i\}_{i=1}^N$ from $\mathbb{R}^n$ to itself. If the maps of an IFS are contractive, then the IFS converges to a unique fixed point called an attractor. Given any initial nonempty bounded input image, the attractor is obtained after a finite number of iterations. Figure 2 shows different iterations of an IFS with three maps for an initial input face image. The final image represents the attractor of the IFS.

D. Partitioned Iterated Function Systems

PIFSs are generalisations of IFSs. In PIFSs, the domains of w_i are restricted. The domain of a PIFS is a subset of the transformation space, not the whole space. However, the transformations w_i are still of the form given in equation 1. Figure 3 displays two PIFSs with their domains.

E. Fractal Model of Face Images

The fractal model of a face image is a collection of maps w_i s with $W = \bigcap_{i=1}^N w_i$. To achieve a good approximation of a face image, it is required that the face image f be close to the attractor of W . In order to obtain a fractal model of a face image, the image is partitioned into non-overlapping smaller blocks (range blocks) R_i and overlapping larger blocks (domain blocks) D_i . A domain pool is prepared from the available domain blocks. For each range block, a search is done through the domain pool to find a domain block whose contractive transformation best approximates the range block. A distance metric such as *root mean square* (RMS) can find the approximation error. A *threshold* value can be considered so that when the calculated error falls below the threshold, the related domain block is accepted as the reference block for the current range block. Therefore, for each range block the parameters of the related transformations are stored. The union of these parameters represent the compressed image of a person's face. Since the aim of this work is to recognise a face image rather than to reconstruct it, a simplification is applied to the compression process to reduce the number of parameters needed to model a face image. Hence, for each map w , only the following parameters e_i and f_i which specify the position of the related domain block are stored. The domain blocks' size is chosen to be twice that of the range blocks. Figure 4 displays the mapping of some domain blocks onto range blocks.

As an example, Figure 5 illustrates the decomposition of a face image using the fractal technique. The original image, and the first, third, and fifth iterates of the reconstructed image are demonstrated for 8x8 range blocks and 16x16 domain blocks.

III. SYSTEM REALISATION

Neural networks are trained to carry out different stages of face compression and recognition. The main reason for employing neural networks is to take advantage of the highly parallel architecture of neural networks in order to achieve real-time recogni-

tion. Two different feedforward networks are designed. The first network implements the search process in which the best approximation of a range block is found among the domain pool. The network accepts two groups of inputs, a domain block and a range block, and calculates the RMS error between the range block and the transformed domain block. Only one network is sufficient to perform the search process. However to speed up the search, more networks with the same architecture can be used in parallel.

The second network has also a feedforward architecture and has been trained to recognise faces. It receives two group of inputs, the fractal model of an input image and the fractal model of a known face from the fractal database. It calculates the degree of similarity between the two fractal face models. If the output is below a threshold, the input face is considered as a known face whose fractal is currently under examination. Similar to the compression stage, more networks with the same architecture can be used to speed up the recognition process.

IV. EXPERIMENTAL RESULTS

Experiments were carried out on a database of known face images to verify the performance of the system. The database consists of 150 front-view face images of 64x64, 256 gray levels. Simulations were done in two steps: compression and recognition.

In the compression stage, the fractal model of all known faces are found and stored. The domain and range block sizes were set to 16x16 and 8x8 pixels, respectively. The compression threshold was selected as 0.002.

In the recognition stage, two groups of experiments were performed. In the first experiment, the known face images were presented to the system and the recognition process was carried out. 100% recognition was achieved in this step. The maximum compression ratio obtained for successful recognition of all known faces was 89:1. Figure 6 demonstrates recognition of the known face number 100. In the second experiment, 50 unknown face images were directed to the system and the recognition process was carried out. None of the input images were recognised by the system. Figure 7 shows the recognition result for an unknown input face image.

The proposed system is less computationally expensive than that for PCA. Moreover, when a new face is added to the database, recalculation of eigenfaces and re-compression of known faces are not necessary.

V. CONCLUSIONS

Feedforward neural networks have been employed to implement a fractal based human face representation and recognition system. An input image is transformed into its fractal model and compared with fractal models of known faces. Simulation results show that the proposed system overcomes the shortcomings of PCA. The maximum compression ratio obtained for the successful recognition of the known faces was observed to be 89:1. A comprehensive face image database is currently being constructed by the authors. This will allow future experiments to be conducted with larger numbers of face images.

ACKNOWLEDGEMENTS

The first author would like to thank DEET, Australia, and FUSA, for providing him with scholarships to undertake this research.

REFERENCES

- [1] M. Barnsley, *Fractals Everywhere*. San Diego: Academic Press, Inc., 1988.
- [2] M.F. Barnsley and L.P. Hurd, *Fractal Image Compression*. AK Peters Ltd., 1993.
- [3] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705-740, May, 1995.
- [4] Y. Fisher (Ed.), *Fractal Image Compression: Theory and Application*. Springer-Verlag New York, Inc., 1995.
- [5] A.E. Jacquin, "Fractal image coding: A review," *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1451-1465, October, 1993.
- [6] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103-108, January, 1990.
- [7] A.Z. Kouzani, F. He, and K. Sammut, "Commonsense knowledge representation and reasoning with fuzzy neural networks," *Proc. 1996 Australian and New Zealand Conference on Intelligent Information Processing Systems*, Adelaide, Australia, pp. 237-240, November, 1996.
- [8] A.Z. Kouzani, F. He, and K. Sammut, "Commonsense knowledge-based face detection," To be presented at *1997 IEEE International Conference on Intelligent Engineering Systems*, Budapest, Hungary, 15-17 September 1997.
- [9] B.B. Mandelbrot, *The Fractal Geometry of Nature*. New York: Freeman and Company, 1983.
- [10] A.J. O'Toole, H. Abdi, K.A. Deffenbacher, and D. Valentin, "Low-dimensional representation of faces in higher dimensions of the face space," *Optical Society of America*, vol. 10, no. 3, pp. 405-411, March, 1993.
- [11] A. Samal and P.A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: A survey," *Pattern Recognition*, vol. 25, no. 1, pp. 65-77, 1992.
- [12] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.
- [13] D. Valentin, H. Abdi, A.J. O'Toole, and G.W. Cottrell, "Connectionist models of face processing: A survey," *Pattern Recognition*, vol. 27, pp. 1209-1230, 1994.



Fig. 1. A contractive transformation of a face image in a plane.



Fig. 2. An attractor of an IFS with three maps.

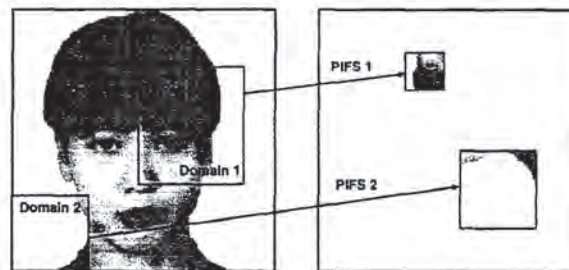


Fig. 3. Partitioned iterated function systems.

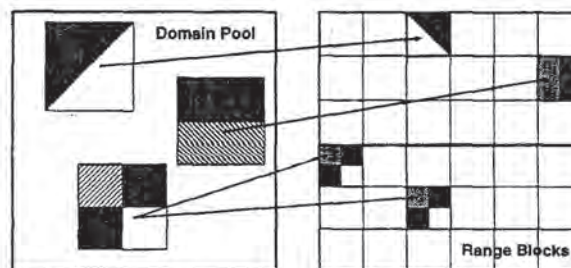


Fig. 4. Mapping of some domain blocks onto range blocks.

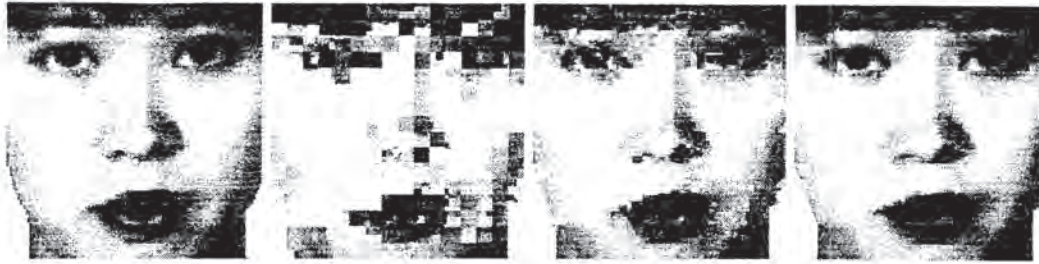


Fig. 5. Fractal decomposition of a sample face image. a) Original image, b) first, c) third, d) fifth iterates of reconstructed image for 8x8 range blocks & 16x16 domain blocks.

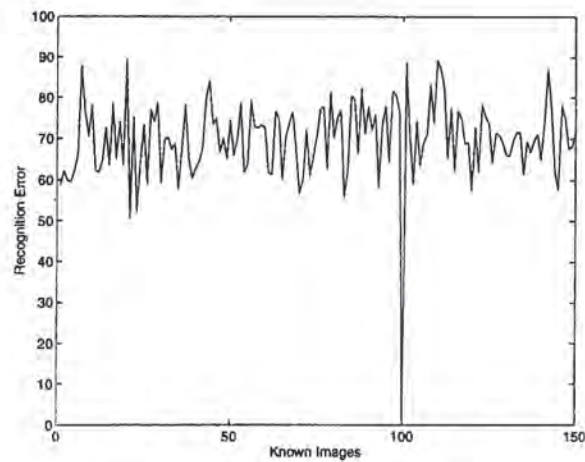


Fig. 6. Recognition result for a known input face image (number 100).

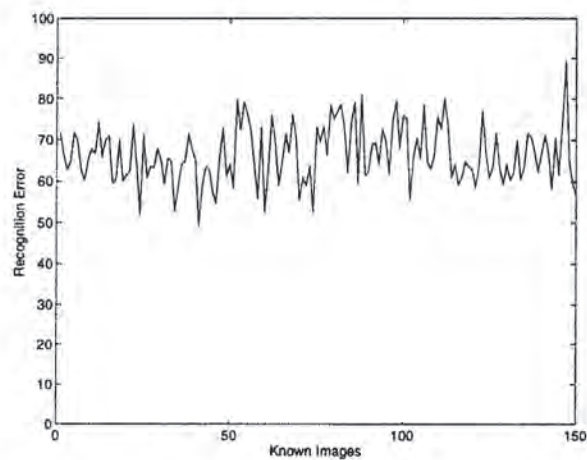


Fig. 7. Recognition result for an unknown input face image.

EXHIBIT 1006B

LIBRARY

CATALOG



 1 of 1
 

PERIODICAL OR NEWSPAPER

Conference proceedings

Full Record

[MARC Tags](#)

Meeting name

[IEEE International Conference on Systems, Man, and Cybernetics.](#)

Main title

Conference proceedings / IEEE International Conference on Systems, Man, and Cybernetics.

Published/Created

New York, NY : Institute of Electrical and Electronic[s] Engineers, c1989-

[Request this Item](#)

 [LC Find It](#)

More Information



LCCN Permalink <https://lccn.loc.gov/93657490>

Publication history 1989-

Description v. : ill. ; 28 cm.
Issued in parts.

Current frequency Annual

ISSN 1062-922X

Linking ISSN	1062-922X
LC classification	Q300 .I485a
Continues	IEEE International Conference on Systems, Man, and Cybernetics. Proceedings of the ... IEEE International Conference on Systems, Man, and Cybernetics (DLC) 93657485 (OCoLC)15064476
Continued by	IEEE International Conference on Systems, Man, and Cybernetics. IEEE International Conference on Systems, Man, and Cybernetics. IEEE International Conference on Systems, Man, and Cybernetics 2163-9590 (DLC) 2011206180 (OCoLC)187462379
Variant title	Some issues have title: IEEE International Conference on Systems, Man, and Cybernetics
Serial key title	Conference proceedings - IEEE International Conference on Systems, Man, and Cybernetics
Abbreviated title	Conf. proc. - IEEE Int. Conf. Syst. Man Cybern.
Related names	IEEE Systems, Man, and Cybernetics Society.
LC Subjects	Cybernetics--Congresses. Systems engineering--Congresses. Human-machine systems--Congresses. Technology and state--Congresses.
Other Subjects	Cybernetics. Human-machine systems. Systems engineering. Technology and state.
Form/Genre	Conference papers and proceedings.
Browse by shelf order	Q300
Notes	Proceedings for 1989, 1991-1992, 1995, are cataloged as monographs in LC. Sponsored by: IEEE Systems, Man, and Cybernetics Society.
Reproduction no./Source	Curran Associates, Inc., 57 Morehouse Lane, Red Hook, NY 12571
Additional formats	Online version: IEEE International Conference on Systems, Man, and

Cybernetics. Conference proceedings 2577-1655 (DLC) 2018200153
 (OCoLC)624332087
 Online version: IEEE International Conference on Systems, Man, and
 Cybernetics. Conference proceedings (OCoLC)1485043935

LCCN	93657490
Invalid LCCN	sn 91032043 92055091 sn 94037075 sn 94037076 2005938111
Dewey class no.	001.53/05
Other system no.	(OCoLC)ocm21479523
Additional Links	http://ieeexplore.ieee.org/xpl/conhome.jsp?punumber=1000738
Type of material	Periodical or Newspaper

Item Availability

>

CALL NUMBER	Q300 .I485a Set 1 Proceedings for 1989, 1991-1992, 1995, and 2005 are cataloged as monographs in LC.
Request in	Jefferson or Adams Building Reading Rooms
Status	Not Charged
Older receipts	1990, 1992-1994, 1996-1998

LIBRARY

CATALOG

PERIODICAL OR NEWSPAPER

Conference proceedings

Full Record

MARC Tags

000	03399cas a2200601 a 4500
001	11423960
005	20250121065659.0
008	900504d198919uunyuar 1 a0eng c
010	___ a 93657490 z sn 91032043 z 92055091 z sn 94037075 z sn 94037076 z 2005938111
022	0_ a 1062-922X l 1062-922X 2 1
035	___ a (OCoLC)ocm21479523
037	___ b Curran Associates, Inc., 57 Morehouse Lane, Red Hook, NY 12571
040	___ a IWA b eng c IWA d MYG d NST d ELM d COO d NSD d NST d OCL d NST d DLC d MYG d DLC d WAU d MYG d DLC d OCLCO d BOS d MYG d OCLCO d DLC d OCLCF d OCLCO d DLC d OCL d OCLCO d DLC
042	___ a pcc a nsdp
050	00 a Q300 b .l485a
082	00 a 001.53/05 2 19
111	2_ a IEEE International Conference on Systems, Man, and Cybernetics.
210	0_ a Conf. proc. - IEEE Int. Conf. Syst. Man Cybern.
222	_0 a Conference proceedings - IEEE International Conference on Systems, Man, and Cybernetics
245	10 a Conference proceedings / c IEEE International Conference on Systems, Man, and Cybernetics.
246	1_ i Some issues have title: a IEEE International Conference on Systems, Man, and Cybernetics
260	___ a New York, NY : b Institute of Electrical and Electronic[s] Engineers, c c1989-

300	—	a v. : b ill. ; c 28 cm.
310	—	a Annual
362	0_	a 1989-
500	—	a Proceedings for 1989, 1991-1992, 1995, are cataloged as monographs in LC. 5 DLC
515	—	a Issued in parts.
550	—	a Sponsored by: IEEE Systems, Man, and Cybernetics Society.
590	—	a SERBIB/SERLOC merged record
650	_0	a Cybernetics v Congresses.
650	_0	a Systems engineering v Congresses.
650	_0	a Human-machine systems v Congresses.
650	_0	a Technology and state v Congresses.
650	_7	a Cybernetics. 2 fast 0 (OCoLC)fst00885777
650	_7	a Human-machine systems. 2 fast 0 (OCoLC)fst00963500
650	_7	a Systems engineering. 2 fast 0 (OCoLC)fst01141455
650	_7	a Technology and state. 2 fast 0 (OCoLC)fst01145265
655	_7	a Conference papers and proceedings. 2 fast 0 (OCoLC)fst01423772
710	2_	a IEEE Systems, Man, and Cybernetics Society.
776	08	i Online version: IEEE International Conference on Systems, Man, and Cybernetics. t Conference proceedings x 2577-1655 w (DLC) 2018200153 w (OCoLC)624332087
776	08	i Online version: a IEEE International Conference on Systems, Man, and Cybernetics. t Conference proceedings w (OCoLC)1485043935
780	00	a IEEE International Conference on Systems, Man, and Cybernetics. t Proceedings of the ... IEEE International Conference on Systems, Man, and Cybernetics w (DLC) 93657485 w (OCoLC)15064476
785	00	a IEEE International Conference on Systems, Man, and Cybernetics. s IEEE International Conference on Systems, Man, and Cybernetics. t IEEE International Conference on Systems, Man, and Cybernetics x 2163-9590 w (DLC) 2011206180 w (OCoLC)187462379
850	—	a CoU a DLC a IaAS a InLP a MCM a MiU a NIC a WaU
859	41	u http://ieeexplore.ieee.org/xpl/conhome.jsp?punumber=1000738
890	—	a Conference proceedings (IEEE International Conference on Systems, Man, and Cybernetics) i 93-657490
906	—	a 7 b cbc c serials d u e ncip f 19 g n-oclcserc
920	—	a Keep 1
984	—	a srvf d 2000-12-16
985	—	e eserial 200406
991	—	b c-GenColl h Q300 i .I485a w SERIALS
991	—	b c-GenColl h Q300 i .I485a p 00058893638 t Copy 1 v 1 * 1998 w CCF

992 — |b B/L |w SERLOC

Request this Item

 LC Find It

Item Availability >

CALL NUMBER	Q300 .I485a Set 1 Proceedings for 1989, 1991-1992, 1995, and 2005 are cataloged as monographs in LC.
Request in	Jefferson or Adams Building Reading Rooms
Status	Not Charged
Older receipts	1990, 1992-1994, 1996-1998

EXHIBIT 1006C



LOCIS: Library of Congress Information System

TELNET locis.loc.gov

No login required

L O C I S : LIBRARY OF CONGRESS INFORMATION SYSTEM

- | | | | |
|----|---|---|-------------------|
| 1 | Library of Congress Catalog | 4 | Braille and Audio |
| 2 | Federal Legislation | 5 | Organizations |
| 3 | Copyright Information | 6 | Foreign Law |
| 7 | Searching Hours and Basics | | |
| 8 | Documentation and Classes | | |
| 9 | Library of Congress General Information | | |
| 12 | Comments and Logoff | | |

LIBRARY OF CONGRESS CATALOG

CHOICE

FILE

- 1 BOOKS: English language books 1968-, French 1973-, German, Portuguese, Spanish 1975-, other European languages 1976-77, non-European languages 1978-79. Some microforms 1984-.

LOC1

The Wayback Machine - <https://web.archive.org/web/19970524160715/http://library.usask.ca:80/hytelnet/us3/us373.html>

LOCIS: Library of Congress Information System

TELNET locis.loc.gov

No login required

L O C I S : LIBRARY OF CONGRESS INFORMATION SYSTEM

- | | | | |
|----|---|---|-------------------|
| 1 | Library of Congress Catalog | 4 | Braille and Audio |
| 2 | Federal Legislation | 5 | Organizations |
| 3 | Copyright Information | 6 | Foreign Law |
| 7 | Searching Hours and Basics | | |
| 8 | Documentation and Classes | | |
| 9 | Library of Congress General Information | | |
| 12 | Comments and Logoff | | |

LIBRARY OF CONGRESS CATALOG

- | CHOICE | | FILE |
|--------|--|------|
| 1 | BOOKS: English language books 1968-, French 1973-, German, Portuguese, Spanish 1975-, other European languages 1976-77, non-European languages 1978-79. Some microforms 1984-. | LOCI |
| 2 | BOOKS earlier than the dates above. Some serials, maps, music, audiovisual items. | PREM |
| 3 | Combination of files 1 and 2 above (LOCI and PREM). | |
| 4 | SERIALS cataloged at LC & some other libraries since 1973. | LOCS |
| 5 | MAPS and other cartographic items cataloged at LC 1968- and some other research libraries 1985-. Atlases (which are books) are in LOCI and PREM. | LOCM |
| 6 | SUBJECT TERMS and cross references from LC Subject Headings. | LCXR |
| 12 | Return to LOCIS MENU screen. | |

libraries.org

A directory of libraries throughout the world

- [Home »](#)
- [Libraries »](#)
- [Guides](#)
- [Documents](#)
- [Vendors »](#)
- [Products »](#)
- [News](#)
- [Procurement](#)
- [Member Login »](#)

found 3 items where the item contains **Library of Congress**. Showing item 2 of 3.

<< prev list next >> search:

Library of Congress

United States -- [Library of Congress](#)

Address: 101 Independence Avenue SE
Washington, District of Columbia
20540
United States

Phone: 202-707-5000

Connect to: [Library Web Site](#) ✓ [Online Catalog](#)

Library details: **Library of Congress** is a **National** library.
This library is affiliated with [Library of Congress \(view map\)](#) . The collection of the library contains **175,000,000** volumes. The library circulates **250,000** items per year.

Other Info: Went live with Voyager on August 15, 1999. Currently planning to implement the Library Collections Access Platform (LCAP) on FOLIO and EBSCO discovery services in FY25.

Permalink:  <https://librarytechnology.org/library/2263>
(Use this link to refer back to this listing.)

Administration: The director of the library is **Dr. Carla Hayden** .

Organizational structure: The library is part of a **publicly-funded** organization. It operates on a **Non-profit** financial model.



Affiliations: The library is a member of the [Association of Research Libraries](#)

Collection size	175,000,000	volumes
Annual Circulation	250,000	transactions

 **Wireless:** The library offers **wireless** access to the Internet. <http://www.loc.gov/rr/wirelessaccess.html>

Library Hours		
Day	Open	Close
Sunday	Closed	
Monday	8:30 am	8:30 pm
Tuesday	8:30 am	8:30 pm
Wednesday	8:30 am	8:30 pm
Thursday	8:30 am	8:30 pm
Friday	8:30 am	5:00 pm
Saturday	8:30 am	5:00 pm



Library of Congress



map location ✓

Related Libraries

- All members of the [Association of Research Libraries](#) members
- [Directory of ARL Member Libraries](#)
- Libraries located in [Washington](#) (District of Columbia)
- [all National libs in District of Columbia](#)

About libraries.org

libraries.org is a free directory of libraries throughout the world.

This directory is a component of [Library Technology Guides](#).

Technology Profile		
	Product Name	Year Contracted
Current Automation System	FOLIO -- EBSCO Information Services	2023
Previous Automation System	Voyager	1998
Discovery Service (w/index)	Ex Libris CDI	
Electronic Resource Management	Electronic Resource Management	
Item ID Type	Barcode	
This library is responsible for the procurement of the library automation system.		

Please [contact Marshall Breeding](#) to report corrections about information listed for this library.

[View Privacy and Security Report](#)

Automation Survey: We are conducting a study on the automation systems used in libraries.

This survey should be completed by a person familiar with your libraries use of

FOLIO -- EBSCO Information Services.

[Respond to the 2024 Library Automation Perceptions Survey](#)

Identifiers	
libraries.org ID	2263
OCLC Symbol	DLC
WorldCat Registry ID	45266

Record History: This listing was created on Aug 9, 1998 and was last modified on Feb 13, 2025.

Updates: Corrections or Updates? Registered members of Library Technology Guides can submit updates to library listings in libraries.org. [Registration](#) is free and easy. Already registered? [login](#). Or, you can report corrections just by sending a message to [Marshall Breeding](#).

[libraries.org](#)

[Update this Entry](#)

[Library Technology Guides](#)

Maintained by [Marshall Breeding](#)

[terms of service](#)

Copyright 1994-2024

EXHIBIT 1006D

[Skip to search](#) [Skip to main content](#)

1997 IEEE International Conference on Systems, Man and Cybernetics : Hyatt Orlando, Orlando, Florida, USA, October 12-15, 1997 : computational cybernetics and simulation / organizing committee, James M. Tien ... [et al.].

Author

IEEE International Conference on Systems, Man, and Cybernetics (1997 : Orlando, Fla.)

Format

Book

Language

English

Published/Created

Piscataway, NJ : IEEE Service Center, c1997.

Description

5 v. (4535 p.) : ill. ; 28 cm.

Available Online

[IEEE Electronic Library \(IEL\)](#) 

Copies in the Library

ReCAP - Remote Storage

TA168 .I567 1997

Available

Location has

Vol. 1-v. 5

Subject(s)

Systems engineering—Congresses

Cybernetics—Congresses

Technology and state—Congresses

Notes

"IEEE catalog number 97CH36088-5"--T.p. verso.

Bibliographic references

Includes bibliographical references and index.

Other title(s)

1997 IEEE International Conference on Systems, Man and Cybernetics
Computational cybernetics and simulation

ISBN

0780340531 (softbound)
078034054X (casebound)

OCLC

37907023

Statement on responsible collection description

Princeton University Library aims to describe library materials in a manner that is respectful to the individuals and communities who create, use, and are represented in the collections we manage.
Read more...

[Ask a Question](#) [Suggest a Correction](#)

[Skip to search](#) [Skip to main content](#)

Staff view

LEADER 01685cam a2200349 a 4500

001 9911918883506421

005 20240812102530.0

008 971125s1997 njua b 101 0 eng d

020 a| 0780340531 (softbound)

020 a| 078034054X (casebound)

035 a| (CStRLIN)NJP97-B66629

035 9| CCU6283TS

035 a| (NjP)1191888-princetondb

035 a| (OCoLC)ocm37907023

035 z| (NjP)Voyager1191888

040 a| NjP b| eng c| NjP

050 4 a| TA168 b| .I567 1997

111 2 a| IEEE International Conference on Systems, Man, and Cybernetics d| (1997 : c| Orlando, Fla.)

245 1 0 a| 1997 IEEE International Conference on Systems, Man and Cybernetics : b| Hyatt Orlando, Orlando, Florida, USA, October 12-15, 1997 : computational cybernetics and simulation / c| organizing committee, James M. Tien ... [et al.].

246 1 4 a| 1997 IEEE International Conference on Systems, Man and Cybernetics

246 3 0 a| Computational cybernetics and simulation

260 a| Piscataway, NJ : b| IEEE Service Center, c| c1997.

300 a| 5 v. (4535 p.) : b| ill. ; c| 28 cm.

504 a| Includes bibliographical references and index.

500 a| "IEEE catalog number 97CH36088-5"--T.p. verso.

650 0 a| Systems engineering v| Congresses. 0| <http://id.loc.gov/authorities/subjects/sh2008111998>

650 0 a| Cybernetics v| Congresses. 0| <http://id.loc.gov/authorities/subjects/sh2008101510>

650 0 a| Technology and state v| Congresses. 0| <http://id.loc.gov/authorities/subjects/sh2008112648>

911 a| 19971204

912 a| 19971125000000.0

914 a| (OCoLC)ocm37907023 b| OCoLC c| match d| 20221117 e| processed f| 37907023

998 a| 11/25/97 s| 9114 n| NjP d| 11/25/97 c| MWC b| MWC i| 971125 l| NJPG

AVA 0| 9911918883506421 8| 22488311310006421 a| 01PRI_INST b| recap c|

Remote Storage d| TA168 .I567 1997 e| available f| 5 g| 0 j| pa k| 0 p| 1 q|

ReCAP t| Vol. 1-v. 5 t|

EXHIBIT 1006E

libraries.org

A directory of libraries throughout the world

- [Home »](#)
- [Libraries »](#)
- [Guides](#)
- [Documents](#)
- [Vendors »](#)
- [Products »](#)
- [News](#)
- [Procurement](#)
- [Member Login »](#)

found 1 item where the item contains **Princeton University Library**.

☐ [list](#) ☐ search:

Princeton University Library

[Princeton University](#)

Address: 1 Washington Road
[Princeton, New Jersey](#)
 08544
 United States

County: [Mercer](#) 

Phone: 609-258-1470

Connect to: [Library Web Site](#) ✓ [Online Catalog](#)

Library details: Princeton University Library is an **Academic** library.

This library is affiliated with [Princeton University](#) ([view map](#)). The collection of the library contains **13,074,645** volumes. The library circulates **212,110** items per year. The library serves a population of **7,972** FTE Enrollment.

Permalink:  <https://librarytechnology.org/library/2168>
 (Use this link to refer back to this listing.)

Administration: The director of the library is **Anne Jarvis**.

Organizational structure: The library is part of a **privately-funded** organization. It operates on a **Non-profit** financial model.

Academic Level: The library Serves an educational institution at the **University** level. The Carnegie Foundation designates this institution as: **Doctoral Universities: Very High Research Activity** (level 15).

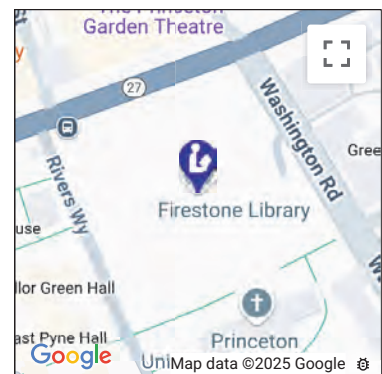


Affiliations: The library is a member of the [Association of Research Libraries](#)

Statistics Academic		
Enrollment	7,972	FTE Students
Collection size	13,074,645	volumes
Annual Circulation	212,110	transactions



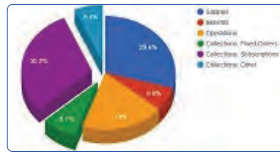
Firestone Library



map location ✓

Related Libraries

- Other libraries with Carnegie class: **Doctoral Universities: Very High Research Activity**
- All members of the [Association of Research Libraries](#) members
- [Directory of ARL Member Libraries](#)
- Libraries located in [Princeton](#) (New Jersey)
- Libraries located in [Mercer](#) county (New Jersey)
- [View map of libraries in Mercer County](#)
- [all Academic libs in New Jersey](#)



[View Detailed Statistics](#)

About libraries.org

libraries.org is a free directory of libraries throughout the world.

This directory is a component of **Library Technology Guides**.

Please **contact Marshall Breeding** to report corrections about information listed for this library.

Technology Profile

	Product Name	Year Contracted
Current Automation System	Alma	2019
Previous Automation System	Voyager	2000
Previous Automation System	NOTIS	1991
Previous Automation System	TOMUS	1987
Previous Automation System	GLIS	1979
Discovery Service (w/index)	Ex Libris CDI	2010
Discovery Interface	Blacklight	2016
Previous Discovery Interface	Voyager	2000
Authentication Service	OpenAthens -- EBSCO	2020
Digital Collection Management	Samvera	2017
Institutional Repository	DSpace	2015
Digital Preservation	LIBSAFE Go	2022
Electronic Resource Management	Alma	
Archives Management	ArchivesSpace	2020
Interlibrary Loan or Resource Sharing	ReShare	
Resource Sharing Networks	BorrowDirect ReCAP	
Web Content Management	Drupal	
Item ID Type	Barcode	

The library's automation system is hosted by the vendor through an **Software-as-a-Service (SaaS)** arrangement.

This library is responsible for the procurement of the library automation system.

[View Privacy and Security Report](#)

Identifiers

libraries.org ID	2168
OCLC Symbol	PUL
WorldCat Registry ID	1791
NCES LIBID	186131
Ringgold ID	6740

Record History: This listing was created on Dec 10, 1997 and was last modified on Apr 10, 2025.

Updates: Corrections or Updates? Registered members of Library Technology Guides can submit updates to library listings in libraries.org. **Registration** is free and easy. Already registered? **login**. Or, you can report corrections just by sending a message to **Marshall Breeding**.

[Basic Search](#)

[Advanced Search](#)

[Library Technology Guides](#)

Maintained by **Marshall Breeding**

[terms of service](#)

Copyright 1994-2024

EXHIBIT 1006F

Fractal face representation and recognition

Publisher: IEEE

[Cite This](#) PDFA.Z. Kouzani ; F. He ; K. Sammut [All Authors](#)

9

Cites in
Papers

97

Full
Text Views**Abstract**[Authors](#)[References](#)[Citations](#)[Keywords](#)[Metrics](#)[More Like This](#)**Abstract:**

This paper presents a face representation and recognition scheme based on the theory of fractals. Each face image is represented by its fractal model which is a small collection of transformation parameters. The transformation is carried out once for known face images. For recognition, the input face image is transformed and its fractal model is then compared against the database of fractal models of known faces. Feedforward neural networks are utilised to implement the compression and recognition parts. Some experimental results are presented. The maximum compression ratio obtained for the successful recognition of known faces was observed to be 89:1 (for a compression threshold of 0.002).

Published in: 1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation

Date of Conference: 12-15 October 1997

DOI: [10.1109/ICSMC.1997.638231](#)

Date Added to IEEE Xplore: 06 August 2002

Publisher: IEEE

Print ISBN: 0-7803-4053-1

Conference Location: Orlando, FL, USA

Print ISSN: 1062-922X

[Authors](#)[References](#)[Citations](#)[Keywords](#)[Metrics](#)

**Need
Full-Text**
access to IEEE Xplore
for your organization?

[CONTACT IEEE TO SUBSCRIBE >](#)**IEEE Personal Account**[CHANGE USERNAME/
PASSWORD](#)**Purchase Details**[PAYMENT OPTIONS](#)
[VIEW PURCHASED](#)**Profile Information**[COMMUNICATIONS
PREFERENCES](#)**Need Help?**[US & CANADA: +1 800
678 4333](#)**Follow**[f](#) [@](#) [in](#) [v](#)

DOCUMENTS

PROFESSION AND
EDUCATION

TECHNICAL INTERESTS

WORLDWIDE: +1 732 981
0060

CONTACT & SUPPORT

About IEEE *Xplore* | Contact Us | Help | Accessibility | Terms of Use | Nondiscrimination Policy | IEEE Ethics Reporting  | Sitemap |
IEEE Privacy Policy

A public charity, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.

© Copyright 2025 IEEE - All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies.



EXHIBIT 1006G

EX1006G
Citations to Kouzani

Source Document:

Kouzani, A. Z., He, F., & Sammut, K. (1997, October). Fractal face representation and recognition. In *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation* (Vol. 2, pp. 1609-1613). ("Kouzani")

Citations: 1999 – 2007 (in chronological order)

Tan, T., & Yan, H. (1999, March). Face recognition by fractal transformations. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)* (Vol. 6, pp. 3537-3540). IEEE.

Schouten, B. A., & de Zeeuw, P. M. (2000, September). Image databases, scale and fractal transforms. In *Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101)* (Vol. 2, pp. 534-537). IEEE.

Ebrahimpour-Komleh, H., Chandran, V., & Sridharan, S. (2001, October). Face recognition using fractal codes. In *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)* (Vol. 3, pp. 58-61). IEEE.

Mozaffari, S., Faez, K., & Ziaratban, M. (2005, August). Character representation and recognition using quad tree-based fractal encoding scheme. In *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)* (pp. 819-823). IEEE.

Abate, A. F., Nappi, M., Riccio, D., & Sabatino, G. (2007, October). 2D and 3D face recognition: A survey. *Pattern recognition letters*, 28(14), 1885-1906.

EXHIBIT 1007A

Detecting Faces in Images: A Survey

Ming-Hsuan Yang, *Member, IEEE*, David J. Kriegman, *Senior Member, IEEE*, and
Narendra Ahuja, *Fellow, IEEE*

Abstract—Images containing faces are essential to intelligent vision-based human computer interaction, and research efforts in face processing include face recognition, face tracking, pose estimation, and expression recognition. However, many reported methods assume that the faces in an image or an image sequence have been identified and localized. To build fully automated systems that analyze the information contained in face images, robust and efficient face detection algorithms are required. Given a single image, the goal of face detection is to identify all image regions which contain a face regardless of its three-dimensional position, orientation, and lighting conditions. Such a problem is challenging because faces are nonrigid and have a high degree of variability in size, shape, color, and texture. Numerous techniques have been developed to detect faces in a single image, and the purpose of this paper is to categorize and evaluate these algorithms. We also discuss relevant issues such as data collection, evaluation metrics, and benchmarking. After analyzing these algorithms and identifying their limitations, we conclude with several promising directions for future research.

Index Terms—Face detection, face recognition, object recognition, view-based recognition, statistical pattern recognition, machine learning.

1 INTRODUCTION

WITH the ubiquity of new information technology and media, more effective and friendly methods for human computer interaction (HCI) are being developed which do not rely on traditional devices such as keyboards, mice, and displays. Furthermore, the ever decreasing price/performance ratio of computing coupled with recent decreases in video image acquisition cost imply that computer vision systems can be deployed in desktop and embedded systems [111], [112], [113]. The rapidly expanding research in face processing is based on the premise that information about a user's identity, state, and intent can be extracted from images, and that computers can then react accordingly, e.g., by observing a person's facial expression. In the last five years, face and facial expression recognition have attracted much attention though they have been studied for more than 20 years by psychophysicists, neuroscientists, and engineers. Many research demonstrations and commercial applications have been developed from these efforts. A first step of any face processing system is detecting the locations in images where faces are present. However, face detection from a single image is a challenging task because of variability in scale, location, orientation (up-right, rotated), and pose (frontal, profile). Facial expression, occlusion, and lighting conditions also change the overall appearance of faces.

We now give a definition of *face detection*: Given an arbitrary image, the goal of face detection is to determine whether or not there are any faces in the image and, if present, return the image location and extent of each face. The challenges associated with face detection can be attributed to the following factors:

- **Pose.** The images of a face vary due to the relative camera-face pose (frontal, 45 degree, profile, upside down), and some facial features such as an eye or the nose may become partially or wholly occluded.
- **Presence or absence of structural components.** Facial features such as beards, mustaches, and glasses may or may not be present and there is a great deal of variability among these components including shape, color, and size.
- **Facial expression.** The appearance of faces are directly affected by a person's facial expression.
- **Occlusion.** Faces may be partially occluded by other objects. In an image with a group of people, some faces may partially occlude other faces.
- **Image orientation.** Face images directly vary for different rotations about the camera's optical axis.
- **Imaging conditions.** When the image is formed, factors such as lighting (spectra, source distribution and intensity) and camera characteristics (sensor response, lenses) affect the appearance of a face.

There are many closely related problems of face detection. *Face localization* aims to determine the image position of a single face; this is a simplified detection problem with the assumption that an input image contains only one face [85], [103]. The goal of *facial feature detection* is to detect the presence and location of features, such as eyes, nose, nostrils, eyebrow, mouth, lips, ears, etc., with the assumption that there is only one face in an image [28], [54]. *Face recognition* or *face identification* compares an input image (probe) against a database (gallery) and reports a match, if

- M.-H. Yang is with Honda Fundamental Research Labs, 800 California Street, Mountain View, CA 94041. E-mail: myang@hrl.com.
- D.J. Kriegman is with the Department of Computer Science and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801. E-mail: kriegman@uiuc.edu.
- N. Ahuja is with the Department of Electrical and Computer Engineering and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801. E-mail: ahuja@vision.ai.uiuc.edu.

Manuscript received 5 May 2000; revised 15 Jan. 2001; accepted 7 Mar. 2001.
Recommended for acceptance by K. Bowyer.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 112058.

Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.

0162-8828/02/\$17.00 © 2002 IEEE

any [163], [133], [18]. The purpose of *face authentication* is to verify the claim of the identity of an individual in an input image [158], [82], while *face tracking* methods continuously estimate the location and possibly the orientation of a face in an image sequence in real time [30], [39], [33]. *Facial expression recognition* concerns identifying the affective states (happy, sad, disgusted, etc.) of humans [40], [35]. Evidently, face detection is the first step in any automated system which solves the above problems. It is worth mentioning that many papers use the term “face detection,” but the methods and the experimental results only show that a single face is localized in an input image. In this paper, we differentiate *face detection* from *face localization* since the latter is a simplified problem of the former. Meanwhile, we focus on face detection methods rather than tracking methods.

While numerous methods have been proposed to detect faces in a single image of intensity or color images, we are unaware of any surveys on this particular topic. A survey of early face recognition methods before 1991 was written by Samal and Iyengar [133]. Chellapa et al. wrote a more recent survey on face recognition and some detection methods [18].

Among the face detection methods, the ones based on learning algorithms have attracted much attention recently and have demonstrated excellent results. Since these data-driven methods rely heavily on the training sets, we also discuss several databases suitable for this task. A related and important problem is how to evaluate the performance of the proposed detection methods. Many recent face detection papers compare the performance of several methods, usually in terms of detection and false alarm rates. It is also worth noticing that many metrics have been adopted to evaluate algorithms, such as learning time, execution time, the number of samples required in training, and the ratio between detection rates and false alarms. Evaluation becomes more difficult when researchers use different definitions for detection and false alarm rates. In this paper, *detection rate* is defined as the ratio between the number of faces correctly detected and the number faces determined by a human. An image region identified as a face by a classifier is considered to be correctly detected if the image region covers more than a certain percentage of a face in the image (See Section 3.3 for details). In general, detectors can make two types of errors: *false negatives* in which faces are missed resulting in low detection rates and *false positives* in which an image region is declared to be face, but it is not. A fair evaluation should take these factors into consideration since one can tune the parameters of one’s method to increase the detection rates while also increasing the number of false detections. In this paper, we discuss the benchmarking data sets and the related issues in a fair evaluation.

With over 150 reported approaches to face detection, the research in face detection has broader implications for computer vision research on object recognition. Nearly all model-based or appearance-based approaches to 3D object recognition have been limited to rigid objects while attempting to robustly perform identification over a broad range of camera locations and illumination conditions. Face detection can be viewed as a two-class recognition problem

in which an image region is classified as being a “face” or “nonface.” Consequently, face detection is one of the few attempts to recognize from images (not abstract representations) a class of objects for which there is a great deal of within-class variability (described previously). It is also one of the few classes of objects for which this variability has been captured using large training sets of images and, so, some of the detection techniques may be applicable to a much broader class of recognition problems.

Face detection also provides interesting challenges to the underlying pattern classification and learning techniques. When a raw or filtered image is considered as input to a pattern classifier, the dimension of the feature space is extremely large (i.e., the number of pixels in normalized training images). The classes of face and nonface images are decidedly characterized by multimodal distribution functions and effective decision boundaries are likely to be nonlinear in the image space. To be effective, either classifiers must be able to extrapolate from a modest number of training samples or be efficient when dealing with a very large number of these high-dimensional training samples.

With an aim to give a comprehensive and critical survey of current face detection methods, this paper is organized as follows: In Section 2, we give a detailed review of techniques to detect faces in a single image. Benchmarking databases and evaluation criteria are discussed in Section 3. We conclude this paper with a discussion of several promising directions for face detection in Section 4.¹

Though we report error rates for each method when available, tests are often done on unique data sets and, so, comparisons are often difficult. We indicate those methods that have been evaluated with a publicly available test set. It can be assumed that a unique data set was used if we do not indicate the name of the test set.

2 DETECTING FACES IN A SINGLE IMAGE

In this section, we review existing techniques to detect faces from a single intensity or color image. We classify single image detection methods into four categories; some methods clearly overlap category boundaries and are discussed at the end of this section.

1. **Knowledge-based methods.** These rule-based methods encode human knowledge of what constitutes a typical face. Usually, the rules capture the relationships between facial features. These methods are designed mainly for face localization.
2. **Feature invariant approaches.** These algorithms aim to find structural features that exist even when the pose, viewpoint, or lighting conditions vary, and then use these to locate faces. These methods are designed mainly for face localization.
3. **Template matching methods.** Several standard patterns of a face are stored to describe the face as a whole or the facial features separately. The correlations between an input image and the stored patterns are

1. An earlier version of this survey paper appeared at <http://vision.ai.uiuc.edu/mhyang/face-detection-survey.html> in March 1999. Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.

TABLE 1
Categorization of Methods for Face Detection in a Single Image

Approach	Representative Works
Knowledge-based	Multiresolution rule-based method [170]
Feature invariant	
– Facial Features	Grouping of edges [87] [178]
– Texture	Space Gray-Level Dependence matrix (SGLD) of face pattern [32]
– Skin Color	Mixture of Gaussian [172] [98]
– Multiple Features	Integration of skin color, size and shape [79]
Template matching	
– Predefined face templates	Shape template [28]
– Deformable Templates	Active Shape Model (ASM) [86]
Appearance-based method	
– Eigenface	Eigenvector decomposition and clustering [163]
– Distribution-based	Gaussian distribution and multilayer perceptron [154]
– Neural Network	Ensemble of neural networks and arbitration schemes [128]
– Support Vector Machine (SVM)	SVM with polynomial kernel [107]
– Naive Bayes Classifier	Joint statistics of local appearance and position [140]
– Hidden Markov Model (HMM)	Higher order statistics with HMM [123]
– Information-Theoretical Approach	Kullback relative information [89] [24]

computed for detection. These methods have been used for both face localization and detection.

4. **Appearance-based methods.** In contrast to template matching, the models (or templates) are learned from a set of training images which should capture the representative variability of facial appearance. These learned models are then used for detection. These methods are designed mainly for face detection.

Table 1 summarizes algorithms and representative works for face detection in a single image within these four categories. Below, we discuss the motivation and general approach of each category. This is followed by a review of specific methods including a discussion of their pros and cons. We suggest ways to further improve these methods in Section 4.

2.1 Knowledge-Based Top-Down Methods

In this approach, face detection methods are developed based on the rules derived from the researcher's knowledge of human faces. It is easy to come up with simple rules to describe the features of a face and their relationships. For example, a face often appears in an image with two eyes that are symmetric to each other, a nose, and a mouth. The relationships between features can be represented by their

relative distances and positions. Facial features in an input image are extracted first, and face candidates are identified based on the coded rules. A verification process is usually applied to reduce false detections.

One problem with this approach is the difficulty in translating human knowledge into well-defined rules. If the rules are detailed (i.e., strict), they may fail to detect faces that do not pass all the rules. If the rules are too general, they may give many false positives. Moreover, it is difficult to extend this approach to detect faces in different poses since it is challenging to enumerate all possible cases. On the other hand, heuristics about faces work well in detecting frontal faces in uncluttered scenes.

Yang and Huang used a hierarchical knowledge-based method to detect faces [170]. Their system consists of three levels of rules. At the highest level, all possible face candidates are found by scanning a window over the input image and applying a set of rules at each location. The rules at a higher level are general descriptions of what a face looks like while the rules at lower levels rely on details of facial features. A multiresolution hierarchy of images is created by averaging and subsampling, and an example is shown in Fig. 1. Examples of the coded rules used to locate face candidates in the lowest resolution include: "the center

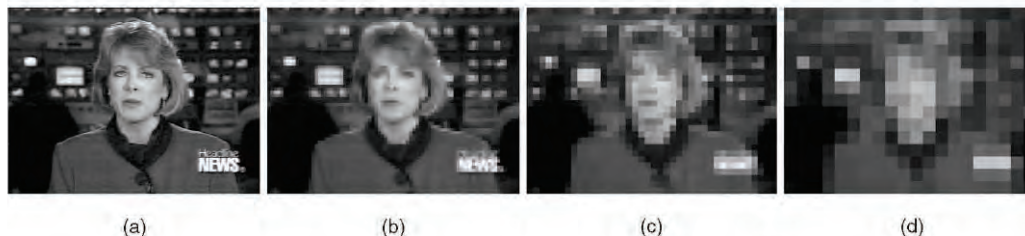


Fig. 1. (a) $n = 1$, original image. (b) $n = 4$. (c) $n = 8$. (d) $n = 16$. Original and corresponding low resolution images. Each square cell consists of $n \times n$ pixels in which the intensity of each pixel is replaced by the average intensity of the pixels in that cell.

Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.



Fig. 2. A typical face used in knowledge-based top-down methods: Rules are coded based on human knowledge about the characteristics (e.g., intensity distribution and difference) of the facial regions [170].

part of the face (the dark shaded parts in Fig. 2) has four cells with a basically uniform intensity," "the upper round part of a face (the light shaded parts in Fig. 2) has a basically uniform intensity," and "the difference between the average gray values of the center part and the upper round part is significant." The lowest resolution (Level 1) image is searched for face candidates and these are further processed at finer resolutions. At Level 2, local histogram equalization is performed on the face candidates received from Level 2, followed by edge detection. Surviving candidate regions are then examined at Level 3 with another set of rules that respond to facial features such as the eyes and mouth. Evaluated on a test set of 60 images, this system located faces in 50 of the test images while there are 28 images in which false alarms appear. One attractive feature of this method is that a coarse-to-fine or focus-of-attention strategy is used to reduce the required computation. Although it does not result in a high detection rate, the ideas of using a multiresolution hierarchy and rules to guide searches have been used in later face detection works [81].

Kotropoulos and Pitas [81] presented a rule-based localization method which is similar to [71] and [170]. First, facial features are located with a projection method that Kanade successfully used to locate the boundary of a face [71]. Let $I(x, y)$ be the intensity value of an $m \times n$ image at position (x, y) , the horizontal and vertical projections of the image are defined as $HI(x) = \sum_{y=1}^n I(x, y)$ and $VI(y) = \sum_{x=1}^m I(x, y)$. The horizontal profile of an input image is obtained first, and then the two local minima, determined by detecting abrupt changes in HI , are said to correspond to the left and right side of the head. Similarly, the vertical profile is obtained and the local minima are determined for the locations of mouth lips, nose tip, and eyes. These detected features constitute a facial candidate. Fig. 3a shows one example where the boundaries

of the face correspond to the local minimum where abrupt intensity changes occur. Subsequently, eyebrow/eyes, nostrils/nose, and the mouth detection rules are used to validate these candidates. The proposed method has been tested using a set of faces in frontal views extracted from the European ACTS M2VTS (MultiModal Verification for Teleservices and Security applications) database [116] which contains video sequences of 37 different people. Each image sequence contains only one face in a uniform background. Their method provides correct face candidates in all tests. The detection rate is 86.5 percent if successful detection is defined as correctly identifying all facial features. Fig. 3b shows one example in which it becomes difficult to locate a face in a complex background using the horizontal and vertical profiles. Furthermore, this method cannot readily detect multiple faces as illustrated in Fig. 3c. Essentially, the projection method can be effective if the window over which it operates is suitably located to avoid misleading interference.

2.2 Bottom-Up Feature-Based Methods

In contrast to the knowledge-based top-down approach, researchers have been trying to find invariant features of faces for detection. The underlying assumption is based on the observation that humans can effortlessly detect faces and objects in different poses and lighting conditions and, so, there must exist properties or features which are invariant over these variabilities. Numerous methods have been proposed to first detect facial features and then to infer the presence of a face. Facial features such as eyebrows, eyes, nose, mouth, and hair-line are commonly extracted using edge detectors. Based on the extracted features, a statistical model is built to describe their relationships and to verify the existence of a face. One problem with these feature-based algorithms is that the image features can be severely corrupted due to illumination, noise, and occlusion. Feature boundaries can be weakened for faces, while shadows can cause numerous strong edges which together render perceptual grouping algorithms useless.

2.2.1 Facial Features

Sirohey proposed a localization method to segment a face from a cluttered background for face identification [145]. It uses an edge map (Canny detector [15]) and heuristics to remove and group edges so that only the ones on the face

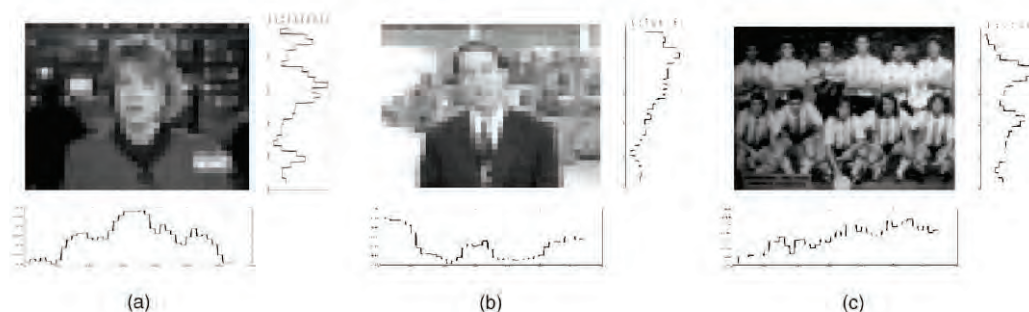


Fig. 3. (a) and (b) $n = 8$. (c) $n = 4$. Horizontal and vertical profiles. It is feasible to detect a single face by searching for the peaks in horizontal and vertical profiles. However, the same method has difficulty detecting faces in complex backgrounds or multiple faces as shown in (b) and (c).
Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.

contour are preserved. An ellipse is then fit to the boundary between the head region and the background. This algorithm achieves 80 percent accuracy on a database of 48 images with cluttered backgrounds. Instead of using edges, Chetverikov and Lerch presented a simple face detection method using blobs and streaks (linear sequences of similarly oriented edges) [20]. Their face model consists of two dark blobs and three light blobs to represent eyes, cheekbones, and nose. The model uses streaks to represent the outlines of the faces, eyebrows, and lips. Two triangular configurations are utilized to encode the spatial relationship among the blobs. A low resolution Laplacian image is generated to facilitate blob detection. Next, the image is scanned to find specific triangular occurrences as candidates. A face is detected if streaks are identified around a candidate.

Graf et al. developed a method to locate facial features and faces in gray scale images [54]. After band pass filtering, morphological operations are applied to enhance regions with high intensity that have certain shapes (e.g., eyes). The histogram of the processed image typically exhibits a prominent peak. Based on the peak value and its width, adaptive threshold values are selected in order to generate two binarized images. Connected components are identified in both binarized images to identify the areas of candidate facial features. Combinations of such areas are then evaluated with classifiers, to determine whether and where a face is present. Their method has been tested with head-shoulder images of 40 individuals and with five video sequences where each sequence consists of 100 to 200 frames. However, it is not clear how morphological operations are performed and how the candidate facial features are combined to locate a face.

Leung et al. developed a probabilistic method to locate a face in a cluttered scene based on local feature detectors and random graph matching [87]. Their motivation is to formulate the face localization problem as a search problem in which the goal is to find the arrangement of certain facial features that is most likely to be a face pattern. Five features (two eyes, two nostrils, and nose/lip junction) are used to describe a typical face. For any pair of facial features of the same type (e.g., left-eye, right-eye pair), their relative distance is computed, and over an ensemble of images the distances are modeled by a Gaussian distribution. A facial template is defined by averaging the responses to a set of multiorientation, multi-scale Gaussian derivative filters (at the pixels inside the facial feature) over a number of faces in a data set. Given a test image, candidate facial features are identified by matching the filter response at each pixel against a template vector of responses (similar to correlation in spirit). The top two feature candidates with the strongest response are selected to search for the other facial features. Since the facial features cannot appear in arbitrary arrangements, the expected locations of the other features are estimated using a statistical model of mutual distances. Furthermore, the covariance of the estimates can be computed. Thus, the expected feature locations can be estimated with high probability. Constellations are then formed only from candidates that lie inside the appropriate locations, and the most face-like constellation is determined. Finding the best constellation is formulated as a random graph matching problem in which the nodes of the

graph correspond to features on a face, and the arcs represent the distances between different features. Ranking of constellations is based on a probability density function that a constellation corresponds to a face versus the probability it was generated by an alternative mechanism (i.e., nonface). They used a set of 150 images for experiments in which a face is considered correctly detected if any constellation correctly locates three or more features on the faces. This system is able to achieve a correct localization rate of 86 percent.

Instead of using mutual distances to describe the relationships between facial features in constellations, an alternative method for modeling faces was also proposed by the Leung et al. [13], [88]. The representation and ranking of the constellations is accomplished using the statistical theory of shape, developed by Kendall [75] and Mardia and Dryden [95]. The shape statistics is a joint probability density function over N feature points, represented by (x_i, y_i) , for the i th feature under the assumption that the original feature points are positioned in the plane according to a general $2N$ -dimensional Gaussian distribution. They applied the same maximum-likelihood (ML) method to determine the location of a face. One advantage of these methods is that partially occluded faces can be located. However, it is unclear whether these methods can be adapted to detect multiple faces effectively in a scene.

In [177], [178], Yow and Cipolla presented a feature-based method that uses a large amount of evidence from the visual image and their contextual evidence. The first stage applies a second derivative Gaussian filter, elongated at an aspect ratio of three to one, to a raw image. Interest points, detected at the local maxima in the filter response, indicate the possible locations of facial features. The second stage examines the edges around these interest points and groups them into regions. The perceptual grouping of edges is based on their proximity and similarity in orientation and strength. Measurements of a region's characteristics, such as edge length, edge strength, and intensity variance, are computed and stored in a feature vector. From the training data of facial features, the mean and covariance matrix of each facial feature vector are computed. An image region becomes a valid facial feature candidate if the Mahalanobis distance between the corresponding feature vectors is below a threshold. The labeled features are further grouped based on model knowledge of where they should occur with respect to each other. Each facial feature and grouping is then evaluated using a Bayesian network. One attractive aspect is that this method can detect faces at different orientations and poses. The overall detection rate on a test set of 110 images of faces with different scales, orientations, and viewpoints is 85 percent [179]. However, the reported false detection rate is 28 percent and the implementation is only effective for faces larger than 60×60 pixels. Subsequently, this approach has been enhanced with active contour models [22], [179]. Fig. 4 summarizes their feature-based face detection method.

Takacs and Wechsler described a biologically motivated face localization method based on a model of retinal feature extraction and small oscillatory eye movements [157]. Their algorithm operates on the conspicuity map or region of interest, with a retina lattice modeled after the magnocellular

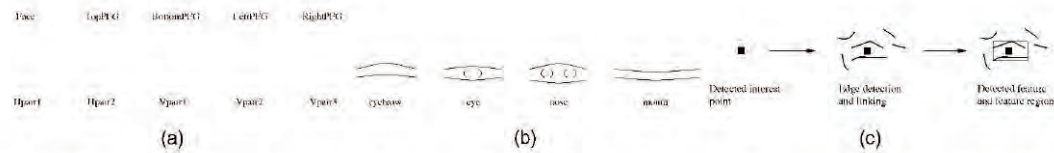


Fig. 4. (a) Yow and Cipolla model a face as a plane with six oriented facial features (eyebrows, eyes, nose, and mouth) [179]. (b) Each facial feature is modeled as pairs of oriented edges. (c) The feature selection process starts with interest points, followed by edge detection and linking, and tested by a statistical model (Courtesy of K.C. Yow and R. Cipolla).

ganglion cells in the human vision system. The first phase computes a coarse scan of the image to estimate the location of the face, based on the filter responses of receptive fields. Each receptive field consists of a number of neurons which are implemented with Gaussian filters tuned to specific orientations. The second phase refines the conspicuity map by scanning the image area at a finer resolution to localize the face. The error rate on a test set of 426 images (200 subjects from the FERET database) is 4.69 percent.

Han et al. developed a morphology-based technique to extract what they call eye-analogue segments for face detection [58]. They argue that eyes and eyebrows are the most salient and stable features of human face and, thus, useful for detection. They define eye-analogue segments as edges on the contours of eyes. First, morphological operations such as closing, clipped difference, and thresholding are applied to extract pixels at which the intensity values change significantly. These pixels become the eye-analogue pixels in their approach. Then, a labeling process is performed to generate the eye-analogue segments. These segments are used to guide the search for potential face regions with a geometrical combination of eyes, nose, eyebrows and mouth. The candidate face regions are further verified by a neural network similar to [127]. Their experiments demonstrate a 94 percent accuracy rate using a test set of 122 images with 130 faces.

Recently, Amit et al. presented a method for shape detection and applied it to detect frontal-view faces in still intensity images [3]. Detection follows two stages: focusing and intensive classification. Focusing is based on spatial arrangements of edge fragments extracted from a simple edge detector using intensity difference. A rich family of such spatial arrangements, invariant over a range of photometric and geometric transformations, is defined. From a set of 300 training face images, particular spatial arrangements of edges which are more common in faces than backgrounds are selected using an inductive method developed in [4]. Meanwhile, the CART algorithm [11] is applied to grow a classification tree from the training images and a collection of false positives identified from generic background images. Given a test image, regions of interest are identified from the spatial arrangements of edge fragments. Each region of interest is then classified as face or background using the learned CART tree. Their experimental results on a set of 100 images from the Olivetti (now AT&T) data set [136] report a false positive rate of 0.2 percent per 1,000 pixels and a false negative rate of 10 percent.

Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.

2.2.2 Texture

Human faces have a distinct texture that can be used to separate them from different objects. Augusteijn and Skufca developed a method that infers the presence of a face through the identification of face-like textures [6]. The texture are computed using second-order statistical features (SGLD) [59] on subimages of 16×16 pixels. Three types of features are considered: skin, hair, and others. They used a cascade correlation neural network [41] for supervised classification of textures and a Kohonen self-organizing feature map [80] to form clusters for different texture classes. To infer the presence of a face from the texture labels, they suggest using votes of the occurrence of hair and skin textures. However, only the result of texture classification is reported, not face localization or detection.

Dai and Nakano also applied SGLD model to face detection [32]. Color information is also incorporated with the face-texture model. Using the face texture model, they design a scanning scheme for face detection in color scenes in which the orange-like parts including the face areas are enhanced. One advantage of this approach is that it can detect faces which are not upright or have features such as beards and glasses. The reported detection rate is perfect for a test set of 30 images with 60 faces.

2.2.3 Skin Color

Human skin color has been used and proven to be an effective feature in many applications from face detection to hand tracking. Although different people have different skin color, several studies have shown that the major difference lies largely between their intensity rather than their chrominance [54], [55], [172]. Several color spaces have been utilized to label pixels as skin including RGB [66], [67], [137], normalized RGB [102], [29], [149], [172], [30], [105], [171], [77], [151], [120], HSV (or HSI) [138], [79], [147], [146], YCrCb [167], [17], YIQ [31], [32], YES [131], CIE XYZ [19], and CIE LUV [173].

Many methods have been proposed to build a skin color model. The simplest model is to define a region of skin tone pixels using Cr, Cb values [17], i.e., $R(Cr, Cb)$, from samples of skin color pixels. With carefully chosen thresholds, $[Cr_1, Cr_2]$ and $[Cb_1, Cb_2]$, a pixel is classified to have skin tone if its values (Cr, Cb) fall within the ranges, i.e., $Cr_1 \leq Cr \leq Cr_2$ and $Cb_1 \leq Cb \leq Cb_2$. Crowley and Coutaz used a histogram $h(r, g)$ of (r, g) values in normalized RGB color space to obtain the probability of obtaining a particular RGB-vector given that the pixel observes skin [29], [30]. In other words, a pixel is classified to belong to skin color if $h(r, g) \geq \tau$.

where τ is a threshold selected empirically from the histogram of samples. Saxe and Foulds proposed an iterative skin identification method that uses histogram intersection in HSV color space [138]. An initial patch of skin color pixels, called the control seed, is chosen by the user and is used to initiate the iterative algorithm. To detect skin color regions, their method moves through the image, one patch at a time, and presents the control histogram and the current histogram from the image for comparison. Histogram intersection [155] is used to compare the control histogram and current histogram. If the match score or number of instances in common (i.e., intersection) is greater than a threshold, the current patch is classified as being skin color. Kjeldsen and Kender defined a color predicate in HSV color space to separate skin regions from background [79]. In contrast to the nonparametric methods mentioned above, Gaussian density functions [14], [77], [173] and a mixture of Gaussians [66], [67], [174] are often used to model skin color. The parameters in a unimodal Gaussian distribution are often estimated using maximum-likelihood [14], [77], [173]. The motivation for using a mixture of Gaussians is based on the observation that the color histogram for the skin of people with different ethnic background does not form a unimodal distribution, but rather a multimodal distribution. The parameters in a mixture of Gaussians are usually estimated using an EM algorithm [66], [174]. Recently, Jones and Rehg conducted a large-scale experiment in which nearly 1 billion labeled skin tone pixels are collected (in normalized RGB color space) [69]. Comparing the performance of histogram and mixture models for skin detection, they find histogram models to be superior in accuracy and computational cost.

Color information is an efficient tool for identifying facial areas and specific facial features if the skin color model can be properly adapted for different lighting environments. However, such skin color models are not effective where the spectrum of the light source varies significantly. In other words, color appearance is often unstable due to changes in both background and foreground lighting. Though the color constancy problem has been addressed through the formulation of physics-based models [45], several approaches have been proposed to use skin color in varying lighting conditions. McKenna et al. presented an adaptive color mixture model to track faces under varying illumination conditions [99]. Instead of relying on a skin color model based on color constancy, they used a stochastic model to estimate an object's color distribution online and adapt to accommodate changes in the viewing and lighting conditions. Preliminary results show that their system can track faces within a range of illumination conditions. However, this method cannot be applied to detect faces in a single image.

Skin color alone is usually not sufficient to detect or track faces. Recently, several modular systems using a combination of shape analysis, color segmentation, and motion information for locating or tracking heads and faces in an image sequence have been developed [55], [173], [172], [99], [147]. We review these methods in the next section.

2.2.4 Multiple Features

Recently, numerous methods that combine several facial features have been proposed to locate or detect faces. Most of them utilize global features such as skin color, size, and shape

to find face candidates, and then verify these candidates using local, detailed features such as eye brows, nose, and hair. A typical approach begins with the detection of skin-like regions as described in Section 2.2.3. Next, skin-like pixels are grouped together using connected component analysis or clustering algorithms. If the shape of a connected region has an elliptic or oval shape, it becomes a face candidate. Finally, local features are used for verification. However, others, such as [7], [63], have used different sets of features.

Yachida et al. presented a method to detect faces in color images using fuzzy theory [19], [169], [168]. They used two fuzzy models to describe the distribution of skin and hair color in CIE XYZ color space. Five (one frontal and four side views) head-shape models are used to abstract the appearance of faces in images. Each shape model is a 2D pattern consisting of $m \times n$ square cells where each cell may contain several pixels. Two properties are assigned to each cell: the skin proportion and the hair proportion, which indicate the ratios of the skin area (or the hair area) within the cell to the area of the cell. In a test image, each pixel is classified as hair, face, hair/face, and hair/background based on the distribution models, thereby generating skin-like and hair-like regions. The head shape models are then compared with the extracted skin-like and hair-like regions in a test image. If they are similar, the detected region becomes a face candidate. For verification, eye-eyebrow and nose-mouth features are extracted from a face candidate using horizontal edges.

Sobottka and Pitas proposed a method for face localization and facial feature extraction using shape and color [147]. First, color segmentation in HSV space is performed to locate skin-like regions. Connected components are then determined by region growing at a coarse resolution. For each connected component, the best fit ellipse is computed using geometric moments. Connected components that are well approximated by an ellipse are selected as face candidates. Subsequently, these candidates are verified by searching for facial features inside of the connected components. Features, such as eyes and mouths, are extracted based on the observation that they are darker than the rest of a face. In [159], [160], a Gaussian skin color model is used to classify skin color pixels. To characterize the shape of the clusters in the binary image, a set of 11 lowest-order geometric moments is computed using Fourier and radial Mellin transforms. For detection, a neural network is trained with the extracted geometric moments. Their experiments show a detection rate of 85 percent based on a test set of 100 images.

The symmetry of face patterns has also been applied to face localization [131]. Skin/nonskin classification is carried out using the class-conditional density function in YES color space followed by smoothing in order to yield contiguous regions. Next, an elliptical face template is used to determine the similarity of the skin color regions based on Hausdorff distance [65]. Finally, the eye centers are localized using several cost functions which are designed to take advantage of the inherent symmetries associated with face and eye locations. The tip of the nose and the center of the mouth are then located by utilizing the distance between the eye centers. One drawback is that it is effective only for a single frontal-view face and when both

eyes are visible. A similar method using color and local symmetry was presented in [151].

In contrast to pixel-based methods, a detection method based on structure, color, and geometry was proposed in [173]. First, multiscale segmentation [2] is performed to extract homogeneous regions in an image. Using a Gaussian skin color model, regions of skin tone are extracted and grouped into ellipses. A face is detected if facial features such as eyes and mouth exist within these elliptic regions. Experimental results show that this method is able to detect faces at different orientations with facial features such as beard and glasses.

Kauth et al. proposed a blob representation to extract a compact, structurally meaningful description of multispectral satellite imagery [74]. A feature vector at each pixel is formed by concatenating the pixel's image coordinates to the pixel's spectral (or textural) components; pixels are then clustered using this feature vector to form coherent connected regions, or "blobs." To detect faces, each feature vector consists of the image coordinates and normalized chrominance, i.e., $X = (x, y, \frac{r}{r+g+b}, \frac{b}{r+g+b})$ [149], [105]. A connectivity algorithm is then used to grow blobs and the resulting skin blob whose size and shape is closest to that of a canonical face is considered as a face.

Range and color have also been employed for face detection by Kim et al. [77]. Disparity maps are computed and objects are segmented from the background with a disparity histogram using the assumption that background pixels have the same depth and they outnumber the pixels in the foreground objects. Using a Gaussian distribution in normalized RGB color space, segmented regions with a skin-like color are classified as faces. A similar approach has been proposed by Darrell et al. for face detection and tracking [33].

2.3 Template Matching

In template matching, a standard face pattern (usually frontal) is manually predefined or parameterized by a function. Given an input image, the correlation values with the standard patterns are computed for the face contour, eyes, nose, and mouth independently. The existence of a face is determined based on the correlation values. This approach has the advantage of being simple to implement. However, it has proven to be inadequate for face detection since it cannot effectively deal with variation in scale, pose, and shape. Multiresolution, multiscale, subtemplates, and deformable templates have subsequently been proposed to achieve scale and shape invariance.

2.3.1 Predefined Templates

An early attempt to detect frontal faces in photographs is reported by Sakai et al. [132]. They used several subtemplates for the eyes, nose, mouth, and face contour to model a face. Each subtemplate is defined in terms of line segments. Lines in the input image are extracted based on greatest gradient change and then matched against the subtemplates. The correlations between subimages and contour templates are computed first to detect candidate locations of faces. Then, matching with the other subtemplates is performed at the candidate positions. In other words, the first phase determines focus of attention or region of interest and the second phase examines the details to determine the existence of a

face. The idea of focus of attention and subtemplates has been adopted by later works on face detection.

Craw et al. presented a localization method based on a shape template of a frontal-view face (i.e., the outline shape of a face) [27]. A Sobel filter is first used to extract edges. These edges are grouped together to search for the template of a face based on several constraints. After the head contour has been located, the same process is repeated at different scales to locate features such as eyes, eyebrows, and lips. Later, Craw et al. describe a localization method using a set of 40 templates to search for facial features and a control strategy to guide and assess the results from the template-based feature detectors [28].

Govindaraju et al. presented a two stage face detection method in which face hypotheses are generated and tested [52], [53], [51]. A face model is built in terms of features defined by the edges. These features describe the curves of the left side, the hair-line, and the right side of a frontal face. The Marr-Hildreth edge operator is used to obtain an edge map of an input image. A filter is then used to remove objects whose contours are unlikely to be part of a face. Pairs of fragmented contours are linked based on their proximity and relative orientation. Corners are detected to segment the contour into feature curves. These feature curves are then labeled by checking their geometric properties and relative positions in the neighborhood. Pairs of feature curves are joined by edges if their attributes are compatible (i.e., if they could arise from the same face). The ratios of the feature pairs forming an edge is compared with the golden ratio and a cost is assigned to the edge. If the cost of a group of three feature curves (with different labels) is low, the group becomes a hypothesis. When detecting faces in newspaper articles, collateral information, which indicates the number of persons in the image, is obtained from the caption of the input image to select the best hypotheses [52]. Their system reports a detection rate of approximately 70 percent based on a test set of 50 photographs. However, the faces must be upright, unoccluded, and frontal. The same approach has been extended by extracting edges in the wavelet domain by Venkatraman and Govindaraju [165].

Tsukamoto et al. presented a qualitative model for face pattern (QMF) [161], [162]. In QMF, each sample image is divided into a number of blocks, and qualitative features are estimated for each block. To parameterize a face pattern, "lightness" and "edgeness" are defined as the features in this model. Consequently, this blocked template is used to calculate "faceness" at every position of an input image. A face is detected if the faceness measure is above a predefined threshold.

Silhouettes have also been used as templates for face localization [134]. A set of basis face silhouettes is obtained using principal component analysis (PCA) on face examples in which the silhouette is represented by an array of bits. These eigen-silhouettes are then used with a generalized Hough transform for localization. A localization method based on multiple templates for facial components was proposed in [150]. Their method defines numerous hypotheses for the possible appearances of facial features. A set of hypotheses for the existence of a face is then defined in terms of the hypotheses for facial components using the Dempster-Shafer theory [34]. Given an image, feature detectors compute

Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.

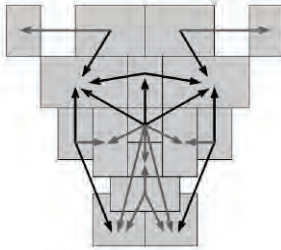


Fig. 5. A 14x16 pixel ratio template for face localization based on Sinha method. The template is composed of 16 regions (the gray boxes) and 23 relations (shown by arrows) [139] (Courtesy of B. Scassellati).

confidence factors for the existence of facial features. The confidence factors are combined to determine the measures of belief and disbelief about the existence of a face. Their system is able to locate faces in 88 images out of 94 images.

Sinha used a small set of spatial image invariants to describe the space of face patterns [143], [144]. His key insight for designing the invariant is that, while variations in illumination change the individual brightness of different parts of faces (such as eyes, cheeks, and forehead), the relative brightness of these parts remain largely unchanged. Determining pairwise ratios of the brightness of a few such regions and retaining just the “directions” of these ratios (i.e., Is one region brighter or darker than the other?) provides a robust invariant. Thus, observed brightness regularities are encoded as a ratio template which is a coarse spatial template of a face with a few appropriately chosen subregions that roughly correspond to key facial features such as the eyes, cheeks, and forehead. The brightness constraints between facial parts are captured by an appropriate set of pairwise brighter-darker relationships between subregions. A face is located if an image satisfies all the pairwise brighter-darker constraints. The idea of using intensity differences between local adjacent regions has later been extended to a wavelet-based representation for pedestrian, car, and face detection [109]. Sinha’s method has been extended and applied to face localization in an active robot vision system [139], [10]. Fig. 5 shows the enhanced template with 23 defined relations. These defined relations are further classified into 11 essential relations (solid arrows) and 12 confirming relations (dashed arrows). Each arrow in the figure indicates a relation, with the head of the arrow denoting the second region (i.e., the denominator of the ratio). A relation is satisfied for face template if the ratio between two regions exceeds a threshold and a face is localized if the number of essential and confirming relations exceeds a threshold.

A hierarchical template matching method for face detection was proposed by Miao et al. [100]. At the first stage, an input image is rotated from -20° to 20° in steps of 5° , in order to handle rotated faces. A multiresolution image hierarchy is formed (See Fig. 1) and edges are extracted using the Laplacian operator. The face template consists of the edges produced by six facial components: two eyebrows, two eyes, one nose, and one mouth. Finally, heuristics are applied to determine the existence of a face. Their experimental results show better results in images containing a single face (frontal or rotated) than in images with multiple faces.

Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.

2.3.2 Deformable Templates

Yuille et al. used deformable templates to model facial features that fit an a priori elastic model to facial features (e.g., eyes) [180]. In this approach, facial features are described by parameterized templates. An energy function is defined to link edges, peaks, and valleys in the input image to corresponding parameters in the template. The best fit of the elastic model is found by minimizing an energy function of the parameters. Although their experimental results demonstrate good performance in tracking nonrigid features, one drawback of this approach is that the deformable template must be initialized in the proximity of the object of interest.

In [84], a detection method based on snakes [73], [90] and templates was developed. An image is first convolved with a blurring filter and then a morphological operator to enhance edges. A modified n -pixel (n is small) snake is used to find and eliminate small curve segments. Each face is approximated by an ellipse and a Hough transform of the remaining snakelets is used to find a dominant ellipse. Thus, sets of four parameters describing the ellipses are obtained and used as candidates for face locations. For each of these candidates, a method similar to the deformable template method [180] is used to find detailed features. If a substantial number of the facial features are found and if their proportions satisfy ratio tests based on a face template, a face is considered to be detected. Lam and Yan also used snakes to locate the head boundaries with a greedy algorithm in minimizing the energy function [85].

Lanitis et al. described a face representation method with both shape and intensity information [86]. They start with sets of training images in which sampled contours such as the eye boundary, nose, chin/cheek are manually labeled, and a vector of sample points is used to represent shape. They used a point distribution model (PDM) to characterize the shape vectors over an ensemble of individuals, and an approach similar to Kirby and Sirovich [78] to represent shape-normalized intensity appearance. A face-shape PDM can be used to locate faces in new images by using active shape model (ASM) search to estimate the face location and shape parameters. The face patch is then deformed to the average shape, and intensity parameters are extracted. The shape and intensity parameters can be used together for classification. Cootes and Taylor applied a similar approach to localize a face in an image [25]. First, they define rectangular regions of the image containing instances of the feature of interest. Factor analysis [5] is then applied to fit these training features and obtain a distribution function. Candidate features are determined if the probabilistic measures are above a threshold and are verified using the ASM. After training this method with 40 images, it is able to locate 35 faces in 40 test images. The ASM approach has also been extended with two Kalman filters to estimate the shape-free intensity parameters and to track faces in image sequences [39].

2.4 Appearance-Based Methods

Contrasted to the template matching methods where templates are predefined by experts, the “templates” in appearance-based methods are learned from examples in images. In general, appearance-based methods rely on techniques from statistical analysis and machine learning to find the relevant

characteristics of face and nonface images. The learned characteristics are in the form of distribution models or discriminant functions that are consequently used for face detection. Meanwhile, dimensionality reduction is usually carried out for the sake of computation efficiency and detection efficacy.

Many appearance-based methods can be understood in a probabilistic framework. An image or feature vector derived from an image is viewed as a random variable $\mathbf{x}$, and this random variable is characterized for faces and nonfaces by the class-conditional density functions $p(\mathbf{x}|\text{face})$ and $p(\mathbf{x}|\text{nonface})$. Bayesian classification or maximum likelihood can be used to classify a candidate image location as face or nonface. Unfortunately, a straightforward implementation of Bayesian classification is infeasible because of the high dimensionality of $\mathbf{x}$, because $p(\mathbf{x}|\text{face})$ and $p(\mathbf{x}|\text{nonface})$ are multimodal, and because it is not yet understood if there are natural parameterized forms for $p(\mathbf{x}|\text{face})$ and $p(\mathbf{x}|\text{nonface})$. Hence, much of the work in an appearance-based method concerns empirically validated parametric and nonparametric approximations to $p(\mathbf{x}|\text{face})$ and $p(\mathbf{x}|\text{nonface})$.

Another approach in appearance-based methods is to find a discriminant function (i.e., decision surface, separating hyperplane, threshold function) between face and nonface classes. Conventionally, image patterns are projected to a lower dimensional space and then a discriminant function is formed (usually based on distance metrics) for classification [163], or a nonlinear decision surface can be formed using multilayer neural networks [128]. Recently, support vector machines and other kernel methods have been proposed. These methods implicitly project patterns to a higher dimensional space and then form a decision surface between the projected face and nonface patterns [107].

2.4.1 Eigenfaces

An early example of employing eigenvectors in face recognition was done by Kohonen [80] in which a simple neural network is demonstrated to perform face recognition for aligned and normalized face images. The neural network computes a face description by approximating the eigenvectors of the image's autocorrelation matrix. These eigenvectors are later known as Eigenfaces.

Kirby and Sirovich demonstrated that images of faces can be linearly encoded using a modest number of basis images [78]. This demonstration is based on the Karhunen-Loève transform [72], [93], [48], which also goes by other names, e.g., principal component analysis [68], and the Hotelling transform [50]. The idea is arguably proposed first by Pearson in 1901 [110] and then by Hotelling in 1933 [62]. Given a collection of n by m pixel training images represented as a vector of size $m \times n$, basis vectors spanning an optimal subspace are determined such that the mean square error between the projection of the training images onto this subspace and the original images is minimized. They call the set of optimal basis vectors eigenpictures since these are simply the eigenvectors of the covariance matrix computed from the vectorized face images in the training set. Experiments with a set of 100 images show that a face image of 91×50 pixels can be effectively encoded using only 50 eigenpictures, while retaining a reasonable likeness (i.e., capturing 95 percent of the variance).

Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.

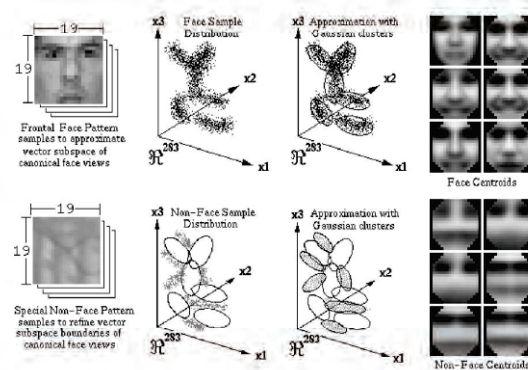


Fig. 6. Face and nonface clusters used by Sung and Poggio [154]. Their method estimates density functions for face and nonface patterns using a set of Gaussians. The centers of these Gaussians are shown on the right (Courtesy of K.-K. Sung and T. Poggio).

Turk and Pentland applied principal component analysis to face recognition and detection [163]. Similar to [78], principal component analysis on a training set of face images is performed to generate the Eigenpictures (here called Eigenfaces) which span a subspace (called the face space) of the image space. Images of faces are projected onto the subspace and clustered. Similarly, nonface training images are projected onto the same subspace and clustered. Since images of faces do not change radically when projected onto the face space, while the projection of nonface images appear quite different. To detect the presence of a face in a scene, the distance between an image region and the face space is computed for all locations in the image. The distance from face space is used as a measure of "faceness," and the result of calculating the distance from face space is a "face map." A face can then be detected from the local minima of the face map. Many works on face detection, recognition, and feature extractions have adopted the idea of eigenvector decomposition and clustering.

2.4.2 Distribution-Based Methods

Sung and Poggio developed a distribution-based system for face detection [152], [154] which demonstrated how the distributions of image patterns from one object class can be learned from positive and negative examples (i.e., images) of that class. Their system consists of two components, distribution-based models for face/nonface patterns and a multilayer perceptron classifier. Each face and nonface example is first normalized and processed to a 19×19 pixel image and treated as a 361-dimensional vector or pattern. Next, the patterns are grouped into six face and six nonface clusters using a modified k -means algorithm, as shown in Fig. 6. Each cluster is represented as a multidimensional Gaussian function with a mean image and a covariance matrix. Fig. 7 shows the distance measures in their method. Two distance metrics are computed between an input image pattern and the prototype clusters. The first distance component is the normalized Mahalanobis distance between the test pattern and the cluster centroid, measured within a lower-dimensional subspace spanned by the cluster's 75 largest eigenvectors. The second distance component is the Euclidean distance between the test pattern and its projection

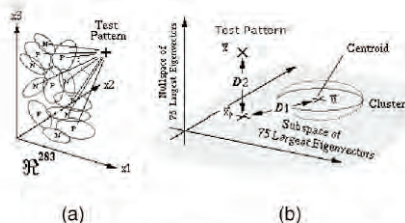


Fig. 7. The distance measures used by Sung and Poggio [154]. Two distance metrics are computed between an input image pattern and the prototype clusters. (a) Given a test pattern, the distance between that image pattern and each cluster is computed. A set of 12 distances between the test pattern and the model's 12 cluster centroids. (b) Each distance measurement between the test pattern and a cluster centroid is a two-value distance metric. D_1 is a Mahalanobis distance between the test pattern's projection and the cluster centroid in a subspace spanned by the cluster's 75 largest eigenvectors. D_2 is the Euclidean distance between the test pattern and its projection in the subspace. Therefore, a distance vector of 24 values is formed for each test pattern and is used by a multilayer perceptron to determine whether the input pattern belongs to the face class or not (Courtesy of K.-K. Sung and T. Poggio).

onto the 75-dimensional subspace. This distance component accounts for pattern differences not captured by the first distance component. The last step is to use a multilayer perceptron (MLP) network to classify face window patterns from nonface patterns using the twelve pairs of distances to each face and nonface cluster. The classifier is trained using standard backpropagation from a database of 47,316 window patterns. There are 4,150 positive examples of face patterns and the rest are nonface patterns. Note that it is easy to collect a representative sample face patterns, but much more difficult to get a representative sample of nonface patterns. This problem is alleviated by a bootstrap method that selectively adds images to the training set as training progress. Starting with a small set of nonface examples in the training set, the MLP classifier is trained with this database of examples. Then, they run the face detector on a sequence of random images and collect all the nonface patterns that the current system wrongly classifies as faces. These false positives are then added to the training database as new nonface examples. This bootstrap method avoids the problem of explicitly collecting a representative sample of nonface patterns and has been used in later works [107], [128].

A probabilistic visual learning method based on density estimation in a high-dimensional space using an eigenspace decomposition was developed by Moghaddam and Pentland [103]. Principal component analysis (PCA) is used to define the subspace best representing a set of face patterns. These principal components preserve the major linear correlations in the data and discard the minor ones. This method decomposes the vector space into two mutually exclusive and complementary subspaces: the principal subspace (or feature space) and its orthogonal complement. Therefore, the target density is decomposed into two components: the density in the principal subspace (spanned by the principal components) and its orthogonal complement (which is discarded in standard PCA) (See Fig. 8). A multivariate Gaussian and a mixture of Gaussians are used to learn the statistics of the local features of a face. These probability densities are then used for object detection based on maximum likelihood estimation. The proposed method has been applied to face localization, coding, and recognition.

Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.

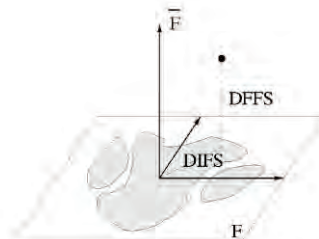


Fig. 8. Decomposition of a face image space into the principal subspace F and its orthogonal complement $\bar{F}$ for an arbitrary density. Every data point x is decomposed into two components: distance in feature space (DIFS) and distance from feature space (DFFS) [103] (Courtesy of B. Moghaddam and A. Pentland).

Compared with the classic eigenface approach [163], the proposed method shows better performance in face recognition. In terms of face detection, this technique has only been demonstrated on localization; see also [76].

In [175], a detection method based on a mixture of factor analyses was proposed. Factor analysis (FA) is a statistical method for modeling the covariance structure of high dimensional data using a small number of latent variables. FA is analogous to principal component analysis (PCA) in several aspects. However, PCA, unlike FA, does not define a proper density model for the data since the cost of coding a data point is equal anywhere along the principal component subspace (i.e., the density is unnormalized along these directions). Further, PCA is not robust to independent noise in the features of the data since the principal components maximize the variances of the input data, thereby retaining unwanted variations. Synthetic and real examples in [36], [37], [9], [7] have shown that the projected samples from different classes in the PCA subspace can often be smeared. For the cases where the samples have certain structure, PCA is suboptimal from the classification standpoint. Hinton et al. have applied FA to digit recognition, and they compare the performance of PCA and FA models [61]. A mixture model of factor analyzers has recently been extended [49] and applied to face recognition [46]. Both studies show that FA performs better than PCA in digit and face recognition. Since pose, orientation, expression, and lighting affect the appearance of a human face, the distribution of faces in the image space can be better represented by a multimodal density model where each modality captures certain characteristics of certain face appearances. They present a probabilistic method that uses a mixture of factor analyzers (MFA) to detect faces with wide variations. The parameters in the mixture model are estimated using an EM algorithm.

A second method in [175] uses Fisher's Linear Discriminant (FLD) to project samples from the high dimensional image space to a lower dimensional feature space. Recently, the Fisherface method [7] and others [156], [181] based on linear discriminant analysis have been shown to outperform the widely used Eigenface method [163] in face recognition on several data sets, including the Yale face database where face images are taken under varying lighting conditions. One possible explanation is that FLD provides a better projection than PCA for pattern classification since it aims to find the most discriminant projection direction. Consequently, the classification results in the projected subspace may be



Fig. 9. Prototype of each face class using Kohonen's SOM by Yang et al. [175]. Each prototype corresponds to the center of a cluster.

superior than other methods. (See [97] for a discussion about training set size.) In the second proposed method, they decompose the training face and nonface samples into several subclasses using Kohonen's Self Organizing Map (SOM) [80]. Fig. 9 shows a prototype of each face class. From these relabeled samples, the within-class and between-class scatter matrices are computed, thereby generating the optimal projection based on FLD. For each subclass, its density is modeled as a Gaussian whose parameters are estimated using maximum-likelihood [36]. To detect faces, each input image is scanned with a rectangular window in which the class-dependent probability is computed. The maximum-likelihood decision rule is used to determine whether a face is detected or not. Both methods in [175] have been tested using the databases in [128], [154] which together consist of 225 images with 619 faces, and experimental results show that these two methods have detection rates of 92.3 percent for MFA and 93.6 percent for the FLD-based method.

2.4.3 Neural Networks

Neural networks have been applied successfully in many pattern recognition problems, such as optical character recognition, object recognition, and autonomous robot driving. Since face detection can be treated as a two class pattern recognition problem, various neural network architectures have been proposed. The advantage of using neural networks for face detection is the feasibility of training a system to capture the complex class conditional density of face patterns. However, one drawback is that the network architecture has to be extensively tuned (number of layers, number of nodes, learning rates, etc.) to get exceptional performance.

An early method using hierarchical neural networks was proposed by Agui et al. [1]. The first stage consists of two parallel subnetworks in which the inputs are intensity values from an original image and intensity values from filtered image using a 3×3 Sobel filter. The inputs to the second stage network consist of the outputs from the subnetworks and extracted feature values such as the standard deviation of the pixel values in the input pattern, a ratio of the number of white pixels to the total number of binarized pixels in a window, and geometric moments. An output value at the second stage indicates the presence of a face in the input region. Experimental results show that this method is able to detect faces if all faces in the test images have the same size.

Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.

Propp and Samal developed one of the earliest neural networks for face detection [117]. Their network consists of four layers with 1,024 input units, 256 units in the first hidden layer, eight units in the second hidden layer, and two output units. A similar hierarchical neural network is later proposed by [70]. The early method by Soulie et al. [148] scans an input image with a time-delay neural network [166] (with a receptive field of 20×25 pixels) to detect faces. To cope with size variation, the input image is decomposed using wavelet transforms. They reported a false negative rate of 2.7 percent and false positive rate of 0.5 percent from a test of 120 images. In [164], Vaillant et al. used convolutional neural networks to detect faces in images. Examples of face and nonface images of 20×20 pixels are first created. One neural network is trained to find approximate locations of faces at some scale. Another network is trained to determine the exact position of faces at some scale. Given an image, areas which may contain faces are selected as face candidates by the first network. These candidates are verified by the second network. Burel and Carel [12] proposed a neural network for face detection in which the large number of training examples of faces and nonfaces are compressed into fewer examples using a Kohonen's SOM algorithm [80]. A multilayer perceptron is used to learn these examples for face/background classification. The detection phase consists of scanning each image at various resolution. For each location and size of the scanning window, the contents are normalized to a standard size, and the intensity mean and variance are scaled to reduce the effects of lighting conditions. Each normalized window is then classified by an MLP.

Feraud and Bernier presented a detection method using autoassociative neural networks [43], [42], [44]. The idea is based on [83] which shows an autoassociative network with five layers is able to perform a nonlinear principal component analysis. One autoassociative network is used to detect frontal-view faces and another one is used to detect faces turned up to 60 degrees to the left and right of the frontal view. A gating network is also utilized to assign weights to frontal and turned face detectors in an ensemble of autoassociative networks. On a small test set of 42 images, they report a detection rate similar to [126]. The method has also been employed in LISTEN [23] and MULTRAK [8].

Lin et al. presented a face detection system using probabilistic decision-based neural network (PDBNN) [91]. The architecture of PDBNN is similar to a radial basis function (RBF) network with modified learning rules and probabilistic interpretation. Instead of converting a whole face image into a training vector of intensity values for the neural network, they first extract feature vectors based on intensity and edge information in the facial region that contains eyebrows, eyes, and nose. The extracted two feature vectors are fed into two PDBNN's and the fusion of the outputs determine the classification result. Based on a set of 23 images provided by Sung and Poggio [154], their experimental results show comparable performance with the other leading neural network-based face detectors [154], [128].

Among all the face detection methods that used neural networks, the most significant work is arguably done by Rowley et al. [127], [126], [128]. A multilayer neural network is used to learn the face and nonface patterns from face/

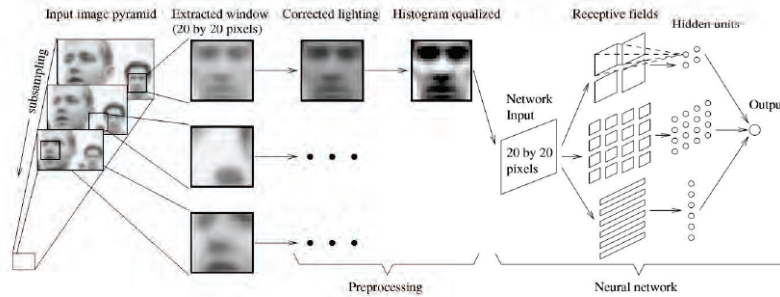


Fig. 10. System diagram of Rowley's method [128]. Each face is preprocessed before feeding it to an ensemble of neural networks. Several arbitration methods are used to determine whether a face exists based on the output of these networks (Courtesy of H. Rowley, S. Baluja, and T. Kanade).

nonface images (i.e., the intensities and spatial relationships of pixels) whereas Sung [152] used a neural network to find a discriminant function to classify face and nonface patterns using distance measures. They also used multiple neural networks and several arbitration methods to improve performance, while Burel and Carel [12] used a single network, and Vaillant et al. [164] used two networks for classification. There are two major components: multiple neural networks (to detect face patterns) and a decision-making module (to render the final decision from multiple detection results). As shown in Fig. 10, the first component of this method is a neural network that receives a 20×20 pixel region of an image and outputs a score ranging from -1 to 1. Given a test pattern, the output of the trained neural network indicates the evidence for a nonface (close to -1) or face pattern (close to 1). To detect faces anywhere in an image, the neural network is applied at all image locations. To detect faces larger than 20×20 pixels, the input image is repeatedly subsampled, and the network is applied at each scale. Nearly 1,050 face samples of various sizes, orientations, positions, and intensities are used to train the network. In each training image, the eyes, tip of the nose, corners, and center of the mouth are labeled manually and used to normalize the face to the same scale, orientation, and position. The second component of this method is to merge overlapping detection and arbitrate between the outputs of multiple networks. Simple arbitration schemes such as logic operators (AND/OR) and voting are used to improve performance. Rowley et al. [127] reported several systems with different arbitration schemes that are less computationally expensive than Sung and Poggio's system and have higher detection rates based on a test set of 24 images containing 144 faces.

One limitation of the methods by Rowley [127] and by Sung [152] is that they can only detect upright, frontal faces. Recently, Rowley et al. [129] extended this method to detect rotated faces using a router network which processes each input window to determine the possible face orientation and then rotates the window to a canonical orientation; the rotated window is presented to the neural networks as described above. However, the new system has a lower detection rate on upright faces than the upright detector. Nevertheless, the system is able to detect 76.9 percent of faces over two large test sets with a small number of false positives.

2.4.4 Support Vector Machines

Support Vector Machines (SVMs) were first applied to face detection by Osuna et al. [107]. SVMs can be considered as a new paradigm to train polynomial function, neural networks, or radial basis function (RBF) classifiers. While most methods for training a classifier (e.g., Bayesian, neural networks, and RBF) are based on minimizing the training error, i.e., *empirical risk*, SVMs operates on another induction principle, called *structural risk minimization*, which aims to minimize an upper bound on the expected generalization error. An SVM classifier is a linear classifier where the separating hyperplane is chosen to minimize the expected classification error of the unseen test patterns. This optimal hyperplane is defined by a weighted combination of a small subset of the training vectors, called support vectors. Estimating the optimal hyperplane is equivalent to solving a linearly constrained quadratic programming problem. However, the computation is both time and memory intensive. In [107], Osuna et al. developed an efficient method to train an SVM for large scale problems, and applied it to face detection. Based on two test sets of 10,000,000 test patterns of 19×19 pixels, their system has slightly lower error rates and runs approximately 30 times faster than the system by Sung and Poggio [153]. SVMs have also been used to detect faces and pedestrians in the wavelet domain [106], [108], [109].

2.4.5 Sparse Network of Winnows

Yang et al. proposed a method that uses SNoW learning architecture [125], [16] to detect faces with different features and expressions, in different poses, and under different lighting conditions [176]. They also studied the effect of learning with primitive as well as with multiscale features. SNoW (Sparse Network of Winnows) is a sparse network of linear functions that utilizes the Winnow update rule [92]. It is specifically tailored for learning in domains in which the potential number of features taking part in decisions is very large, but may be unknown a priori. Some of the characteristics of this learning architecture are its sparsely connected units, the allocation of features and links in a data driven way, the decision mechanism, and the utilization of an efficient update rule. In training the SNoW-based face detector, 1,681 face images from Olivetti [136], UMIST [56], Harvard [57], Yale [7], and FERET [115] databases are used to capture the variations in face patterns. To compare with other methods, they report results with two readily

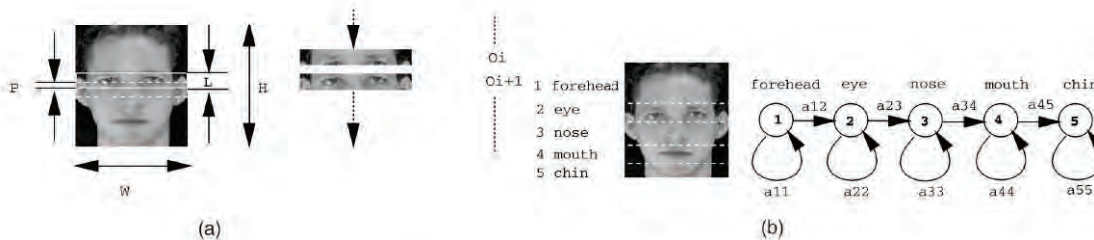


Fig. 11. Hidden Markov model for face localization. (a) Observation vectors: To train an HMM, each face sample is converted to a sequence of observation vectors. Observation vectors are constructed from a window of $W \times L$ pixels. By scanning the window vertically with P pixels of overlap, an observation sequence is constructed. (b) Hidden states: When an HMM with five states is trained with sequences of observation vectors, the boundaries between states are shown in (b) [136].

available data sets which contain 225 images with 619 faces [128]. With an error rate of 5.9 percent, this technique performs as well as other methods evaluated on the data set 1 in [128], including those using neural networks [128], Kullback relative information [24], naive Bayes classifier [140] and support vector machines [107], while being computationally more efficient. See Table 4 for performance comparisons with other face detection methods.

2.4.6 Naive Bayes Classifier

In contrast to the methods in [107], [128], [154] which model the global appearance of a face, Schneiderman and Kanade described a naive Bayes classifier to estimate the joint probability of local appearance and position of face patterns (subregions of the face) at multiple resolutions [140]. They emphasize local appearance because some local patterns of an object are more unique than others; the intensity patterns around the eyes are much more distinctive than the pattern found around the cheeks. There are two reasons for using a naive Bayes classifier (i.e., no statistical dependency between the subregions). First, it provides better estimation of the conditional density functions of these subregions. Second, a naive Bayes classifier provides a functional form of the posterior probability to capture the joint statistics of local appearance and position on the object. At each scale, a face image is decomposed into four rectangular subregions. These subregions are then projected to a lower dimensional space using PCA and quantized into a finite set of patterns, and the statistics of each projected subregion are estimated from the projected samples to encode local appearance. Under this formulation, their method decides that a face is present when the likelihood ratio is larger than the ratio of prior probabilities. With an error rate of 93.0 percent on data set 1 in [128], the proposed Bayesian approach shows comparable performance to [128] and is able to detect some rotated and profile faces. Schneiderman and Kanade later extend this method with wavelet representations to detect profile faces and cars [141].

A related method using joint statistical models of local features was developed by Rickert et al. [124]. Local features are extracted by applying multiscale and multiresolution filters to the input image. The distribution of the features vectors (i.e., filter responses) is estimated by clustering the data and then forming a mixture of Gaussians. After the model is learned and further refined, test images are classified by computing the likelihood of their feature vectors

with respect to the model. Their experimental results on face and car detection show interesting and good results.

2.4.7 Hidden Markov Model

The underlying assumption of the Hidden Markov Model (HMM) is that patterns can be characterized as a parametric random process and that the parameters of this process can be estimated in a precise, well-defined manner. In developing an HMM for a pattern recognition problem, a number of hidden states need to be decided first to form a model. Then, one can train HMM to learn the transitional probability between states from the examples where each example is represented as a sequence of observations. The goal of training an HMM is to maximize the probability of observing the training data by adjusting the parameters in an HMM model with the standard Viterbi segmentation method and Baum-Welch algorithms [122]. After the HMM has been trained, the output probability of an observation determines the class to which it belongs.

Intuitively, a face pattern can be divided into several regions such as the forehead, eyes, nose, mouth, and chin. A face pattern can then be recognized by a process in which these regions are observed in an appropriate order (from top to bottom and left to right). Instead of relying on accurate alignment as in template matching or appearance-based methods (where facial features such as eyes and noses need to be aligned well with respect to a reference point), this approach aims to associate facial regions with the states of a continuous density Hidden Markov Model. HMM-based methods usually treat a face pattern as a sequence of observation vectors where each vector is a strip of pixels, as shown in Fig. 11a. During training and testing, an image is scanned in some order (usually from top to bottom) and an observation is taken as a block of pixels, as shown in Fig. 11a. For face patterns, the boundaries between strips of pixels are represented by probabilistic transitions between states, as shown in Fig. 11b, and the image data within a region is modeled by a multivariate Gaussian distribution. An observation sequence consists of all intensity values from each block. The output states correspond to the classes to which the observations belong. After the HMM has been trained, the output probability of an observation determines the class to which it belongs. HMMs have been applied to both face recognition and localization. Samaria [136] showed that the states of the HMM he trained corresponds to facial regions, as shown in

Fig. 11b. In other words, one state is responsible for characterizing the observation vectors of human foreheads, and another state is responsible for characterizing the observation vectors of human eyes. For face localization, an HMM is trained for a generic model of human faces from a large collection of face images. If the face likelihood obtained for each rectangular pattern in the image is above a threshold, a face is located.

Samaria and Young applied 1D and pseudo 2D HMMs to facial feature extraction and face recognition [135], [136]. Their HMMs exploit the structure of a face to enforce constraints on the state transitions. Since significant facial regions such as hair, forehead, eyes, nose, and mouth occur in the natural order from top to bottom, each of these regions is assigned to a state in a one-dimensional continuous HMM. Fig. 11b shows these five hidden states. For training, each image is uniformly segmented, from top to bottom into five states (i.e., each image is divided into five nonoverlapping regions of equal size). The uniform segmentation is then replaced by the Viterbi segmentation and the parameters in the HMM are reestimated using the Baum-Welch algorithm. As shown in Fig. 11a, each face image of width W and height H is divided into overlapping blocks of height L and width W . There are P rows of overlap between consecutive blocks in the vertical direction. These blocks form an observation sequence for the image, and the trained HMM is used to determine the output state. Similar to [135], Nefian and Hayes applied HMMs and the Karhunen Loève Transform (KLT) to face localization and recognition [104]. Instead of using raw intensity values, the observation vectors consist of the (KLT) coefficients computed from the input vectors. Their experimental results on face recognition show a better recognition rate than [135]. On the MIT database, which contains 432 images each with a single face, this pseudo 2D HMM system has a success rate of 90 percent.

Rajagopalan et al. proposed two probabilistic methods for face detection [123]. In contrast to [154], which uses a set of multivariate Gaussians to model the distribution of face patterns, the first method in [123] uses higher order statistics (HOS) for density estimation. Similar to [154], both the unknown distributions of faces and nonfaces are clustered using six density functions based on higher order statistics of the patterns. As in [152], a multilayer perceptron is used for classification, and the input vector consists of twelve distance measures (i.e., log probability) between the image pattern and the twelve model clusters. The second method in [123] uses an HMM to learn the face to nonface and nonface to face transitions in an image. This approach is based on generating an observation sequence from the image and learning the HMM parameters corresponding to this sequence. The observation sequence to be learned is first generated by computing the distance of the subimage to the centers of the 12 face and nonface cluster centers estimated in the first method. After the learning completes, the optimal state sequence is further processed for binary classification. Experimental results show that both HOS and HMM methods have a higher detection rate than [128], [154], but with more false alarms.

2.4.8 Information-Theoretical Approach

The spatial property of face pattern can be modeled through different aspects. The contextual constraint, among others, is a powerful one and has often been applied to texture segmentation. The contextual constraints in a face pattern are usually specified by a small neighborhood of pixels. Markov random field (MRF) theory provides a convenient and consistent way to model context-dependent entities such as image pixels and correlated features. This is achieved by characterizing mutual influences among such entities using conditional MRF distributions. According to the Hammersley-Clifford theorem, an MRF can be equivalently characterized by a Gibbs distribution and the parameters are usually maximum a posteriori (MAP) estimates [119]. Alternatively, the face and nonface distributions can be estimated using histograms. Using Kullback relative information, the Markov process that maximizes the information-based discrimination between the two classes can be found and applied to detection [89], [24].

Low applied Kullback relative information [26] to face detection by associating a probability function $p(x)$ to the event that the template is a face and $q(x)$ to the event that the template is not a face [89]. A face training database consisting of nine views of 100 individuals is used to estimate the face distribution. The nonface probability density function is estimated from a set of 143,000 nonface templates using histograms. From the training sets, the most informative pixels (MIP) are selected to maximize the Kullback relative information between $p(x)$ and $q(x)$ (i.e., to give the maximum class separation). As it turns out, the MIP distribution focuses on the eye and mouth regions and avoids the nose area. The MIP are then used to obtain linear features for classification and representation using the method of Fukunaga and Koontz [47]. To detect faces, a window is passed over the input image, and the distance from face space (DFFS) as defined in [114] is calculated. If the DFFS to the face subspace is lower than the distance to the nonface subspace, a face is assumed to exist within the window.

Kullback relative information is also employed by Colme-narez and Huang to maximize the information-based discrimination between positive and negative examples of faces [24]. Images from the training set of each class (i.e., face and nonface class) are analyzed as observations of a random process and are characterized by two probability functions. They used a family of discrete Markov processes to model the face and background patterns and to estimate the probability model. The learning process is converted into an optimization problem to select the Markov process that maximizes the information-based discrimination between the two classes. The likelihood ratio is computed using the trained probability model and used to detect the faces.

Qian and Huang [119] presented a method that employs the strategies of both view-based and model-based methods. First, a visual attention algorithm, which uses high-level domain knowledge, is applied to reduce the search space. This is achieved by selecting image areas in which targets may appear based on the region maps generated by a region detection algorithm (water-shed method). Within the selected regions, faces are detected with a combination of template matching and feature matching methods using a hierarchical Markov random field and maximum a posteriori estimation.

Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.

2.4.9 Inductive Learning

Inductive learning algorithms have also been applied to locate and detect faces. Huang et al. applied Quinlan's C4.5 algorithm [121] to learn a decision tree from positive and negative examples of face patterns [64]. Each training example is an 8×8 pixel window and is represented by a vector of 30 attributes which is composed of entropy, mean, and standard deviation of the pixel intensity values. From these examples, C4.5 builds a classifier as a decision tree whose leaves indicate class identity and whose nodes specify tests to perform on a single attribute. The learned decision tree is then used to decide whether a face exists in the input example. The experiments show a localization accuracy rate of 96 percent on a set of 2,340 frontal face images in the FERET data set.

Duta and Jain [38] presented a method to learn the face concept using Mitchell's Find-S algorithm [101]. Similar to [154], they conjecture that the distribution of face patterns $p(x|face)$ can be approximated by a set of Gaussian clusters and that the distance from a face instance to one of the cluster centroids should be smaller than a fraction of the maximum distance from the points in that cluster to its centroid. The Find-S algorithm is then applied to learn the thresholding distance such that faces and nonfaces can be differentiated. This method has several distinct characteristics. First, it does not use negative (nonface) examples, while [154], [128] use both positive and negative examples. Second, only the central portion of a face is used for training. Third, feature vectors consist of images with 32 intensity levels or textures, while [154] uses full-scale intensity values as inputs. This method achieves a detection rate of 90 percent on the first CMU data set.

2.5 Discussion

We have reviewed and classified face detection methods into four major categories. However, some methods can be classified into more than one category. For example, template matching methods usually use a face model and subtemplates to extract facial features [132], [27], [180], [143], [51], and then use these features to locate or detect faces. Furthermore, the boundary between knowledge-based methods and some template matching methods is blurry since the latter usually implicitly applies human knowledge to define the face templates [132], [28], [143]. On the other hand, face detection methods can also be categorized otherwise. For example, these methods can be classified based on whether they rely on local features [87], [140], [124] or treat a face pattern as whole (i.e., holistic) [154], [128]. Nevertheless, we think the four major classes categorize most methods sufficiently and appropriately.

3 FACE IMAGE DATABASES AND PERFORMANCE EVALUATION

Most face detection methods require a training data set of face images and the databases originally developed for face recognition experiments can be used as training sets for face detection. Since these databases were constructed to empirically evaluate recognition algorithms in certain domains, we first review the characteristics of these databases and their applicability to face detection. Although numerous face

detection algorithms have been developed, most of them have not been tested on data sets with a large number of images. Furthermore, most experimental results are reported using different test sets. In order to compare methods fairly, a few benchmark data sets have recently been compiled. We review these benchmark data sets and discuss their characteristics. There are still a few issues that need to be carefully considered in performance evaluation even when the methods use the same test set. One issue is that researchers have different interpretations of what a "successful detection" is. Another issue is that different training sets are used, particularly, for appearance-based methods. We conclude this section with a discussion of these issues.

3.1 Face Image Database

Although many face detection methods have been proposed, less attention has been paid to the development of an image database for face detection research. The FERET database consists of monochrome images taken in different frontal views and in left and right profiles [115]. Only the upper torso of an individual (mostly head and necks) appears in an image on a uniform and uncluttered background. The FERET database has been used to assess the strengths and weaknesses of different face recognition approaches [115]. Since each image consists of an individual on a uniform and uncluttered background, it is not suitable for face detection benchmarking. This is similar to many databases that were created for the development and testing of face recognition algorithms. Turk and Pentland created a face database of 16 people [163] (available at <ftp://whitechapel.media.mit.edu/pub/images/>). The images are taken in frontal view with slight variability in head orientation (tilted upright, right, and left) on a cluttered background. The face database from AT&T Cambridge Laboratories (formerly known as the Olivetti database) consists of 10 different images for forty distinct subjects. (available at <http://www.uk.research.att.com/facedatabase.html>) [136]. The images were taken at different times, varying the lighting, facial expressions, and facial details (glasses). The Harvard database consists of cropped, masked frontal face images taken from a wide variety of light sources [57]. It was used by Halinan for a study on face recognition under the effect of varying illumination conditions. With 16 individuals, the Yale face database (available at <http://cvc.yale.edu/>) contains 10 frontal images per person, each with different facial expressions, with and without glasses, and under different lighting conditions [7]. The M2VTS multimodal database from the European ACTS projects was developed for access control experiments using multimodal inputs [116]. It contains sequences of face images of 37 people. The five sequences for each subject were taken over one week. Each image sequence contains images from right profile (-90 degree) to left profile (90 degree) while the subject counts from "0" to "9" in their native languages. The UMIST database consists of 564 images of 20 people with varying pose. The images of each subject cover a range of poses from right profile to frontal views [56]. The Purdue AR database contains over 3,276 color images of 126 people (70 males and 56 females) in frontal view [96]. This database is designed for face recognition experiments under several mixing factors, such as facial expressions, illumination conditions, and occlusions. All the faces appear with different facial expression (neutral, smile, anger, and scream), illumination

Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.

TABLE 2
Face Image Database

Data Set	Location	Description
MIT Database [163]	ftp://whitechapel.media.mit.edu/pub/images/	Faces of 16 people, 27 of each person under various illumination conditions, scale and head orientation.
FERET Database [115]	http://www.nist.gov/humanid/feret	A large collection of male and female faces. Each image contains a single person with certain expression.
UMIST Database [56]	http://images.ee.umist.ac.uk/danny/database.html	564 images of 20 subjects. Each subject covers a range of poses from profile to frontal views.
University of Bern Database	ftp://iamftp.unibe.ch/pub/Images/FaceImages/	300 frontal face images of 30 people (10 images per person) and 150 profile face images (5 images per person).
Yale Database [7]	http://cvc.yale.edu	Face images with expressions, glasses under different illumination conditions.
AT&T (Olivetti) Database [136]	http://www.uk.research.att.com	40 subjects, 10 images per subject.
Harvard Database [57]	ftp://ftp.hrl.harvard.edu/pub/faces/	Cropped, masked face images under a wide range of lighting conditions.
M2VTS Database [116]	http://poseidon.csd.auth.gr/M2VTS/index.html	A multimodal database containing various image sequences.
Purdue AR Database [96]	http://rvll.ecn.purdue.edu/~aleix/aleix_face_DB.html	3,276 face images with different facial expressions and occlusions under different illuminations.

(left light source, right light source, and sources from both sides), and occlusion (wearing sunglasses or scarf). The images were taken during two sessions separated by two weeks. All the images were taken by the same camera setup under tightly controlled conditions of illumination and pose. This face database has been applied to image and video indexing as well as retrieval [96]. Table 2 summarizes the characteristics of the abovementioned face image databases.

3.2 Benchmark Test Sets for Face Detection

The abovementioned databases are designed mainly to measure performance of face recognition methods and, thus, each image contains only one individual. Therefore, such databases can be best utilized as training sets rather than test sets. The tacit reason for comparing classifiers on test sets is that these data sets represent problems that systems might face in the real world and that superior performance on these benchmarks may translate to superior performance on other

real-world tasks. Toward this end, researchers have compiled a wide collection of data sets from a wide variety of images. Sung and Poggio created two databases for face detection [152], [154]. The first set consists of 301 frontal and near-frontal mugshots of 71 different people. These images are high quality digitized images with a fair amount of lighting variation. The second set consists of 23 images with a total of 149 face patterns. Most of these images have complex background with faces taking up only a small amount of the total image area. The most widely-used face detection database has been created by Rowley et al. [127], [130] (available at <http://www.cs.cmu.edu/~har/faces.html>). It consists of 130 images with a total of 507 frontal faces. This data set includes 23 images of the second data set used by Sung and Poggio [154]. Most images contain more than one face on a cluttered background and, so, this is a good test set to assess algorithms which detect upright frontal faces. Fig. 12 shows some images in the data set collected by Sung and



Fig. 12. Sample images in Sung and Poggio's data set [154]. Some images are scanned from newspapers and, thus, have low resolution. Though most faces in the images are upright and frontal. Some faces in the images appear in different pose.

Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.



Fig. 13. Sample images in Rowley et al.'s data set [128]. Some images contain hand-drawn cartoon faces. Most images contain more than one face and the face size varies significantly.

Poggio [154], and Fig. 13 shows images from the data set collected by Rowley et al. [128].

Rowley et al. also compiled another database of images for detecting 2D faces with frontal pose and rotation in image plane [129]. It contains 50 images with a total of 223 faces, of

which 210 are at angles of more than 10 degrees. Fig. 14 shows some rotated images in this data set. To measure the performance of detection methods on faces with profile views, Schneiderman and Kanade gathered a set of 208 images where each image contains faces with facial



Fig. 14. Sample images of Rowley et al.'s data set [129] which contains images with in-plane rotated faces against complex background.

Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.

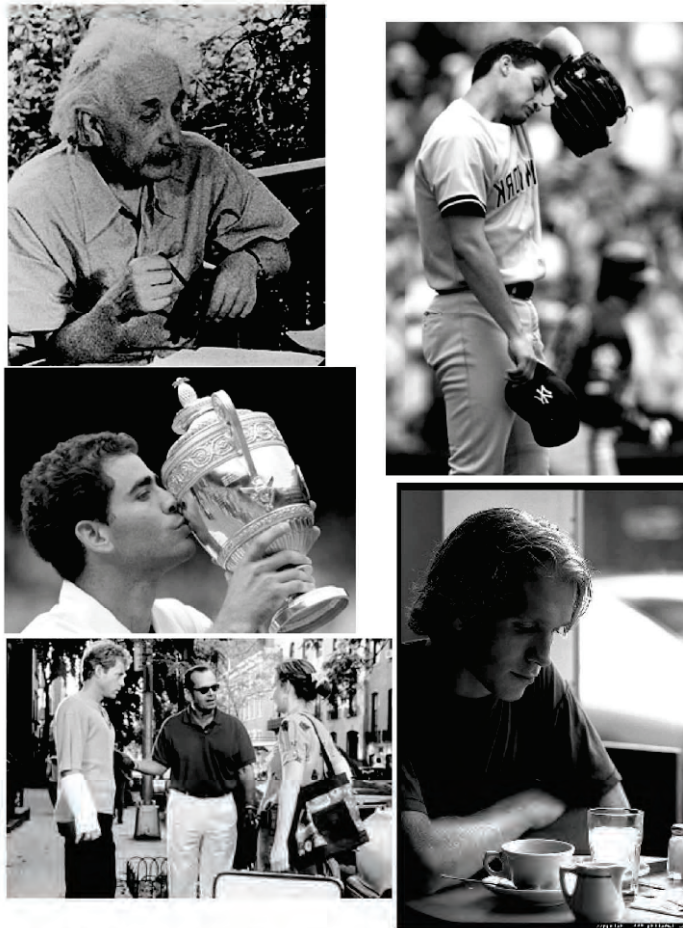


Fig. 15. Sample images of profile faces from Schneiderman and Kanade's data set [141]. This data set contains images with faces in profile views and some with facial expressions.

expressions and in profile views [141]. Fig. 15 shows some images in the test set.

Recently, Kodak compiled an image database as a common test bed for direct benchmarking of face detection and recognition algorithms [94]. Their database has 300 digital photos that are captured in a variety of resolutions and face size ranges from as small as 13×13 pixels to as

large as 300×300 pixels. Table 3 summarizes the characteristics of the abovementioned test sets for face detection.

3.3 Performance Evaluation

In order to obtain a fair empirical evaluation of face detection methods, it is important to use a standard and representative test set for experiments. Although many face detection methods have been developed over the past

TABLE 3
Test Sets for Face Detection

Data Set	Location	Description
MIT Test Set [154]	http://www.cs.cmu.edu/~har	Two sets of high and low resolution gray scale images with multiple faces in complex background.
CMU Test Set [128]	http://www.cs.cmu.edu/~har	130 gray scale images with a total of 507 frontal faces.
CMU Profile Face Test Set [141]	ftp://eyes.ius.cs.cmu.edu/usr20/ftp/testing_face_images.tar.gz	208 gray scale images with faces in profile views.
Kodak Data Set [94]	Eastman Kodak Corporation	Faces of multiple size, pose and under varying illumination in color images. Designed for face detection and recognition.

TABLE 4
Experimental Results on Images from Test Set 1 (125 Images with 483 Faces) and
Test Set 2 (23 Images with 136 Faces) (See Text for Details)

Method	Test Set 1		Test Set 2	
	Detection Rate	False Detections	Detection Rate	False Detections
Distribution based [154]	N/A	N/A	81.9%	13
Neural network [128]	92.5%	862	90.3%	42
Naive Bayes classifier [140]	93.0%	88	91.2%	12
Kullback relative information [24]	98.0%	12758	N/A	N/A
Support vector machine [107]	N/A	N/A	74.2%	20
Mixture of factor analyzers [175]	92.3%	82	89.4%	3
Fisher linear discriminant [175]	93.6%	74	91.5%	1
SNoW with primitive features [176]	94.2%	84	93.6%	3
SNoW with multi-scale features [176]	94.8%	78	94.1%	3
Inductive learning [38]	90%	N/A	N/A	N/A

decade, only a few of them have been tested on the same data set. Table 4 summarizes the reported performance among several appearance-based face detection methods on two standard data sets described in the previous section.

Although Table 4 shows the performance of these methods on the same test set, such an evaluation may not characterize how well these methods will compare in the field. There are a few factors that complicate the assessment of these appearance-based methods. First, the reported results are based on different training sets and different tuning parameters. The number and variety of training examples have a direct effect on the classification performance. However, this factor is often ignored in performance evaluation, which is an appropriate criteria if the goal is to evaluate the systems rather than the learning methods. The second factor is the training time and execution time. Although the training time is usually ignored by most systems, it may be important for real-time applications that require online training on different data sets. Third, the number of scanning windows in these methods vary because they are designed to operate in different environments (i.e., to detect faces within a size range). For example, Colmenarez and Huang argued that their method scans more windows than others and, thus, the number of false detections is higher than others [24]. Furthermore, the criteria adopted in reporting the detection rates is usually not clearly described in most systems. Fig. 16a shows a test image and Fig. 16b shows some subimages to be classified as a face or nonface. Suppose that all the subimages in Fig. 16b are classified as face patterns, some criteria may consider all of them as “successful” detections. However, a more strict criterion (e.g., each successful detection must contain all the visible eyes and mouths in an image) may classify most of them as false alarms. It is clear that a uniform criteria should be adopted to assess different classifiers. In [128], Rowley et al. adjust the criteria until the experimental results match their intuition of what a correct detection is, i.e., the square window should contain the eyes and also the mouth. The criteria they eventually use is that the center of the detected bounding box must be within four pixels and the scale must be within a factor of 1.2 (their scale step size) of ground truth (recorded manually).

Finally, the evaluation criteria may and should depend on the purpose of the detector. If the detector is going to be used to count people, then the sum of false positives and false

negatives is appropriate. On the other hand, if the detector is to be used to verify that an individual is who he/she claims to be (validation), then it may be acceptable for the face detector to have additional false detections since it is unlikely that these false detections will be acceptable images of the individual, i.e., the validation process will reject the false detections. In other words, the penalty or cost of one type of error should be properly weighted such that one can build an optimal classifier using Bayes decision rule (See Sections 2.2-2.4 in [36]). This argument is supported by a recent study which points out the accuracy of the classifier (i.e., detection rate in face detection) is not an appropriate goal for many of the real-world task [118]. One reason is that classification accuracy assumes equal misclassification costs. This assumption is problematic because for most real-world problems one type of classification error is much more expensive than another. In some face detection applications, it is important that all the existing faces are detected. Another reason is accuracy maximization assumes that the class distribution is known for the target environment. In other words, we assume the test data sets represent the “true” working environment for the face detectors. However, this assumption is rarely justified.

When detection methods are used within real systems, it is important to consider what computational resources are required, particularly, time and memory. Accuracy may need to be sacrificed for for speed.



Fig. 16. (a) Test image. (b) Detection results. Different criteria lead to different detection results. Suppose all the subimages in (b) are classified as face patterns by a classifier. A loose criterion may declare all the faces as “successful” detections, while a more strict one would declare most of them as nonfaces.

Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.

The scope of the considered techniques in evaluation is also important. In this survey, we discuss at least four different forms of the face detection problem:

1. Localization in which there is a single face and the goal is to provide a suitable estimate of position; scale to be used as input for face recognition.
2. In a cluttered monochrome scene, detect all faces.
3. In color images, detect (localize) all faces.
4. In a video sequence, detect and localize all faces.

An evaluation protocol should be carefully designed when assessing these different detection situations. It should be noted that there is a potential risk of using a universal though modest sized standard test set. As researchers develop new methods or "tweak" existing ones to get better performance on the test set, they engage in a subtle form of the unacceptable practice of "testing on the training set." As a consequence, the latest methods may perform better against this hypothetical test set but not actually perform better in practice. This can be obviated by having a sufficiently large and representative universal test set. Alternatively, methods could be evaluated on a smaller test set if that test set is randomly chosen (generated) each time the method is evaluated.

In summary, fair and effective performance evaluation requires careful design of protocols, scope, and data sets. Such issues have attracted much attention in numerous vision problems [21], [60], [142], [115]. However, performing this evaluation or trying to declare a "winner" is beyond the scope of this survey. Instead, we hope that either a consortium of researchers engaged in face detection or a third party will take on this task. Until then, we hope that when applicable, researchers will report the result of their methods on the publicly available data sets described here. As a first step toward this goal, we have collected sample face detection codes and evaluation tools at <http://vision.ai.uiuc.edu/mhyang/face-detection-survey.html>.

4 DISCUSSION AND CONCLUSION

This paper attempts to provide a comprehensive survey of research on face detection and to provide some structural categories for the methods described in over 150 papers. When appropriate, we have reported on the relative performance of methods. But, in so doing, we are cognizant that there is a lack of uniformity in how methods are evaluated and, so, it is imprudent to explicitly declare which methods indeed have the lowest error rates. Instead, we urge members of the community to develop and share test sets and to report results on already available test sets. We also feel the community needs to more seriously consider systematic performance evaluation: This would allow users of the face detection algorithms to know which ones are competitive in which domains. It will also spur researchers to produce truly more effective face detection algorithms.

Although significant progress has been made in the last two decades, there is still work to be done, and we believe that a robust face detection system should be effective under full variation in:

- lighting conditions,
- orientation, pose, and partial occlusion,
- facial expression, and

- presence of glasses, facial hair, and a variety of hair styles.

Face detection is a challenging and interesting problem in and of itself. However, it can also be seen as a one of the few attempts at solving one of the grand challenges of computer vision, the recognition of object classes. The class of faces admits a great deal of shape, color, and albedo variability due to differences in individuals, nonrigidity, facial hair, glasses, and makeup. Images are formed under variable lighting and 3D pose and may have cluttered backgrounds. Hence, face detection research confronts the full range of challenges found in general purpose, object class recognition. However, the class of faces also has very apparent regularities that are exploited by many heuristic or model-based methods or are readily "learned" in data-driven methods. One expects some regularities when defining classes in general, but they may not be so apparent. Finally, though faces have tremendous within-class variability, face detection remains a two class recognition problem (face versus nonface).

ACKNOWLEDGMENTS

The authors would like to thank Kevin Bowyer and the anonymous reviewers for their comments and suggestions. The authors also thank Baback Moghaddam, Henry Rowley, Brian Scassellati, Henry Schneiderman, Kah-Kay Sung, and Kin Choong Yow for providing images. M.-H. Yang was supported by ONR grant N00014-00-1-009 and Ray Ozzie Fellowship. D. J. Kriegman was supported in part by US National Science Foundation ITR CCR 00-86094 and the National Institute of Health R01-EY 12691-01. N. Ahuja was supported in part by US Office of Naval Research grant N00014-00-1-009.

REFERENCES

- [1] T. Agui, Y. Kokubo, H. Nagashashi, and T. Nagao, "Extraction of Face Recognition from Monochromatic Photographs Using Neural Networks," *Proc. Second Int'l Conf. Automation, Robotics, and Computer Vision*, vol. 1, pp. 18.8.1-18.8.5, 1992.
- [2] N. Ahuja, "A Transform for Multiscale Image Segmentation by Integrated Edge and Region Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 1211-1235, Sept. 1996.
- [3] Y. Amit, D. Geman, and B. Jedynak, "Efficient Focusing and Face Detection," *Face Recognition: From Theory to Applications*, H. Wechsler, P.J. Phillips, V. Bruce, F. Fogelman-Soulie, and T.S. Huang, eds., vol. 163, pp. 124-156, 1998.
- [4] Y. Amit, D. Geman, and K. Wilder, "Joint Induction of Shape Features and Tree Classifiers," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 11, pp. 1300-1305, Nov. 1997.
- [5] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley, 1984.
- [6] M.F. Augusteijn and T.L. Skujca, "Identification of Human Faces through Texture-Based Feature Recognition and Neural Network Technology," *Proc. IEEE Conf. Neural Networks*, pp. 392-398, 1993.
- [7] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.
- [8] O. Bernier, M. Collobert, R. Féraud, V. Lemarie, J.E. Viallet, and D. Collobert, "MULTRAK: A System for Automatic Multiperson Localization and Tracking in Real-Time," *Proc. IEEE Int'l Conf. Image Processing*, pp. 136-140, 1998.
- [9] C.M. Bishop, *Neural Networks for Pattern Recognition*. Oxford Univ. Press, 1995.

- [10] C. Breazeal and B. Scassellati, "A Context-Dependent Attention System for a Social Robot," *Proc. 16th Int'l Joint Conf. Artificial Intelligence*, vol. 2, pp. 1146-1151, 1999.
- [11] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Wadsworth, 1984.
- [12] G. Burel and D. Carel, "Detection and Localization of Faces on Digital Images," *Pattern Recognition Letters*, vol. 15, no. 10, pp. 963-967, 1994.
- [13] M.C. Burl, T.K. Leung, and P. Perona, "Face Localization via Shape Statistics," *Proc. First Int'l Workshop Automatic Face and Gesture Recognition*, pp. 154-159, 1995.
- [14] J. Cai, A. Goshtasby, and C. Yu, "Detecting Human Faces in Color Images," *Proc. 1998 Int'l Workshop Multi-Media Database Management Systems*, pp. 124-131, 1998.
- [15] J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-698, June 1986.
- [16] A. Carleson, C. Cumby, J. Rosen, and D. Roth, "The SNoW Learning Architecture," Technical Report UIUCDCS-R-99-2101, Univ. of Illinois at Urbana-Champaign Computer Science Dept., 1999.
- [17] D. Chai and K.N. Ngan, "Locating Facial Region of a Head-and-Shoulders Color Image," *Proc. Third Int'l Conf. Automatic Face and Gesture Recognition*, pp. 124-129, 1998.
- [18] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705-740, 1995.
- [19] Q. Chen, H. Wu, and M. Yachida, "Face Detection by Fuzzy Matching," *Proc. Fifth IEEE Int'l Conf. Computer Vision*, pp. 591-596, 1995.
- [20] D. Chetverikov and A. Lerch, "Multiresolution Face Detection," *Theoretical Foundations of Computer Vision*, vol. 69, pp. 131-140, 1993.
- [21] K. Cho, P. Meer, and J. Cabrera, "Performance Assessment through Bootstrap," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 11, pp. 1185-1198, Nov. 1997.
- [22] R. Cipolla and A. Blake, "The Dynamic Analysis of Apparent Contours," *Proc. Third IEEE Int'l Conf. Computer Vision*, pp. 616-623, 1990.
- [23] M. Collobert, R. Féraud, G.L. Tourneur, O. Bernier, J.E. Viallet, Y. Mahieux, and D. Collobert, "LISTEN: A System for Locating and Tracking Individual Speakers," *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition*, pp. 283-288, 1996.
- [24] A.J. Colmenarez and T.S. Huang, "Face Detection with Information-Based Maximum Discrimination," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 782-787, 1997.
- [25] T.F. Cootes and C.J. Taylor, "Locating Faces Using Statistical Feature Detectors," *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition*, pp. 204-209, 1996.
- [26] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley Interscience, 1991.
- [27] I. Craw, H. Ellis, and J. Lishman, "Automatic Extraction of Face Features," *Pattern Recognition Letters*, vol. 5, pp. 183-187, 1987.
- [28] I. Craw, D. Tock, and A. Bennett, "Finding Face Features," *Proc. Second European Conf. Computer Vision*, pp. 92-96, 1992.
- [29] J.L. Crowley and J.M. Bedrune, "Integration and Control of Reactive Visual Processes," *Proc. Third European Conf. Computer Vision*, vol. 2, pp. 47-58, 1994.
- [30] J.L. Crowley and F. Berard, "Multi-Modal Tracking of Faces for Video Communications," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 640-645, 1997.
- [31] Y. Dai and Y. Nakano, "Extraction for Facial Images from Complex Background Using Color Information and SGLD Matrices," *Proc. First Int'l Workshop Automatic Face and Gesture Recognition*, pp. 238-242, 1995.
- [32] Y. Dai and Y. Nakano, "Face-Texture Model Based on SGLD and Its Application in Face Detection in a Color Scene," *Pattern Recognition*, vol. 29, no. 6, pp. 1007-1017, 1996.
- [33] T. Darrell, G. Gordon, M. Harville, and J. Woodfill, "Integrated Person Tracking Using Stereo, Color, and Pattern Detection," *Int'l J. Computer Vision*, vol. 37, no. 2, pp. 175-185, 2000.
- [34] A. Dempster, "A Generalization of Bayesian Theory," *J. Royal Statistical Soc.*, vol. 30, pp. 205-247, 1978.
- [35] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski, "Classifying Facial Actions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974-989, Oct. 2000.
- [36] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*. New York: John Wiley, 1973.
- [37] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*. New York: Wiley-Interscience, 2001.
- [38] N. Duta and A.K. Jain, "Learning the Human Face Concept from Black and White Pictures," *Proc. Int'l Conf. Pattern Recognition*, pp. 1365-1367, 1998.
- [39] G.J. Edwards, C.J. Taylor, and T. Cootes, "Learning to Identify and Track Faces in Image Sequences," *Proc. Sixth IEEE Int'l Conf. Computer Vision*, pp. 317-322, 1998.
- [40] I.A. Essa and A. Pentland, "Facial Expression Recognition Using a Dynamic Model and Motion Energy," *Proc. Fifth IEEE Int'l Conf. Computer Vision*, pp. 360-367, 1995.
- [41] S. Fahlman and C. Lebiere, "The Cascade-Correlation Learning Architecture," *Advances in Neural Information Processing Systems 2*, D.S. Touretsky, ed., pp. 524-532, 1990.
- [42] R. Féraud, "PCA, Neural Networks and Estimation for Face Detection," *Face Recognition: From Theory to Applications*, H. Wechsler, P.J. Phillips, V. Bruce, F. Fogelman-Soulie, and T.S. Huang, eds., vol. 163, pp. 424-432, 1998.
- [43] R. Féraud and O. Bernier, "Ensemble and Modular Approaches for Face Detection: A Comparison," *Advances in Neural Information Processing Systems 10*, M.I. Jordan, M.J. Kearns, and S.A. Solla, eds., pp. 472-478, MIT Press, 1998.
- [44] R. Féraud, O.J. Bernier, J.-E. Villet, and M. Collobert, "A Fast and Accurate Face Detector Based on Neural Networks," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 42-53, Jan. 2001.
- [45] D. Forsyth, "A Novel Approach to Color Constancy," *Int'l J. Computer Vision*, vol. 5, no. 1, pp. 5-36, 1990.
- [46] B.J. Frey, A. Colmenarez, and T.S. Huang, "Mixtures of Local Subspaces for Face Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 32-37, 1998.
- [47] F. Fukunaga and W. Koontz, "Applications of the Karhunen-Loève Expansion to Feature Selection and Ordering," *IEEE Trans. Computers*, vol. 19, no. 5, pp. 311-318, 1970.
- [48] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic, 1972.
- [49] Z. Ghahramani and G.E. Hinton, "The EM Algorithm for Mixtures of Factor Analyzers," Technical Report CRG-TR-96-1, Dept. Computer Science, Univ. of Toronto, 1996.
- [50] R.C. Gonzalez and P.A. Wintz, *Digital Image Processing*. Reading: Addison Wesley, 1987.
- [51] V. Govindaraju, "Locating Human Faces in Photographs," *Int'l J. Computer Vision*, vol. 19, no. 2, pp. 129-146, 1996.
- [52] V. Govindaraju, D.B. Sher, R.K. Srihari, and S.N. Srihari, "Locating Human Faces in Newspaper Photographs," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 549-554, 1989.
- [53] V. Govindaraju, S.N. Srihari, and D.B. Sher, "A Computational Model for Face Location," *Proc. Third IEEE Int'l Conf. Computer Vision*, pp. 718-721, 1990.
- [54] H.P. Graf, T. Chen, E. Petajan, and E. Cosatto, "Locating Faces and Facial Parts," *Proc. First Int'l Workshop Automatic Face and Gesture Recognition*, pp. 41-46, 1995.
- [55] H.P. Graf, E. Cosatto, D. Gibbon, M. Kocheisen, and E. Petajan, "Multimodal System for Locating Heads and Faces," *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition*, pp. 88-93, 1996.
- [56] D.B. Graham and N.M. Allinson, "Characterizing Virtual Eigen-signatures for General Purpose Face Recognition," *Face Recognition: From Theory to Applications*, H. Wechsler, P.J. Phillips, V. Bruce, F. Fogelman-Soulie, and T.S. Huang, eds., vol. 163, pp. 446-456, 1998.
- [57] P. Hallinan, "A Deformable Model for Face Recognition Under Arbitrary Lighting Conditions," PhD thesis, Harvard Univ., 1995.
- [58] C.-C. Han, H.-Y.M. Liao, K.-C. Yu, and L.-H. Chen, "Fast Face Detection via Morphology-Based Pre-Processing," *Proc. Ninth Int'l Conf. Image Analysis and Processing*, pp. 469-476, 1998.
- [59] R.M. Haralick, K. Shanmugam, and I. Dinstein, "Texture Features for Image Classification," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610-621, 1973.
- [60] M. Heath, S. Sarkar, T. Sanocki, and K. Bowyer, "A Robust Visual Method for Assessing the Relative Performance of Edge Detection Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 12, pp. 1338-1359, Dec. 1997.
- [61] G.E. Hinton, P. Dayan, and M. Revow, "Modeling the Manifolds of Images of Handwritten Digits," *IEEE Trans. Neural Networks*, vol. 8, no. 1, pp. 65-74, 1997.

Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.

- [62] H. Hotelling, "Analysis of a Complex of Statistical Variables into Principal Components," *J. Educational Psychology*, vol. 24, pp. 417-441, pp. 498-520, 1933.
- [63] K. Hotta, T. Kurita, and T. Mishima, "Scale Invariant Face Detection Method Using Higher-Order Local Autocorrelation Features Extracted from Log-Polar Image," *Proc. Third Int'l Conf. Automatic Face and Gesture Recognition*, pp. 70-75, 1998.
- [64] J. Huang, S. Gutta, and H. Wechsler, "Detection of Human Faces Using Decision Trees," *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition*, pp. 248-252, 1996.
- [65] D. Huttenlocher, G. Klanderman, and W. Rucklidge, "Comparing Images Using the Hausdorff Distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, pp. 850-863, 1993.
- [66] T.S. Jebara and A. Pentland, "Parameterized Structure from Motion for 3D Adaptive Feedback Tracking of Faces," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 144-150, 1997.
- [67] T.S. Jebara, K. Russell, and A. Pentland, "Mixtures of Eigenfeatures for Real-Time Structure from Texture," *Proc. Sixth IEEE Int'l Conf. Computer Vision*, pp. 128-135, 1998.
- [68] I.T. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 1986.
- [69] M.J. Jones and J.M. Rehg, "Statistical Color Models with Application to Skin Detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 274-280, 1999.
- [70] P. Juell and R. Marsh, "A Hierarchical Neural Network for Human Face Detection," *Pattern Recognition*, vol. 29, no. 5, pp. 781-787, 1996.
- [71] T. Kanade, "Picture Processing by Computer Complex and Recognition of Human Faces," PhD thesis, Kyoto Univ., 1973.
- [72] K. Karhunen, "Über Lineare Methoden in der Wahrscheinlichkeitsrechnung," *Annales Academiae Scientiarum Fennicae, Series AI: Mathematica-Physica*, vol. 37, pp. 3-79, 1946. (Translated by RAND Corp., Santa Monica, Calif., Report T-131, Aug. 1960).
- [73] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Proc. First IEEE Int'l Conf. Computer Vision*, pp. 259-269, 1987.
- [74] R. Kauth, A. Pentland, and G. Thomas, "Blob: An Unsupervised Clustering Approach to Spatial Preprocessing of MSS Imagery," *Proc. 11th Int'l Symp. Remote Sensing of the Environment*, pp. 1309-1317, 1977.
- [75] D.G. Kendall, "Shape Manifolds, Procrustean Metrics, and Complex Projective Shapes," *Bull. London Math. Soc.*, vol. 16, pp. 81-121, 1984.
- [76] C. Kervrann, F. Davoine, P. Perez, H. Li, R. Forchheimer, and C. Labit, "Generalized Likelihood Ratio-Based Face Detection and Extraction of Mouth Features," *Proc. First Int'l Conf. Audio- and Video-Based Biometric Person Authentication*, pp. 27-34, 1997.
- [77] S.-H. Kim, N.-K. Kim, S.C. Ahn, and H.-G. Kim, "Object Oriented Face Detection Using Range and Color Information," *Proc. Third Int'l Conf. Automatic Face and Gesture Recognition*, pp. 76-81, 1998.
- [78] M. Kirby and L. Sirovich, "Application of the Karhunen-Loève Procedure for the Characterization of Human Faces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103-108, Jan. 1990.
- [79] R. Kjeldsen and J. Kender, "Finding Skin in Color Images," *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition*, pp. 312-317, 1996.
- [80] T. Kohonen, *Self-Organization and Associative Memory*. Springer 1989.
- [81] C. Kotropoulos and I. Pitas, "Rule-Based Face Detection in Frontal Views," *Proc. Int'l Conf. Acoustics, Speech and Signal Processing*, vol. 4, pp. 2537-2540, 1997.
- [82] C. Kotropoulos, A. Tefas, and I. Pitas, "Frontal Face Authentication Using Variants of Dynamic Link Matching Based on Mathematical Morphology," *Proc. IEEE Int'l Conf. Image Processing*, pp. 122-126, 1998.
- [83] M.A. Kramer, "Nonlinear Principal Component Analysis Using Autoassociative Neural Networks," *Am. Inst. Chemical Eng. J.*, vol. 37, no. 2, pp. 233-243, 1991.
- [84] Y.H. Kwon and N. da Vitoria Lobo, "Face Detection Using Templates," *Proc. Int'l Conf. Pattern Recognition*, pp. 764-767, 1994.
- [85] K. Lam and H. Yan, "Fast Algorithm for Locating Head Boundaries," *J. Electronic Imaging*, vol. 3, no. 4, pp. 351-359, 1994.
- [86] A. Lanitis, C.J. Taylor, and T.F. Cootes, "An Automatic Face Identification System Using Flexible Appearance Models," *Image and Vision Computing*, vol. 13, no. 5, pp. 393-401, 1995.
- [87] T.K. Leung, M.C. Burl, and P. Perona, "Finding Faces in Cluttered Scenes Using Random Labeled Graph Matching," *Proc. Fifth IEEE Int'l Conf. Computer Vision*, pp. 637-644, 1995.
- [88] T.K. Leung, M.C. Burl, and P. Perona, "Probabilistic Affine Invariants for Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 678-684, 1998.
- [89] M.S. Lew, "Information Theoretic View-Based and Modular Face Detection," *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition*, pp. 198-203, 1996.
- [90] F. Leymarie and M.D. Levine, "Tracking Deformable Objects in the Plan Using an Active Contour Model," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 6, pp. 617-634, June 1993.
- [91] S.-H. Lin, S.-Y. Kung, and L.-J. Lin, "Face Recognition/Detection by Probabilistic Decision-Based Neural Network," *IEEE Trans. Neural Networks*, vol. 8, no. 1, pp. 114-132, 1997.
- [92] N. Littlestone, "Learning Quickly when Irrelevant Attributes Abound: A New Linear-Threshold Algorithm," *Machine Learning*, vol. 2, pp. 285-318, 1988.
- [93] M.M. Loève, *Probability Theory*. Princeton, N.J.: Van Nostrand, 1955.
- [94] A.C. Loui, C.N. Judice, and S. Liu, "An Image Database for Benchmarking of Automatic Face Detection and Recognition Algorithms," *Proc. IEEE Int'l Conf. Image Processing*, pp. 146-150, 1998.
- [95] K.V. Mardia and I.L. Dryden, "Shape Distributions for Landmark Data," *Advanced Applied Probability*, vol. 21, pp. 742-755, 1989.
- [96] A. Martinez and R. Benavente, "The AR Face Database," Technical Report CVC 24, Purdue Univ., 1998.
- [97] A. Martinez and A. Kak, "PCA versus LDA," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228-233, Feb. 2001.
- [98] S. McKenna, S. Gong, and Y. Raja, "Modelling Facial Colour and Identity with Gaussian Mixtures," *Pattern Recognition*, vol. 31, no. 12, pp. 1883-1892, 1998.
- [99] S. McKenna, Y. Raja, and S. Gong, "Tracking Colour Objects Using Adaptive Mixture Models," *Image and Vision Computing*, vol. 17, nos. 3/4, pp. 223-229, 1998.
- [100] J. Miao, B. Yin, K. Wang, L. Shen, and X. Chen, "A Hierarchical Multiscale and Multiangle System for Human Face Detection in a Complex Background Using Gravity-Center Template," *Pattern Recognition*, vol. 32, no. 7, pp. 1237-1248, 1999.
- [101] T. Mitchell, *Machine Learning*. McGraw Hill, 1997.
- [102] Y. Miyake, H. Saitoh, H. Yaguchi, and N. Tsukada, "Facial Pattern Detection and Color Correction from Television Picture for Newspaper Printing," *J. Imaging Technology*, vol. 16, no. 5, pp. 165-169, 1990.
- [103] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696-710, July 1997.
- [104] A.V. Nefian and M. H. H III, "Face Detection and Recognition Using Hidden Markov Models," *Proc. IEEE Int'l Conf. Image Processing*, vol. 1, pp. 141-145, 1998.
- [105] N. Oliver, A. Pentland, and F. Berard, "LAFER: Lips and Face Real Time Tracker," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 123-129, 1997.
- [106] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio, "Pedestrian Detection Using Wavelet Templates," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 193-199, 1997.
- [107] E. Osuna, R. Freund, and F. Girosi, "Training Support Vector Machines: An Application to Face Detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 130-136, 1997.
- [108] C. Papageorgiou, M. Oren, and T. Poggio, "A General Framework for Object Detection," *Proc. Sixth IEEE Int'l Conf. Computer Vision*, pp. 555-562, 1998.
- [109] C. Papageorgiou and T. Poggio, "A Trainable System for Object Recognition," *Int'l J. Computer Vision*, vol. 38, no. 1, pp. 15-33, 2000.
- [110] K. Pearson, "On Lines and Planes of Closest Fit to Systems of Points in Space," *Philosophical Magazine*, vol. 2, pp. 559-572, 1901.
- [111] A. Pentland, "Looking at People," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 107-119, Jan. 2000.
- [112] A. Pentland, "Perceptual Intelligence," *Comm. ACM*, vol. 43, no. 3, pp. 35-44, 2000.
- [113] A. Pentland and T. Choudhury, "Face Recognition for Smart Environments," *IEEE Computer*, pp. 50-55, 2000.
- [114] A. Pentland, B. Moghaddam, and T. Starner, "View-Based and Modular Eigenspaces for Face Recognition," *Proc. Fourth IEEE Int'l Conf. Computer Vision*, pp. 84-91, 1994.

Authorized licensed use limited to: Ingrid Hsieh-Yee. Downloaded on April 11, 2025 at 23:01:10 UTC from IEEE Xplore. Restrictions apply.

- [115] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1034, Oct. 2000.
- [116] S. Pigeon and L. Vandendrope, "The M2VTS Multimodal Face Database," *Proc. First Int'l Conf. Audio- and Video-Based Biometric Person Authentication*, 1997.
- [117] M. Propp and A. Samal, "Artificial Neural Network Architectures for Human Face Detection," *Intelligent Eng. Systems through Artificial Neural Networks*, vol. 2, 1992.
- [118] F. Provost and T. Fawcett, "Robust Classification for Imprecise Environments," *Machine Learning*, vol. 42, no. 3, pp. 203-231, 2001.
- [119] R.J. Qian and T.S. Huang, "Object Detection Using Hierarchical MRF and MAP Estimation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 186-192, 1997.
- [120] R.J. Qian, M.I. Sezan, and K.E. Matthews, "A Robust Real-Time Face Tracking Algorithm," *Proc. IEEE Int'l Conf. Image Processing*, pp. 131-135, 1998.
- [121] J.R. Quinlan, C4. 5: *Programs for Machine Learning*. Kluwer Academic, 1993.
- [122] L.R. Rabiner and B.-H. Jung, *Fundamentals of Speech Recognition*. Prentice Hall, 1993.
- [123] A. Rajagopalan, K. Kumar, J. Karlekar, R. Manivasakan, M. Patil, U. Desai, P. Poonacha, and S. Chaudhuri, "Finding Faces in Photographs," *Proc. Sixth IEEE Int'l Conf. Computer Vision*, pp. 640-645, 1998.
- [124] T. Rikert, M. Jones, and P. Viola, "A Cluster-Based Statistical Model for Object Detection," *Proc. Seventh IEEE Int'l Conf. Computer Vision*, vol. 2, pp. 1046-1053, 1999.
- [125] D. Roth, "Learning to Resolve Natural Language Ambiguities: A Unified Approach," *Proc. 15th Nat'l Conf. Artificial Intelligence*, pp. 806-813, 1998.
- [126] H. Rowley, S. Baluja, and T. Kanade, "Human Face Detection in Visual Scenes," *Advances in Neural Information Processing Systems 8*, D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, eds., pp. 875-881, 1996.
- [127] H. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 203-208, 1996.
- [128] H. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23-38, Jan. 1998.
- [129] H. Rowley, S. Baluja, and T. Kanade, "Rotation Invariant Neural Network-Based Face Detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 38-44, 1998.
- [130] H.A. Rowley, "Neural Network-Based Face Detection," PhD thesis, Carnegie Mellon Univ., 1999.
- [131] E. Saber and A.M. Tekalp, "Frontal-View Face Detection and Facial Feature Extraction Using Color, Shape and Symmetry Based Cost Functions," *Pattern Recognition Letters*, vol. 17, no. 8, pp. 669-680, 1998.
- [132] T. Sakai, M. Nagao, and S. Fujibayashi, "Line Extraction and Pattern Detection in a Photograph," *Pattern Recognition*, vol. 1, pp. 233-248, 1969.
- [133] A. Samal and P.A. Iyengar, "Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey," *Pattern Recognition*, vol. 25, no. 1, pp. 65-77, 1992.
- [134] A. Samal and P.A. Iyengar, "Human Face Detection Using Silhouettes," *Int'l J. Pattern Recognition and Artificial Intelligence*, vol. 9, no. 6, pp. 845-867, 1995.
- [135] F. Samaria and S. Young, "HMM Based Architecture for Face Identification," *Image and Vision Computing*, vol. 12, pp. 537-583, 1994.
- [136] F.S. Samaria, "Face Recognition Using Hidden Markov Models," PhD thesis, Univ. of Cambridge, 1994.
- [137] S. Satoh, Y. Nakamura, and T. Kanade, "Name-It: Naming and Detecting Faces in News Videos," *IEEE Multimedia*, vol. 6, no. 1, pp. 22-35, 1999.
- [138] D. Saxe and R. Foulds, "Toward Robust Skin Identification in Video Images," *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition*, pp. 379-384, 1996.
- [139] B. Scassellati, "Eye Finding via Face Detection for a Foveated, Active Vision System," *Proc. 15th Nat'l Conf. Artificial Intelligence*, 1998.
- [140] H. Schneiderman and T. Kanade, "Probabilistic Modeling of Local Appearance and Spatial Relationships for Object Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 45-51, 1998.
- [141] H. Schneiderman and T. Kanade, "A Statistical Method for 3D Object Detection Applied to Faces and Cars," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 746-751, 2000.
- [142] J.A. Shufelt, "Performance Evaluation and Analysis of Monocular Building Extraction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 4, pp. 311-326, Apr. 1997.
- [143] P. Sinha, "Object Recognition via Image Invariants: A Case Study," *Investigative Ophthalmology and Visual Science*, vol. 35, no. 4, pp. 1735-1740, 1994.
- [144] P. Sinha, "Processing and Recognizing 3D Forms," PhD thesis, Massachusetts Inst. of Technology, 1995.
- [145] S.A. Sirohey, "Human Face Segmentation and Identification," Technical Report CS-TR-3176, Univ. of Maryland, 1993.
- [146] J. Sobottka and I. Pitas, "Segmentation and Tracking of Faces in Color Images," *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition*, pp. 236-241, 1996.
- [147] K. Sobottka and I. Pitas, "Face Localization and Feature Extraction Based on Shape and Color Information," *Proc. IEEE Int'l Conf. Image Processing*, pp. 483-486, 1996.
- [148] F. Soulie, E. Viennet, and B. Lamy, "Multi-Modular Neural Network Architectures: Pattern Recognition Applications in Optical Character Recognition and Human Face Recognition," *Int'l J. Pattern Recognition and Artificial Intelligence*, vol. 7, no. 4, pp. 721-755, 1993.
- [149] T. Starner and A. Pentland, "Real-Time ASL Recognition from Video Using HMM's," Technical Report 375, Media Lab, Massachusetts Inst. of Technology, 1996.
- [150] Y. Sumi and Y. Ohta, "Detection of Face Orientation and Facial Components Using Distributed Appearance Modeling," *Proc. First Int'l Workshop Automatic Face and Gesture Recognition*, pp. 254-259, 1995.
- [151] Q.B. Sun, W.M. Huang, and J.K. Wu, "Face Detection Based on Color and Local Symmetry Information," *Proc. Third Int'l Conf. Automatic Face and Gesture Recognition*, pp. 130-135, 1998.
- [152] K.-K. Sung, "Learning and Example Selection for Object and Pattern Detection," PhD thesis, Massachusetts Inst. of Technology, 1996.
- [153] K.-K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection," Technical Report AI Memo 1521, Massachusetts Inst. of Technology AI Lab, 1994.
- [154] K.-K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 39-51, Jan. 1998.
- [155] M.J. Swain and D.H. Ballard, "Color Indexing," *Int'l J. Computer Vision*, vol. 7, no. 1, pp. 11-32, 1991.
- [156] D.L. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 891-896, Aug. 1996.
- [157] B. Takacs and H. Wechsler, "Face Location Using a Dynamic Model of Retinal Feature Extraction," *Proc. First Int'l Workshop Automatic Face and Gesture Recognition*, pp. 243-247, 1995.
- [158] A. Tefas, C. Kotropoulos, and I. Pitas, "Variants of Dynamic Link Architecture Based on Mathematical Morphology for Frontal Face Authentication," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 814-819, 1998.
- [159] J.C. Terrillon, M. David, and S. Akamatsu, "Automatic Detection of Human Faces in Natural Scene Images by Use of a Skin Color Model and Invariant Moments," *Proc. Third Int'l Conf. Automatic Face and Gesture Recognition*, pp. 112-117, 1998.
- [160] J.C. Terrillon, M. David, and S. Akamatsu, "Detection of Human Faces in Complex Scene Images by Use of a Skin Color Model and Invariant Fourier-Mellin Moments," *Proc. Int'l Conf. Pattern Recognition*, pp. 1350-1355, 1998.
- [161] A. Tsukamoto, C.-W. Lee, and S. Tsuji, "Detection and Tracking of Human Face with Synthesized Templates," *Proc. First Asian Conf. Computer Vision*, pp. 183-186, 1993.
- [162] A. Tsukamoto, C.-W. Lee, and S. Tsuji, "Detection and Pose Estimation of Human Face with Synthesized Image Models," *Proc. Int'l Conf. Pattern Recognition*, pp. 754-757, 1994.
- [163] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.
- [164] R. Vaillant, C. Monroq, and Y. Le Cun, "An Original Approach for the Localisation of Objects in Images," *IEE Proc. Vision, Image and Signal Processing*, vol. 141, pp. 245-250, 1994.
- [165] M. Venkatraman and V. Govindaraju, "Zero Crossings of a Non-Orthogonal Wavelet Transform for Object Location," *Proc. IEEE Int'l Conf. Image Processing*, vol. 3, pp. 57-60, 1995.

- [166] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang, "Phoneme Recognition Using Time-Delay Neural Networks," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 328-339, May 1989.
- [167] H. Wang and S.-F. Chang, "A Highly Efficient System for Automatic Face Region Detection in MPEG Video," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, no. 4, pp. 615-628, 1997.
- [168] H. Wu, Q. Chen, and M. Yachida, "Face Detection from Color Images Using a Fuzzy Pattern Matching Method," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 6, pp. 557-563, June 1999.
- [169] H. Wu, T. Yokoyama, D. Pramadihanto, and M. Yachida, "Face and Facial Feature Extraction from Color Image," *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition*, pp. 345-350, 1996.
- [170] G. Yang and T. S. Huang, "Human Face Detection in Complex Background," *Pattern Recognition*, vol. 27, no. 1, pp. 53-63, 1994.
- [171] J. Yang, R. Stiefelhofen, U. Meier, and A. Waibel, "Visual Tracking for Multimodal Human Computer Interaction," *Proc. ACM Human Factors in Computing Systems Conf. (CHI 98)*, pp. 140-147, 1998.
- [172] J. Yang and A. Waibel, "A Real-Time Face Tracker," *Proc. Third Workshop Applications of Computer Vision*, pp. 142-147, 1996.
- [173] M.-H. Yang and N. Ahuja, "Detecting Human Faces in Color Images," *Proc. IEEE Int'l Conf. Image Processing*, vol. 1, pp. 127-130, 1998.
- [174] M.-H. Yang and N. Ahuja, "Gaussian Mixture Model for Human Skin Color and Its Application in Image and Video Databases," *Proc. SPIE: Storage and Retrieval for Image and Video Databases VII*, vol. 3656, pp. 458-466, 1999.
- [175] M.-H. Yang, N. Ahuja, and D. Kriegman, "Mixtures of Linear Subspaces for Face Detection," *Proc. Fourth Int'l Conf. Automatic Face and Gesture Recognition*, pp. 70-76, 2000.
- [176] M.-H. Yang, D. Roth, and N. Ahuja, "A SNoW-Based Face Detector," *Advances in Neural Information Processing Systems 12*, S.A. Solla, T. K. Leen, and K.-R. Müller, eds., pp. 855-861, MIT Press, 2000.
- [177] K.C. Yow and R. Cipolla, "A Probabilistic Framework for Perceptual Grouping of Features for Human Face Detection," *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition*, pp. 16-21, 1996.
- [178] K.C. Yow and R. Cipolla, "Feature-Based Human Face Detection," *Image and Vision Computing*, vol. 15, no. 9, pp. 713-735, 1997.
- [179] K.C. Yow and R. Cipolla, "Enhancing Human Face Detection Using Motion and Active Contours," *Proc. Third Asian Conf. Computer Vision*, pp. 515-522, 1998.
- [180] A. Yuille, P. Hallinan, and D. Cohen, "Feature Extraction from Faces Using Deformable Templates," *Int'l J. Computer Vision*, vol. 8, no. 2, pp. 99-111, 1992.
- [181] W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant Analysis of Principal Components for Face Recognition," *Proc. Third Int'l Conf. Automatic Face and Gesture Recognition*, pp. 336-341, 1998.



Ming-Hsuan Yang received the PhD degree in computer science from the University of Illinois at Urbana-Champaign in 2000. He studied computer science and power mechanical engineering at the National Tsing-Hua University, Taiwan; computer science and brain theory at the University of Southern California; artificial intelligence and electrical engineering at the University of Texas at Austin. In 1999, he received the Ray Ozzie fellowship for his

research work on vision-based human computer interaction. His research interests include computer vision, computer graphics, pattern recognition, cognitive science, neural computation, and machine learning. He is a member of the IEEE and IEEE Computer Society.



David J. Kriegman received the BSE degree in electrical engineering and computer science from Princeton University in 1983 and was awarded the Charles Ira Young Medal for Electrical Engineering Research. He received the MS degree in 1984 and the PhD degree in 1989 in electrical engineering from Stanford University. From 1990-1998, he was on the faculty of the Electrical Engineering and Computer Science Departments at Yale University. In 1998, he joined the Computer Science Department and Beckman Institute at the University of Illinois at Urbana-Champaign where he is an associate professor. Dr. Kriegman was chosen for a US National Science Foundation Young Investigator Award in 1992 and has received best paper awards at the 1996 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) and the 1998 European Conference on Computer Vision. He has served as program cochair of CVPR 2000, he is the associate editor-in-chief of the *IEEE Transactions Pattern Analysis and Machine Intelligence*, and is currently an associate editor of the *IEEE Transactions on Robotics and Automation*. He has published more than 90 papers on object recognition, illumination modeling, face recognition, structure from motion, geometry of curves and surfaces, mobile robot navigation, and robot planning. He is a senior member of the IEEE and the IEEE Computer Society.



Narendra Ahuja (F '92) received the BE degree with honors in electronics engineering from the Birla Institute of Technology and Science, Pilani, India, in 1972, the ME degree with distinction in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1974, and the PhD degree in computer science from the University of Maryland, College Park, in 1979. From 1974 to 1975, he was the scientific officer in the Department of Electronics, Government of India, New Delhi. From 1975 to 1979, he was at the Computer Vision Laboratory, University of Maryland, College Park. Since 1979, he has been with the University of Illinois at Urbana-Champaign where he is currently the Donald Biggar Willet Professor in the Department of Electrical and Computer Engineering, the Coordinated Science Laboratory, and the Beckman Institute. His interests are in computer vision, robotics, image processing, image synthesis, sensors, and parallel algorithms. His current research emphasizes the integrated use of multiple image sources of scene information to construct three-dimensional descriptions of scenes; the use of integrated image analysis for realistic image synthesis; parallel architectures and algorithms and special sensors for computer vision; extraction and representation of spatial structure, e.g., in images and video; and use of the results of image analysis for a variety of applications including visual communication, image manipulation, video retrieval, robotics, and scene navigation. He received the 1999 Emanuel R. Piore award of the IEEE and the 1998 Technology Achievement Award of the International Society for Optical Engineering. He was selected as an associate (1998-99) and Beckman Associate (1990-91) in the University of Illinois Center for Advanced Study. He received the University Scholar Award (1985), Presidential Young Investigator Award (1984), National Scholarship (1967-72), and President's Merit Award (1966). He has coauthored the books *Pattern Models* (Wiley, 1983) and *Motion and Structure from Image Sequences* (Springer-Verlag, 1992), and coedited the book *Advances in Image Understanding*, (IEEE Press, 1996). He is a fellow of the IEEE and a member of the IEEE Computer Society, American Association for Artificial Intelligence, International Association for Pattern Recognition, Association for Computing Machinery, American Association for the Advancement of Science, and International Society for Optical Engineering. He is a member of the Optical Society of America. He is on the editorial boards of the *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *Computer Vision, Graphics, and Image Processing*, the *Journal of Mathematical Imaging and Vision*; the *Journal of Pattern Analysis and Applications*; *International Journal of Imaging Systems and Technology*; and the *Journal of Information Science and Technology*, and a guest coeditor of the Artificial Intelligence Journal's special issue on vision.

► For more information on this or any other computing topic, please visit our Digital Library at <http://computer.org/publications/dlib>.

EXHIBIT 1007B

LIBRARY

CATALOG



 1 of 1
 

PERIODICAL OR NEWSPAPER

IEEE transactions on pattern analysis and machine intelligence

Full Record

[MARC Tags](#)

Main title

IEEE transactions on pattern analysis and machine intelligence.

Published/Created

[New York] IEEE Computer Society

<2004-> : Los Alamitos, CA : IEEE Computer Society

[Request this Item](#)

 [LC Find It](#)

More Information



LCCN Permalink

<https://lccn.loc.gov/80640328>

Publication history

v. PAMI-1- Jan. 1979-

Description

volumes 28 cm

Vol. designation PAMI dropped with <v. 10, no. 6, Nov. 1988>.

Current frequency

Monthly, 1989-

Former frequency

Quarterly, 1979

Bimonthly, 1980-1988

ISSN	0162-8828
Linking ISSN	0162-8828
LC classification	Q327 .I19
Variant title	Institute of Electrical and Electronics Engineers transactions on pattern analysis and machine intelligence TPAMI Transactions on pattern analysis and machine intelligence
Serial key title	IEEE transactions on pattern analysis and machine intelligence
Abbreviated title	IEEE trans. pattern anal. mach. intell. IEEE Trans Pattern Anal Mach Intell
Related names	IEEE Computer Society . Institute of Electrical and Electronics Engineers .
LC Subjects	Pattern perception--Periodicals . Pattern recognition systems--Periodicals . Artificial intelligence--Periodicals .
Other Subjects	Pattern Recognition, Automated. Artificial Intelligence. Artificial intelligence. Pattern perception. Pattern recognition systems. Kunstmatige intelligentie.
Form/Genre	Periodicals . Periodicals . Periodical . Electronic journals
Browse by shelf order	Q327
Reproduction no./Source	IEEE Service Center, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08854
Additional formats	Online version: IEEE transactions on pattern analysis and machine intelligence (Online) 1939-3539 (DLC) 2007213712 (OCoLC)44581019 CD-ROM version: IEEE transactions on pattern analysis and machine intelligence 2160-9292 (DLC) 2011200092 (OCoLC)712654733

LCCN	80640328
Invalid LCCN	sn 79004711
CODEN	ITPIDJ
Dewey class no.	001.53/4
NLM class no.	W1 IE4465
National bib agency no.	9885960
Other system no.	(OCoLC)ocm04253074
Additional Links	http://www.computer.org/tpami/
Type of material	Periodical or Newspaper
Content type	text
Media type	unmediated
Carrier type	volume
Source of description	Latest issue consulted: Vol. 26, no. 1 (Jan. 2004).

Item Availability



CALL NUMBER	Q327 .I19 Set 1
Request in	Jefferson or Adams Building Reading Rooms
Status	Not Charged
Older receipts	v.1-v.11 (1979-1989), v.12:no.7-v.17:no.3 (1990:July-1995:Mar.), v.17:no.5-v.17:no.6 (1995:May-1995:June) v.17:no.8-v.18:no.9 (1995:Aug.-1996:Sept.), v.18:no.11-v.19:no.4 (1996:Nov.-1997:Apr.) v.19:no.8 (1997:Aug.), v.20-v.34:no.5 (1998-2012:May), v.34:no.7- v.36:no.1 (2012:July-2014:Jan.), v.36:no.4-v.36:no.5 (2014:Apr.-2014:May), v.37:no.1 (2015:Jan.), v.38:no.1 (2016:Jan.)

CALL NUMBER	Q327 .I19 Copy 1 Unbound issues
Request in	Shelved at Jefferson or Adams Building Reading Rooms v. 19 1997 Aug c.1 Shelved at Jefferson or Adams Building Reading Rooms
Status	Not Charged
Latest receipts	v. 34, no. 6 (2012 June)

LIBRARY

CATALOG

PERIODICAL OR NEWSPAPER

IEEE transactions on pattern analysis and machine intelligence

Full Record

MARC Tags

000	04189cas a2200865 4500
001	11235899
005	20240918091529.0
008	780926c19799999caumr p 0 a0eng c
010	__ a 80640328 z sn 79004711
016	7_ a 9885960 2 DNLM
022	0_ a 0162-8828 l 0162-8828 2 1
030	__ a ITPIDJ
032	__ a 492090 b USPS
035	__ a (OCoLC)ocm04253074
037	__ b IEEE Service Center, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08854
040	__ a NSD b eng c NSD d COO d NSD d DLC d OCL d NSD d OCL d NST d OCL d NST d AIP d NST d AIP d NST d AIP d NST d NSD d OCL d MYG d CUS d OCL d MYG d IEEEEX d OCLCO d LYU d OCLCO d NLN d NLGGC d LVB d NSD d DLC d OCLCO d DLC d TULIB d OCLCO d OCLCO d OCLCF d OCLCO d IUL d OCLCO d OCLCO d OCLCO d OCLCO d OCLCA d GILDS d OCLCO d OCLCA d AU@
042	__ a nsdp a pcc
050	00 a Q327 b .I19
060	00 a W1 b IE4465
082	04 a 001.53/4
210	0_ a IEEE trans. pattern anal. mach. intell.

210	10	a IEEE Trans Pattern Anal Mach Intell 2 dnlm
222	_0	a IEEE transactions on pattern analysis and machine intelligence
245	00	a IEEE transactions on pattern analysis and machine intelligence.
246	3_	a Institute of Electrical and Electronics Engineers transactions on pattern analysis and machine intelligence
246	3_	a TPAMI
246	3_	a Transactions on pattern analysis and machine intelligence
260	__	a [New York] b IEEE Computer Society
260	3_	3 <2004-> : a Los Alamitos, CA : b IEEE Computer Society
300	__	a volumes c 28 cm
310	__	a Monthly, b 1989-
321	__	a Quarterly, b 1979
321	__	a Bimonthly, b 1980-1988
336	__	a text b txt 2 rdacontent
337	__	a unmediated b n 2 rdamedia
338	__	a volume b nc 2 rdacarrier
362	0_	a v. PAMI-1- Jan. 1979-
515	__	a Vol. designation PAMI dropped with <v. 10, no. 6, Nov. 1988>.
588	__	a Latest issue consulted: Vol. 26, no. 1 (Jan. 2004).
590	__	a SERBIB/SERLOC merged record
650	_0	a Pattern perception v Periodicals.
650	_0	a Pattern recognition systems v Periodicals.
650	_0	a Artificial intelligence v Periodicals.
650	_6	a Reconnaissance des formes (Informatique) v Périodiques.
650	_6	a Perception des structures v Périodiques.
650	_6	a Intelligence artificielle v Périodiques.
650	_7	a Artificial intelligence. 2 fast 0 (OCoLC)fst00817247
650	_7	a Pattern perception. 2 fast 0 (OCoLC)fst01055254
650	_7	a Pattern recognition systems. 2 fast 0 (OCoLC)fst01055266
650	17	a Kunstmatige intelligentie. 2 gtt
650	12	a Pattern Recognition, Automated.
650	22	a Artificial Intelligence.
655	_7	a Periodicals. 2 fast 0 (OCoLC)fst01411641
655	_7	a Periodicals. 2 lcgft
655	_2	a Periodical.
655	_0	a Electronic journals
710	2_	a IEEE Computer Society.

710	2_	a Institute of Electrical and Electronics Engineers.
776	08	i Online version: t IEEE transactions on pattern analysis and machine intelligence (Online) x 1939-3539 w (DLC) 2007213712 w (OCoLC)44581019
776	08	i CD-ROM version: t IEEE transactions on pattern analysis and machine intelligence x 2160-9292 w (DLC) 2011200092 w (OCoLC)712654733
850	—	a CU-RivA a CU-SB a CaBRC a CaBVaS a CaMWU a CaNBFU a CaNSHD a CaOKR a CaOLU a CaOTDT a CaOTU a CaQMU a CoFS a CtW a CtY-KS a DLC a DeU a FU a GU a ICU a InU a KyU a MB a MChB a MMeT a MdU a MiDW a MnU a MoKU a MsU a NBC a NN a NRU a NcD a NjP a NjR a OC a OkU a PPT a PPI a ScU a TU a TxArU a TxU a ViW a WU-E
859	41	u http://www.computer.org/tpami/
853	20	8 1 a v. b no. u 12 v r i (year) j (month) w m
863	41	8 1.1 a <23> b <1> i <2001> j <01>
890	—	a IEEE transactions on pattern analysis and machine intelligence. (IEEE Computer Society) New York. i 80-640328
906	—	a 7 b cbc c serials d 2 e ncip f 19 g n-oclcserc
920	—	a Keep 1
923	—	d 2002-10-21 s Group Registration
984	—	a srvf
985	—	e eserial 200406
991	—	b c-GenColl h Q327 i .I19 w SERIALS
991	—	b c-GenColl h Q327 i .I19 p 00034418615 t Copy 1 v 8 * 1986 w CCF
991	—	b c-GenColl h Q327 i .I19 p 00001372464 t Copy 1 v 3 * 1981 w CCF
991	—	b c-GenColl h Q327 i .I19 p 0001364357A t Copy 1 v 14 * 8888 w CCF
992	—	b SER w SERLOC

Request this Item



Item Availability



CALL NUMBER

Q327 .I19

Set 1

Request in

Jefferson or Adams Building Reading Rooms

Status

Not Charged

Older receipts v.1-v.11 (1979-1989), v.12:no.7-v.17:no.3 (1990:July-1995:Mar.),
v.17:no.5-v.17:no.6 (1995:May-1995:June)
v.17:no.8-v.18:no.9 (1995:Aug.-1996:Sept.), v.18:no.11-v.19:no.4
(1996:Nov.-1997:Apr.)
v.19:no.8 (1997:Aug.), v.20-v.34:no.5 (1998-2012:May), v.34:no.7-
v.36:no.1 (2012:July-2014:Jan.), v.36:no.4-v.36:no.5
(2014:Apr.-2014:May), v.37:no.1 (2015:Jan.), v.38:no.1 (2016:Jan.)

CALL NUMBER [Q327 .I19](#)
Copy 1
Unbound issues

Request in Shelved at Jefferson or Adams Building Reading Rooms
v. 19 1997 Aug c.1 Shelved at Jefferson or Adams Building Reading
Rooms

Status Not Charged

Latest receipts v. 34, no. 6 (2012 June)

EXHIBIT 1007C

libraries.org

A directory of libraries throughout the world

- [Home »](#)
- [Libraries »](#)
- [Guides](#)
- [Documents](#)
- [Vendors »](#)
- [Products »](#)
- [News](#)
- [Procurement](#)
- [Member Login »](#)

found 3 items where the item contains **Library of Congress**. Showing item 2 of 3.

<< prev list next >> search:

Library of Congress

United States -- Library of Congress

Address: 101 Independence Avenue SE
Washington, District of Columbia
20540
United States

Phone: 202-707-5000

Connect to: [Library Web Site](#) ✓ [Online Catalog](#)

Library details: Library of Congress is a **National** library.
This library is affiliated with [Library of Congress \(view map\)](#) . The collection of the library contains **175,000,000** volumes. The library circulates **250,000** items per year.

Other Info: Went live with Voyager on August 15, 1999. Currently planning to implement the Library Collections Access Platform (LCAP) on FOLIO and EBSCO discovery services in FY25.

Permalink:  <https://librarytechnology.org/library/2263>
(Use this link to refer back to this listing.)

Administration: The director of the library is **Dr. Carla Hayden** .

Organizational structure: The library is part of a **publicly-funded** organization. It operates on a **Non-profit** financial model.



Affiliations: The library is a member of the [Association of Research Libraries](#)

Collection size	175,000,000	volumes
Annual Circulation	250,000	transactions

 **Wireless:** The library offers **wireless** access to the Internet. <http://www.loc.gov/rr/wirelessaccess.html>

Library Hours		
Day	Open	Close
Sunday	Closed	
Monday	8:30 am	8:30 pm
Tuesday	8:30 am	8:30 pm
Wednesday	8:30 am	8:30 pm
Thursday	8:30 am	8:30 pm
Friday	8:30 am	5:00 pm
Saturday	8:30 am	5:00 pm



Library of Congress



map location ✓

Related Libraries

- All members of the [Association of Research Libraries](#) members
- [Directory of ARL Member Libraries](#)
- Libraries located in [Washington](#) (District of Columbia)
- [all National libs in District of Columbia](#)

About libraries.org

[libraries.org](#) is a free directory of libraries throughout the world.

This directory is a component of [Library Technology Guides](#).

Technology Profile		
	Product Name	Year Contracted
Current Automation System	FOLIO -- EBSCO Information Services	2023
Previous Automation System	Voyager	1998
Discovery Service (w/index)	Ex Libris CDI	
Electronic Resource Management	Electronic Resource Management	
Item ID Type	Barcode	
This library is responsible for the procurement of the library automation system.		

Please [contact Marshall Breeding](#) to report corrections about information listed for this library.

[View Privacy and Security Report](#)

Automation Survey: We are conducting a study on the automation systems used in libraries.

This survey should be completed by a person familiar with your libraries use of

FOLIO -- EBSCO Information Services.

[Respond to the 2024 Library Automation Perceptions Survey](#)

Identifiers	
libraries.org ID	2263
OCLC Symbol	DLC
WorldCat Registry ID	45266

Record History: This listing was created on Aug 9, 1998 and was last modified on Feb 13, 2025.

Updates: Corrections or Updates? Registered members of Library Technology Guides can submit updates to library listings in libraries.org. [Registration](#) is free and easy. Already registered? [login](#). Or, you can report corrections just by sending a message to [Marshall Breeding](#).

[libraries.org](#)

[Update this Entry](#)

[Library Technology Guides](#)

Maintained by [Marshall Breeding](#)

[terms of service](#)

Copyright 1994-2024

EXHIBIT 1007D

IEEE TRANSACTIONS ON
PATTERN ANALYSIS AND
MACHINE INTELLIGENCE

January 2002 (Vol. 24, No. 1)

Members- Download the complete issue

 [ZIP of all PDF files](#)

PAPERS		Free	Requires Subscription	Purchase
pp. 1-18 ♦ Integrated Position Estimation Using Aerial Image Sequences <i>D.-G. Sim, R.-H. Park, R.-C. Kim, S.U. Lee, I.-C. Kim</i>		 ABSTRACT	 PDF	 BUY ARTICLE
pp. 19-33 ♦ The Radiometry of Multiple Images <i>Q.-Tuan Luong, Pascal Fua, Yvan G. Leclerc</i>		 ABSTRACT	 PDF	 BUY ARTICLE
pp. 34-58 ♦ Detecting Faces in Images: A Survey <i>Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja</i>		 ABSTRACT	 PDF	 BUY ARTICLE
pp. 59-74 ♦ Piecewise Linear Skeletonization Using Principal Curves <i>Balazs Kegl, Adam Krzyzak</i>		 ABSTRACT	 PDF	 BUY ARTICLE
pp. 75-89 ♦ Object Tracking with Bayesian Estimation of Dynamic Layer Representations <i>Hai Tao, Harpreet S. Sawhney, Rakesh Kumar</i>		 ABSTRACT	 PDF	 BUY ARTICLE
pp. 90-102 ♦ ICP Registration Using Invariant Features <i>Gregory C. Sharp, Sang W. Lee, David K. Wehe</i>		 ABSTRACT	 PDF	 BUY ARTICLE
pp. 103-113 ♦ Multihierarchical Graph Search <i>Juan-Antonio Fernandez-Madrigal, Javier Gonzalez</i>		 ABSTRACT	 PDF	 BUY ARTICLE

The Wayback Machine - https://web.archive.org/web/20020302204214/http://www.computer.org:80/tpami/tp2002/i1toc.htm



[HOME](#)[INDEX](#)[SEARCH](#)[HELP](#)[CONTACT](#)[CART](#)



IEEE TRANSACTIONS ON
PATTERN ANALYSIS AND
MACHINE INTELLIGENCE

January 2002 (Vol. 24, No. 1)

Members- Download the complete issue
 [ZIP of all PDF files](#)

PAPERS

pp. 1-18 • Integrated Position Estimation Using Aerial Image Sequences

D.-G. Sim, R.-H. Park, R.-C. Kim, S.U. Lee, I.-C. Kim

Free

Requires
Subscription

Purchase

 ABSTRACT

 PDF

 BUY ARTICLE

pp. 19-33 • The Radiometry of Multiple Images

Q.-Tuan Luong, Pascal Fua, Yvan G. Leclerc

 ABSTRACT

 PDF

 BUY ARTICLE

pp. 34-58 • Detecting Faces in Images: A Survey

Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja

 ABSTRACT

 PDF

 BUY ARTICLE

pp. 59-74 • Piecewise Linear Skeletonization Using Principal Curves

Balazs Kegl, Adam Krzyzak

 ABSTRACT

 PDF

 BUY ARTICLE

pp. 75-89 • Object Tracking with Bayesian Estimation of Dynamic Layer Representations

Hai Tao, Harpreet S. Sawhney, Rakesh Kumar

 ABSTRACT

 PDF

 BUY ARTICLE

pp. 90-102 • ICP Registration Using Invariant Features

Gregory C. Sharp, Sang W. Lee, David K. Wehe

 ABSTRACT

 PDF

 BUY ARTICLE

pp. 103-113 • Multihierarchical Graph Search

Juan-Antonio Fernandez-Madrigal, Javier Gonzalez

 ABSTRACT

 PDF

 BUY ARTICLE

pp. 114-124 • Reconstruction of Three-Dimensional Objects through Matching of Their Parts

Georgios Papaioannou, Evaggelia-Aggeliki Karabassi, Theoharis Theoharis

 ABSTRACT

 PDF

 BUY ARTICLE

SHORT PAPERS

pp. 125-131 • Optical Flow in Log-Mapped Image Plane-A New Approach

Mohammed Yeasin

Free

Requires
Subscription

Purchase

 ABSTRACT

 PDF

 BUY ARTICLE

pp. 132-139 • Extraction and Optimization of B-Spline PBD Templates for Recognition of Connected Handwritten Digit Strings

Zhongkang Lu, Zheru Chi, Wan-Chi Siu

 ABSTRACT

 PDF

 BUY ARTICLE

pp. 139-144 • A Note on Park and Chin's Algorithm

Ronaldo Fumio Hashimoto, Junior Barrera

 ABSTRACT

 PDF

 BUY ARTICLE

The full text of IEEE Transactions on Pattern Analysis and Machine Intelligence is available to members of the IEEE Computer Society who have an [online subscription](#) and a [web-account](#).

This site and all contents (unless otherwise noted) are Copyright ©2001 ,
Institute of Electrical and Electronics Engineers, Inc. All rights reserved

EXHIBIT 1007E

Try the [Copyright Public Records System \(CPRS\)](#) pilot with enhanced search features and filters.

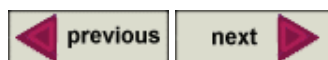
[Help](#)[Search](#)[History](#)[Titles](#)[Start Over](#)

Public Catalog

Copyright Catalog (1978 to present) at DC4

Search Request: Left Anchored Title = IEEE transactions on pattern analysis and machine intelligence

Search Results: Displaying 15 of 43 entries



Labeled View

IEEE Transactions on pattern analysis and machine intelligence : a...

Type of Work: Serial

Title: IEEE Transactions on pattern analysis and machine intelligence : a publication of the IEEE Computer Society / [associate editors, Prijono Hardjowirogo ... et al.].

Serial Publication Year: 2001

Serial Key Title: IEEE Transactions on pattern analysis and machine intelligence

Imprint: New York : IEEE, 1979-.

ISSN: 0162-8828

Description: print material.

Frequency: Quarterly; bimonthly per v. 2, no. 1; monthly per v. 13, 1991.

Publication History: Vol. PAMI-1, no 1-

Copyright Claimant: Institute of Electrical and Electronics Engineers, Inc. (employer for hire)

Issues Registered: v. 23, no. 1, Jan01.; Pub. 2001-01-02; Reg. 2001-06-15; TX0005329958
v. 23, no. 2, Feb01.; Pub. 2001-02-06; Reg. 2001-06-15; TX0005329958
v. 23, no. 3, Mar01.; Pub. 2001-03-01; Reg. 2001-06-15; TX0005328279
v. 23, no. 4, Apr01.; Pub. 2001-04-05; Reg. 2001-06-15; TX0005328279
v. 23, no. 5, May01.; Pub. 2001-05-07; Reg. 2001-09-07; TX0005376694
v. 23, no. 6, Jun01.; Pub. 2001-05-31; Reg. 2001-09-07; TX0005376694
v. 23, no. 7, Jul01.; Pub. 2001-07-02; Reg. 2001-09-07; TX0005376694
v. 23, no. 8, Aug01.; Pub. 2001-08-02; Reg. 2001-09-07; TX0005376694
v. 23, no. 9, Sep01.; Pub. 2001-09-04; Reg. 2002-01-10; TX0005520186
v. 23, no. 10, Oct01.; Pub. 2001-10-09; Reg. 2002-01-10; TX0005520186
v. 23, no. 11, Nov01.; Pub. 2001-11-05; Reg. 2002-01-10; TX0005525796
v. 23, no. 12, Dec01.; Pub. 2001-12-04; Reg. 2002-01-10; TX0005525796
v. 24, no. 1, Jan02. Created 2001; Pub. 2001-12-20; Reg. 2002-03-04; TX0005486059

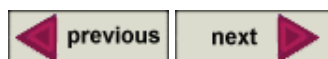
Other Title: Transactions on pattern analysis and machine intelligence

Names: [Hardjowirogo, Prijono](#)

[Institute of Electrical and Electronics Engineers, Inc. Computer Society](#)

[Institute of Electrical and Electronics Engineers, Inc.](#)

[IEEE Computer Society](#)



Save, Print and Email ([Help Page](#))

Select Download Format	Full Record ▾	Format for Print/Save
Enter your email address:		Email

[Help](#) [Search](#) [History](#) [Titles](#) [Start Over](#)

[Contact Us](#) | [Request Copies](#) | [Get a Search Estimate](#) | [Frequently Asked Questions \(FAQs\) about Copyright](#) | [Copyright Office](#)
[Home Page](#) | [Library of Congress Home Page](#)

EXHIBIT 1007F

EX1007F
Citations to Yang

Source Document:

Yang, M. H., Kriegman, D. J., & Ahuja, N. (2002, January). Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1), 34-58.

Citations: May to August 2002 (in chronological order)

Gorodnichy, D. O. (2002, May). On importance of nose for face tracking. In Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition (pp. 188-193). IEEE.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002, July). Visual features of intermediate complexity and their use in classification. *Nature neuroscience*, 5(7), 682-687.

Hadid, A., Pietikainen, M., & Martinkauppi, B. (2002, August). Color-based face detection using skin locus model and hierarchical filtering. In *2002 International Conference on Pattern Recognition* (Vol. 4, pp. 196-200). IEEE.

Mariano, V. Y., Min, J., Park, J. H., Kasturi, R., Mihalcik, D., Li, H., ... & Drayer, T. (2002, August). Performance evaluation of object detection algorithms. In *2002 International Conference on Pattern Recognition* (Vol. 3, pp. 965-969). IEEE.

Yuan, Q., Gao, W., & Yao, H. (2002, August). Robust frontal face detection in complex environment. In *2002 International Conference on Pattern Recognition* (Vol. 1, pp. 25-28). IEEE.

EXHIBIT 1008A

Face Identification by Fitting a 3D Morphable Model Using Linear Shape and Texture Error Functions

Sami Romdhani, Volker Blanz, and Thomas Vetter

University of Freiburg, Institut für Informatik,
Georges-Köhler-Allee 52, 79110 Freiburg, Germany,
{romdhani, volker, vetter}@informatik.uni-freiburg.de

Abstract. This paper presents a novel algorithm aiming at analysis and identification of faces viewed from different poses and illumination conditions. Face analysis from a single image is performed by recovering the shape and textures parameters of a 3D Morphable Model in an analysis-by-synthesis fashion. The shape parameters are computed from a shape error estimated by optical flow and the texture parameters are obtained from a texture error. The algorithm uses linear equations to recover the shape and texture parameters irrespective of pose and lighting conditions of the face image. Identification experiments are reported on more than 5000 images from the publicly available CMU-PIE database which includes faces viewed from 13 different poses and under 22 different illuminations. Extensive identification results are available on our web page for future comparison with novel algorithms.

1 Introduction

Images of faces do not depend on identity only, unfortunately. They are also influenced by illumination, by the pose from which the face is viewed and by its expression. Hence a face analysis system is confronted with the perplexing task of separating different sources of image variation. The recent FERET tests revealed that recognition performance drops significantly when pose and/or directed light are present in the input images [16]. Solving these two problems has become a major research issue. In this paper we present a novel algorithm able to identify faces in presence of combined pose and illumination variation.

A common approach for addressing these problems is to build an explicit generative model of the variations of face images. Then, analysis of a novel face image requires the fitting of an image generated by the model to the novel image, thereby coding the novel image in terms of model parameters. This fitting problem is in fact an optimisation problem: Its goal is to find the model parameters which minimise the difference between the image produced by the model and the novel image. The minimisation can be performed by a standard gradient descent algorithm [15] [6] or by a Levenberg-Marquardt [17] which is computationally expensive as it computes at each iteration the gradient of the image difference. This predicament lead first Gleicher [12], then Sclaroff and Isidoro [20] and Cootes *et al.* [8] to derive an optimisation-based algorithm in the specific context of model registration. This new approach, referred to as *Image Difference Decomposition* (IDD), makes the assumption that there is a linear dependency between the image difference (that is, the difference between the model image and the

input image) and the parameter update. Gleicher [12] applies this method to recover the projective transformation parameters of a small image region under motion. Sclaroff and Isidoro [20] endeavour to track the non-rigid motion of an image region. Cootes *et al.* [8] use IDD to fit a 2D face model to a novel face image.

Based on a theory of the complete image formation process (Section 2 and 3), we show that, in fact, the error function depends non-linearly on the parameters and the linear assumption made by IDD holds only on a small domain of variations (and not on the general problem which interests us, that of combined variation of pose and illumination). To approximate the non-linear image variations, induced by only the azimuth angle, Cootes *et al.* [9] had to learn five linear relations. If the number on non-linear degrees of freedom increases (elevation angle, direction of light, etc.), the number of linear relations required would grow exponentially. To overcome this problem, we use a 3D model that is able to describe faces viewed under arbitrary pose and illuminated from any direction. To fit the model, we do not assume a linear relation as in IDD, but we exploit, in Section 4, the fact that the error function is a non-linear mixing of two linear parts: A shape error function which linearly depends on the shape parameters of the model and a texture error which also depends linearly on the texture parameters. Hence, a *Linear Shape and Texture* (LiST) fitting algorithm is proposed, in Section 5, which uses these two linear functions to iteratively update the shape and texture parameters. Fitting and identification experiments are presented, in Section 7, on more than 5000 images from the publicly available CMU-PIE database which show faces viewed from 13 different poses and under 22 different lighting.

2 Morphable Model

In this section we briefly outline the Morphable Model. A detailed description can be found in [6] and [7]. The 3D Morphable Face Model is a model of faces represented in 3D where shape information is separated from texture information. The shape and texture models are learnt from a set of 200 exemplar faces which were acquired by a 3D *Cyberware*TM scanner. This scanner allows the sampling of texture with little interference from any external lighting. Hence texture values can be regarded as albedo values which are characteristic of identity only. Building a 3D Morphable Face Model requires to transform the shape and texture spaces into vector spaces for which any convex combination of exemplar shapes and textures describes a realistic human face. This is achieved by setting the exemplar faces in full correspondence with respect to a reference shape. Correspondences between all exemplar face and the reference face are established by an optical flow algorithm [2]. This produces a 3D deformation field for each exemplar faces which is used to warp the textures onto the reference shape yielding a shape-free texture. This scheme introduces a consistent labelling of vertices across the whole set of exemplar faces. The geometry of a face is represented by a shape vector $\mathbf{s} = (X_1, Y_1, Z_1, X_2, \dots, Y_N, Z_N) \in \mathcal{R}^{3N}$ that contains the X, Y, Z coordinates of its N vertices. For simplicity the R, G, B texture values are sampled at the same N points, therefore the texture vector of a face is represented by $\mathbf{t} = (R_1, G_1, B_1, R_2, \dots, G_N, B_N) \in \mathcal{R}^{3N}$. Realistic novel shape and texture vectors, $\mathbf{s}$ and $\mathbf{t}$ can be generated by convex combination of the M exemplars, $\mathbf{s}^{ex}$ and $\mathbf{t}^{ex}$:

$$\mathbf{s} = \sum_{i=1}^M a_i \mathbf{s}_i^{ex}, \quad \mathbf{t} = \sum_{i=1}^M b_i \mathbf{t}_i^{ex}, \quad \sum_{i=1}^M a_i = \sum_{i=1}^M b_i = 1 \quad (1)$$

To take advantage of the inherent regularity of human faces and reduce the dimensionality of the shape and texture spaces, a popular statistical technique, Principal Component Analysis, is used. PCA is applied separately on the shape and texture spaces, thereby ignoring the correlation between shape and texture as opposed to other techniques [8]. We describe the application of PCA to shapes, its application to textures is straightforward. After subtracting their mean, $\bar{\mathbf{s}}$, the exemplars are arranged in a data matrix $\mathbf{A}$ and the eigenvectors of its covariance matrix $\mathbf{C}$ are computed using the Singular Value Decomposition [18] of $\mathbf{A}$:

$$\begin{aligned} \bar{\mathbf{s}} &= \frac{1}{M} \sum_{i=1}^M \mathbf{s}_i^{ex}, \quad \mathbf{a}_i = \mathbf{s}_i^{ex} - \bar{\mathbf{s}}, \quad \mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_M) = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T, \\ \mathbf{C} &= \frac{1}{M} \mathbf{A}\mathbf{A}^T = \frac{1}{M} \mathbf{U}\mathbf{\Lambda}^2\mathbf{U}^T \end{aligned} \quad (2)$$

The M columns of the orthogonal matrix $\mathbf{U}$ are the eigenvectors of the covariance matrix $\mathbf{C}$, and $\sigma_i^2 = \frac{\lambda_i^2}{M}$ are its eigenvalues, where the λ_i are the elements of the diagonal matrix $\mathbf{\Lambda}$, arranged in decreasing order. Now, instead of describing a novel shape and texture as a convex combination of exemplars, as in Equation 1, we express them as a linear combination of shape and texture eigenvectors, $\mathbf{s}_i$ and $\mathbf{t}_i$, respectively:

$$\mathbf{s} = \bar{\mathbf{s}} + \sum_{i=1}^{M-1} \alpha_i \sigma_{s,i} \mathbf{s}_i, \quad \mathbf{t} = \bar{\mathbf{t}} + \sum_{i=1}^{M-1} \beta_i \sigma_{t,i} \mathbf{t}_i, \quad (3)$$

To achieve dimensionality reduction, the last eigenvalues (which are very small compared to the first eigenvalues) are set to zero. Hence, Equation 3, is reformulated as:

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{S}\boldsymbol{\alpha}, \quad \mathbf{t} = \bar{\mathbf{t}} + \mathbf{T}\boldsymbol{\beta} \quad (4)$$

where the columns of $\mathbf{S}$ and $\mathbf{T}$ are the most significant eigenvectors $\mathbf{s}_i$ and $\mathbf{t}_i$, re-scaled by their standard deviation. The Morphable Model shape and texture coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ constitute a pose and illumination invariant low-dimensional coding of the identity of a face as they are derived from characteristics pertaining only to the identity (3D shape and albedo). They will be used to identify faces in Section 7.2.

An additional and important benefit of PCA is that it provides an estimate of the probability densities of the shapes and textures, assuming that these densities are Gaussian:

$$p(\mathbf{s}) \sim e^{-\frac{1}{2}\|\boldsymbol{\alpha}\|^2}, \quad p(\mathbf{t}) \sim e^{-\frac{1}{2}\|\boldsymbol{\beta}\|^2}. \quad (5)$$

3 Face Rendering

The Morphable Face Model is a generative model which can produce photo-realistic face images. In this section, we summarise the graphical rendering process which transforms face shape and texture vectors, $\mathbf{s}$ and $\mathbf{t}$, into a face image I . For more information see [6], or [7]. In this section, we use a hat $\hat{\cdot}$ to denote entities pertaining to a single vertex.

Shape Transformation. Let us denote by $\hat{\mathbf{s}}$, the (X, Y, Z) coordinates of a vertex of the shape $\mathbf{s}$. The first rigid transformation is a translation by $\hat{\mathbf{t}}_{3d}$ and a rotation about the axes X, Y, Z of angles ϕ, γ and θ . The translation $\hat{\mathbf{t}}_{3d}$ is such as it sets the origin at the centre of the 3D shape $\mathbf{s}$. We denote by $\hat{\mathbf{w}}$ the result of this transformation:

$$\hat{\mathbf{w}} = \hat{R}_\phi \hat{R}_\gamma \hat{R}_\theta (\hat{\mathbf{v}} + \hat{\mathbf{t}}_{3d}), \quad \hat{\mathbf{t}}_{3d} = \sum_j^N \hat{\mathbf{s}}_j \quad (6)$$

The 2D image coordinates of the vertex, denoted by $\hat{\mathbf{s}}_{2d}$, are then determined using an orthographic projection (as opposed to [6]) which is legitimate when the distance from the camera to the rendered head is much larger than its dimension. In practise this introduces only a minor error. After the projection a 2D translation, $\hat{\mathbf{t}}_{2d}$ and a re-scaling of f are applied:

$$\hat{\mathbf{s}}_{2d} = f \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \hat{\mathbf{w}} + \hat{\mathbf{t}}_{2d} \quad (7)$$

Equations 4, 6 and 7 can be expressed as one equation for the entire set of vertices, clarifying the dependency between the 2D image coordinates vector $\mathbf{s}_{2d}$ and the shape parameters α :

$$\mathbf{s}_{2d} = f \mathbf{P} \mathbf{R} (\bar{\mathbf{s}} + \mathbf{S} \alpha + \mathbf{t}_{3d}) + \mathbf{t}_{2d} \quad (8)$$

where $\mathbf{t}_{3d}$ and $\mathbf{t}_{2d}$ are translation vectors of length $3N$ and $2N$ respectively formed by concatenating $\hat{\mathbf{t}}_{3d}$ and $\hat{\mathbf{t}}_{2d}$. $\mathbf{R}$ is a $3N \times 3N$ block diagonal matrix which performs the rotation $R_\phi R_\gamma R_\theta$ of all N vertices. $\mathbf{P}$ is the $2N \times 3N$ orthographic projection matrix for the ensemble of vertices. Note that, for rendering, a test of visibility must still be performed using a z-buffer method [11].

Illumination and Colour Transformation. We assume that the face is illuminated by ambient light and one directed light. In the same fashion as the shape transformation aforementioned, we denote by $\hat{\mathbf{t}}$ the R, G, B colour values of one vertex of the texture $\mathbf{t}$ and by $\hat{\mathbf{n}}$ its unit-length normal (computed by use of the shape $\hat{\mathbf{s}}$). The illumination effects are modelled using Phong illumination [11]:

$$\hat{\mathbf{t}}_l = (\hat{A}_{amb} + \langle \hat{\mathbf{n}}, \mathbf{l} \rangle \hat{A}_{dir}) \hat{\mathbf{t}} + k_s \langle \mathbf{v}, \hat{\mathbf{r}} \rangle^\nu \hat{A}_{dir} \mathbf{1}, \quad \hat{\mathbf{r}} = 2\langle \hat{\mathbf{n}}, \mathbf{l} \rangle - \mathbf{l}, \quad \|\mathbf{r}\| = 1 \quad (9)$$

where $\hat{\mathbf{t}}_l$ is the resulting colour of the vertex, $\hat{A}_{amb}$ is a diagonal matrix containing the R, G, B colour values of the ambient light, $\mathbf{l}$ is the direction of the light, $\hat{A}_{dir}$ the diagonal matrix holding its colour, k_s , the specular reflectance, ν , the angular distribution of the specularities of human skin (or *shininess*), $\mathbf{v}$, the viewing direction, $\mathbf{1}$, a vector of ones and $\mathbf{r}$ is the direction of maximum specular reflection.

The second colour transformation allows to render various types of images such as photograph from different sources, grey-scale images or paintings [6] [7]. It uses a diagonal matrix, $\mathbf{G}$, storing the R, G, B colour gains, a vector $\hat{\mathbf{o}}$ containing the R, G, B offsets and a matrix $M(c)$ which interpolates between the original colour values and the overall luminance depending on the colour contrast c . The resulting colour of the vertex rendered in the image at the position $\hat{\mathbf{s}}_{2d}$ is $\hat{\mathbf{t}}_M$:

$$\hat{\mathbf{t}}_M = G M(c) \hat{\mathbf{t}}_l + \hat{\mathbf{o}}, \quad M(c) = c I + (1 - c) \begin{pmatrix} 0.3 & 0.59 & 0.11 \\ 0.3 & 0.59 & 0.11 \\ 0.3 & 0.59 & 0.11 \end{pmatrix} \quad (10)$$

Again, the dependency of the rendered texture over the model parameters β is clarified by summarising Equations 4, 9 and 10 as one:

$$\mathbf{t}_M = (A_{amb} + A_{dir}) (\bar{\mathbf{t}} + T\beta) + \mathbf{a}_{spec} + \mathbf{o} \quad (11)$$

where A_{amb} and A_{dir} are a $3N \times 3N$ block diagonal matrix representing the effects of colour transformation, G and $M(c)$, and, respectively, the ambient light $\hat{A}_{amb}$ and the lambertian part of the directed lights; $\mathbf{a}_{spec}$ contains the specular part of the light and $\mathbf{o}$ the R, G, B offset.

Texture Warping. The image of a face is produced by setting the colour of the vertices, $\mathbf{t}_M$, at the positions $\mathbf{s}_{2d}$. Naturally, this is performed only for the visible vertices. As the values of $\mathbf{s}_{2d}$ are non-discrete, a rasterisation is performed. Also, to obtain a smooth and continuous image, a 2D to 2D mapping is achieved, whereby the colour of pixels between vertices is interpolated using the topology provided by a triangular mesh. These three steps constitute a warping operation, denoted by the function $\mathcal{W}(\cdot, \cdot)$:

$$I = \mathcal{W}(\mathbf{s}_{2d}, \mathbf{t}_M) \quad (12)$$

4 Inversion of the Face Image Formation Process

In Section 2 and 3 the problem of the *generation* of a photo-realistic face image is described. Algorithms of this kind found many applications in the field of Computer Graphics. Here, we are interested in the inverse problem, that of *recovering* the shape, texture, rigid transformation, colour transformation and lighting parameters from one image of a face, I_{input} . This vision problem is generally referred to as fitting. In this section, the theory motivating a novel fitting algorithm is detailed and in the next section the fitting algorithm is formally described.

Explaining an input image in terms of model parameters is usually performed using an *Analysis-by-Synthesis* loop whereby, at each iteration, an image is generated using the current estimate of the model parameters, then the difference between the model image and the input image is computed, and an update of the model parameters which reduces this difference is sought for. This is a minimisation problem whose cost function is the image difference between the image rendered by the model and the input image. The image difference, $\delta \mathbf{t}$ can be computed on the reference grid obtained from the shape estimate:

$$\mathbf{t}_I = \mathcal{W}^{-1}(\mathbf{s}_{2d}, I_{input}) \quad \delta \mathbf{t} = \mathbf{t}_M - \mathbf{t}_I \quad (13)$$

where $\mathbf{t}_I$ is a texture vector formed by the R, G, B values sampled from the input image at the pixels given by the current shape estimate $\mathbf{s}_{2d}$. Such an error function was

introduced by the Image Difference Decomposition algorithm [12] and used to update the parameters using a linear equation:

$$\delta \mathbf{p} = A \delta \mathbf{t} \quad (14)$$

where $\mathbf{p}$ regroups the model parameters which are to be recovered and A is a matrix learnt off-line. However this linear dependency is an assumption which holds in a small domain of variation of I_{input} . In the presence of pose and illumination variation, the image formation process, described in Section 3, allows the expression of Equation 13 as:

$$\delta \mathbf{t} = \frac{\partial \mathbf{t}_M}{\partial \alpha} \delta \alpha + \frac{\partial \mathbf{t}_M}{\partial \beta} \delta \beta + \frac{\partial \mathbf{t}_M}{\partial \mathbf{p}_i} \delta \mathbf{p}_i - \frac{\partial \mathcal{W}^{-1}(\mathbf{s}_{2d}, I_{input})}{\partial \mathbf{s}_{2d}} \left(\frac{\partial \mathbf{s}_{2d}}{\partial \alpha} \delta \alpha + \frac{\partial \mathbf{s}_{2d}}{\partial \mathbf{p}_r} \delta \mathbf{p}_r \right) \quad (15)$$

Several of these Jacobi matrices are not constant due to the non-linear influence of the shape and the illumination on the texture (see Equation 9 and 11). This makes Equation 13 non-linear and the assumption leading to Equation 14 violated. The non-constancy of the derivatives obliges their re-computation at each iteration of the fitting algorithm. However, we can still benefit from the regularity of the problem by realising that some of the derivatives are relatively inexpensive to compute: If α and $\mathbf{p}_i$ are maintained constant, $\frac{\partial \mathbf{t}_M}{\partial \beta}$ is constant and can be computed at relatively low cost and used to update β . Hence, $\delta \beta$ is obtained by solving a simple system of linear equations. Similarly, α , $\mathbf{p}_i$ and $\mathbf{p}_r$ are updated separately while maintaining the other parameters constant. This process is repeated iteratively until global convergence is obtained. The recovery of each of the parameters is detailed in the next two sections. It should be noted that each parameter update is biased because the other parameters are kept constant. To overcome this problem a regularisation technique, detailed in Section 4.3, is used.

4.1 Shape Fitting

Updating the shape parameters α from an image difference $\delta \mathbf{t}$ is difficult due to the combined influence of: (i) the shape error (ii) the normals which depends non-linearly on the shape and (iii) the texture error (see Equation 9 and 11). On the other hand, recovering the shape parameters α from a shape error is easier: if f , ϕ , γ and θ are kept constant the relation between the shape $\mathbf{s}_{2d}$ and α is linear (from Equation 8):

$$\frac{\partial \mathbf{s}_{2d}}{\partial \alpha} = f P R S \quad (16)$$

So, the derivative is the 2D projection of the shape PCA matrix rotated and scaled by the current estimate of the pose and scale parameters, ϕ , γ , θ , and f . Note that this gradient is not learnt off-line nor numerically estimated on-line, it simply results from a matrix multiplication. Hence, updating α from a shape error $\delta \mathbf{s}_{2d}$ requires only to solve a linear system of equations. Ideally the shape error would be directly obtained from the image (as for the texture error described in the next section). However, as the image does not contain any shape information, the shape error is estimated by applying an optical flow algorithm [2] between the input image and the model image rendered using the current parameters. We denote by $\mathbf{s}_{2d}^{of}$ the shape recovered by optical flow in the image frame.

Rotation, Translation and Scale Parameters Update. Equation 16 is used to update the shape model parameter α but is not appropriate to update the shape transformation parameters due to the non-linear interaction of the parameters. Therefore, to recover these parameters, a Levenberg-Marquardt optimisation [18] is used to minimise the geometrical error between s_{2d}^{of} and the model points [14]:

$$\arg \min_{f, \phi, \gamma, \theta, \mathbf{t}_{2d}} \| \mathbf{s}_{2d}^{of} - (f P R (\bar{\mathbf{s}} + S \alpha + \mathbf{t}_{3d}) + \mathbf{t}_{2d}) \|^2 = (\tilde{f}, \tilde{\phi}, \tilde{\gamma}, \tilde{\theta}, \tilde{\mathbf{t}}_{2d}) \quad (17)$$

The parameters α and $\mathbf{t}_{3d}$ are not optimised. The current value of α is used and $\mathbf{t}_{3d}$ is uniquely determined by α (as it is the centre of mass of $\mathbf{s}$). This optimisation is very fast as the search space is very low-dimensional (6).

Shape Model Parameters Update. A Shape error, δs_{2d} , is defined as the difference between the shape recovered by optical flow $\mathbf{s}_{2d}^{of}$ and the shape rendered using the new rotation and scaling parameters $\tilde{f}, \tilde{\phi}, \tilde{\gamma}, \tilde{\theta}, \tilde{\mathbf{t}}_{2d}$ and the current estimate of α . If the rotation and scaling parameters are kept constant, the dependency between δs_{2d} and α is linear. Hence, the update of α which explains the remaining of δs_{2d} (not accounted for by the rotation, translation and scaling) is obtained in a single step, by solving an over-constrained linear system of equations:

$$\delta s_{2d} = f P R S \delta \alpha \quad (18)$$

4.2 Texture Fitting

Illumination and Colour Transformation Update. The update on p_i is performed while maintaining the other parameters constant. Unfortunately $\delta \mathbf{t}$ does not depend linearly on the p_i and, similarly to the p_r update aforementioned, we have to resort to a non-linear Levenberg-Marquardt minimisation [18] to recover the illumination and colour transformation parameters. The problem to be solved is:

$$\min_{A_{amb}, A_{dir}, G, c, \mathbf{o}} \| \delta \mathbf{t} \|^2 \quad (19)$$

with $\delta \mathbf{t}$ defined by Equation 13. The minimisation can be performed on the ensemble of the visible vertices or on a subset thereof to decrease computation time. Again, this optimisation is fast due to the low-dimensionality of the search space (11).

Texture Model Parameters Update. If α , p_i and p_r are kept constant Equation 15 becomes:

$$\delta \mathbf{t} = (A_{amb} + A_{dir}) T \delta \beta \quad (20)$$

Similarly to the shape parameter update, the texture parameter update is obtained in a single step by solving an over-constrained linear system of equations performed by computing the pseudo-inverse of $(A_{amb} + A_{dir}) T$. Again, to speed-up the processing, only a subset of the pixels can be used.

4.3 Model Constraints

The fitting algorithm exploits the fact that the shape and image differences depend linearly on, respectively, the shape and texture model parameters, if the other parameters are maintained constant. This restriction introduces a bias in $\frac{\partial \mathbf{s}_{2d}}{\partial \boldsymbol{\alpha}}$ and in $\frac{\partial I}{\partial \boldsymbol{\beta}}$. Therefore updating $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ solely based on Equations 18 and 20 would lead to inaccuracies. These equations must be traded off with the prior probabilities of the parameters given by Equation 5. In the remaining of this section, we discuss only the shape parameters. The same is performed on the texture parameters. Equation 18 solves the following problem:

$$\min_{\delta \boldsymbol{\alpha}} \|\delta \mathbf{s}_{2d} - A \delta \boldsymbol{\alpha}\|^2 \quad (21)$$

where the matrix A is equal to $f P R S$. Now, instead of solving this problem, we favour solving the following problem:

$$\min_{\delta \boldsymbol{\alpha}} (\|\delta \mathbf{s}_{2d} - A \delta \boldsymbol{\alpha}\|^2 + \eta \|\boldsymbol{\alpha}_{cur} + \delta \boldsymbol{\alpha}\|^2) \quad (22)$$

where $\boldsymbol{\alpha}_{cur}$ is the shape model parameter vector before the update. This formula is motivated by the fact that the prior probability of $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{cur} + \delta \boldsymbol{\alpha}$ directly depends on its norm. The parameter η trades off the prior probability and the fitting quality. Applying singular value decomposition to A yields $A = U \text{diag}(w_i) V^T$, where w_i are the singular values of A . It can be shown that Equation 22 is minimal for:

$$\delta \boldsymbol{\alpha} = V \text{diag}\left(\frac{w_i}{w_i^2 + \eta}\right) U^T \delta \mathbf{s}_{2d} - V \text{diag}\left(\frac{\eta}{w_i^2 + \eta}\right) V^T \boldsymbol{\alpha}_{cur} \quad (23)$$

Note that (i) if $\eta = 0$, the update computed by Equation 23 is the same as the one obtained by Equation 18 and that (ii) computing the update using Equation 18 is not much more expensive as computing the update using Equation 23.

Value of η . The value of η depends on the magnitude of $\|\delta \mathbf{s}_{2d}\|^2$ relative to the magnitude of $\|\delta \boldsymbol{\alpha}\|^2$ (Equation 22). As $\delta \boldsymbol{\alpha}$ also depends on η (Equation 23), it appears, at first, that these values must be tuned through an iterative process. However, $\|\delta \mathbf{s}_{2d}\|^2$ and $\|\delta \boldsymbol{\alpha}\|^2$ are related by the singular values of A :

$$w_n^2 < \frac{\|\delta \mathbf{s}_{2d}\|^2}{\|\delta \boldsymbol{\alpha}\|^2} < w_1^2 \quad (24)$$

where w_n and w_1 are, respectively, the lowest and the largest singular values of A . Hence, the following algorithm is used to tune η : At the beginning of the fitting algorithm, A is strongly biased due to the fact that the other parameters (f , ϕ , γ and θ) are far from their optimum, hence we set η to the square of one of the first singular values. Then, as the other parameters reach their optimum, the confidence in A increases, and η is decreased on the spectrum of the singular values of A . More specific detail about this process is given in the next Section.

5 LiST – A Linear Shape and Texture Fitting Algorithm

Based on the inversion of the image formation process detailed in the previous section, we present, in this section, the fitting algorithm which is summarised in Figure 1. The key feature of this fitting algorithm is the exploitation of linear relations to update the shape and texture parameters. Therefore we refer to this algorithm as the *Linear Shape and Texture* (LiST) fitting algorithm.

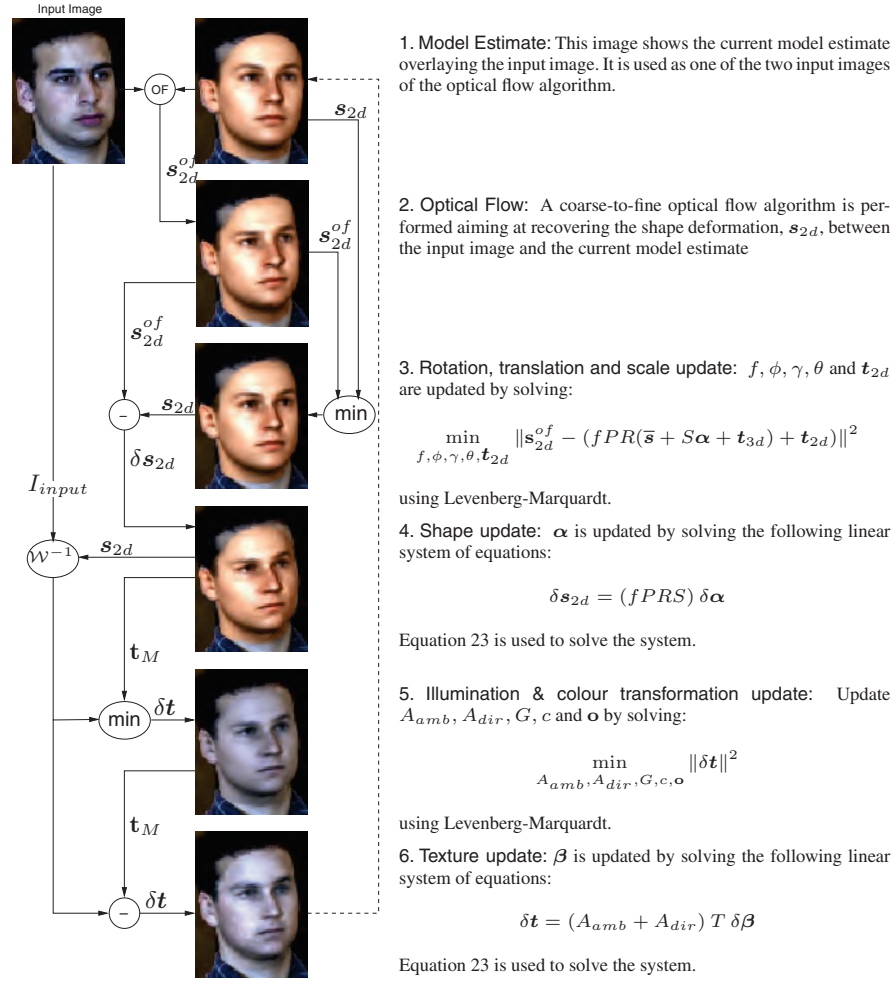


Fig. 1. This figure visually demonstrates the six steps of the first iteration of the LiST fitting algorithm.

1. **Model Estimate:** Before starting the algorithm, the parameters are initialised to their average values ($\alpha = \beta = 0$) except for the pose parameters f, ϕ, γ, θ and t_{2d} , which must be estimated by use of other means (manually, using a face-pose detector or using additional 3D camera and head position information as done in our experiments (see Section 7)). Also, the direction of the light must be known as this parameter is not recovered by the algorithm. Based on these parameters, a model image is generated overlaying the input image. It is used as input of the optical flow algorithm.
2. **Optical Flow:** A coarse-to-fine optical flow algorithm [2] is performed to recover the shape deformation between the input image and the current model estimate. For the first iterations there is a large difference between the two images (mostly due to the fact that the illumination and colour parameters are not recovered yet). This may lead to inaccuracies in the optical flow algorithm. Therefore the algorithm is not applied on the raw RGB values but on the images filtered by a Laplacian kernel. Also, the accuracy of the flow recovered at one pixel depends on the magnitude of the Laplacian at that pixel. Therefore only the flow of the salient pixels whose Laplacian is higher than a threshold (the salient pixels) is used.
3. **Rotation, translation and scale update:** f, ϕ, γ, θ and t_{2d} are updated by minimising the difference between the shape recovered by optical flow and the model shape over the set of salient pixels. This is achieved by solving Equation 17. As the cost function is non-linear, the minimisation is carried out by a Levenberg-Marquardt algorithm [18].
4. **Shape update:** α is updated by solving the linear Equation 18 by use of Equation 23. The value of η is chosen based on the iteration number. At the beginning of the algorithm, the shape derivative is strongly biased due to the fact that the other parameters are far from their optimum. Hence a strong weight on the prior probability must be chosen. Therefore, at the first iteration, η is set to $w_{i_1}^2$, where w_i are the singular values of the gradient matrix, $i_1 = 0.1 n$ and n is the number of singular values (i.e. the number of shape parameters). At the second iteration, $i_2 = 0.2 \cdot n$, and so forth. At the tenth iteration and thereafter, the cost function is not influenced by the prior probabilities anymore.
5. **Illumination and colour transformation update:** Similarly to step 3, A_{amb}, A_{dir}, G, c and $\mathbf{o}$ are updated by solving Equation 19 using a Levenberg-Marquardt algorithm [18].
6. **Texture update:** Similarly to step 4, the texture parameters β are updated by solving the linear Equation 20 by use of Equation 23. The selection of η is the same as explained at step 4. After this step, all the parameters have been successively updated and a new iteration is started at step 1. The iterative algorithm is stopped when $\|\delta t\|^2$ obtained after this iteration becomes stationary.

6 Discussion

Building on our earlier 3D morphable model fitting using stochastic gradient descent (SGD) [5], we now developed LiST, an algorithm which takes advantage of the linear parts of the model. Applied to the problem of face recognition, we achieved with LiST a

similar performance than the SGD algorithm in a more efficient manner (5 times faster). LiST departs from SGD in that it estimates different shape and texture errors which are used to update the shape and texture model parameters separately. The accuracy of the shape estimate depends on the goodness of the texture estimate and vice-versa. Therefore, it is paramount to interleave shape and texture update as the convergence of one helps the convergence of the other. The ability of LiST to recover out of the image plane rotation and directed illumination (see experiments in Section 7) originates from the fact that it is based on a model which closely approximates physical reality.

LiST bears some resemblance with the Active Shape Model (ASM) [10] and with the Vectorizer [3]. The ASM recovers the 2D shape of objects, represented as a sparse number of landmark points, using an equation similar to the Equation 18. Apart from the shape representation (dense physically-based 3D shape as opposed to sparse 2D shape), the main difference between LiST and ASM lies in the shape error estimation. As the ASM lacks a texture model, the shape is recovered by searching the image for edges along the normals of the 2D shape. By contrast, LiST uses the full estimate of the texture model to constraint the optical flow shape recovery, thereby using more image information than the ASM.

The Vectorizer [3] recovers the shape using optical flow between the input image and a model-generated image in a similar fashion to steps 1 and 2 of LiST. However, the Vectorizer lacks a shape model and hence the shape yielded by optical flow is not constrained by a shape prior probability as is achieved by LiST via Equations 18 and 23.

As mentioned in Section 4, the main difference between Image Difference Decomposition algorithms and LiST is the utilisation by LiST of an explicit shape error which allows us to take advantage of the linearities of the shape model to recover the shape parameters. On the other hand, in IDD the shape error is implicitly represented in the texture error which is, as explained in Section 4, a source of non-linearities. As these non-linearities are modelled by a constant Jacobi matrix, IDD is not designed to be used on a general problem such as face image fitting exhibiting pose and light variations. To model only one non-linear degree of freedom (the azimuth angle) Cootes *et al.* had to learn five constant gradient matrices [9]. In presence of combined pose and illumination variation the number of constant Jacobi matrices required to approximate the non-linearities would increase exponentially.

7 Experiments

In this sections, the fitting and identification performances are investigated on the Pose Illumination and Expression (PIE) database from CMU [21] which exhibits combined variations of pose and directed light. None of the 68 individuals in the PIE database is in the training set of 3D scans. We selected two portions of the database to carry out our tests: The first part includes pose variations at ambient light: each individual is imaged from 13 different poses (884 images overall). The second portion presents variation with respect to both pose and directed illumination and with an ambient light. Each individual is imaged at 3 poses and illuminated from 22 different directions (4488 images overall). The cameras and flashes are distributed in an hemisphere in front of the subject as shown in Figure 2.

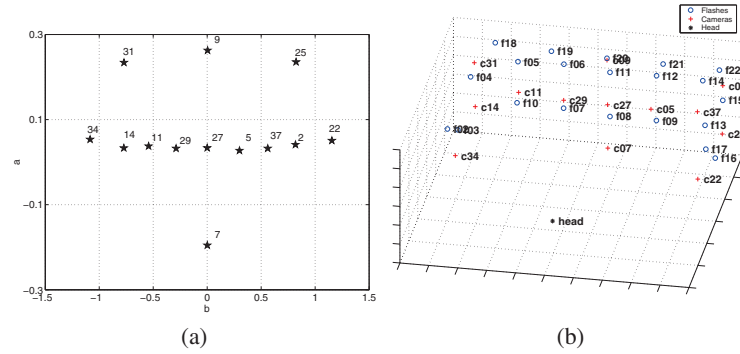


Fig. 2. PIE database camera and light positions. (a) A plot of the azimuth and altitude angles of the cameras, along with the camera ID number. 9 of the 13 cameras sample a half circle at roughly head height ranging from a full left to a full right profile view (± 60 degrees); 2 cameras were placed above and below the central camera; and 2 cameras were positioned in the corners of the room. (b) 3D locations of the cameras, the flashes and the head. Courtesy of [21] and [13].

7.1 Fitting

Before fitting a face image, the face must be detected, its pose estimated and the direction of the light must be known. During execution of the algorithm, it refines the estimate of the face location and pose but leaves the approximation of the light direction constant. The initial value of all the other parameters is their average value. For the experiments on the PIE database, the face pose was estimated automatically from the geometrical information provided by the CMU database. Indeed, the 3D position of the cameras, the light sources and the head were recorded at the time of the data acquisition [21]. These data allow us to recover an initial estimate of the azimuth of the face (ϕ) and the direction of the light $\mathbf{l}$. The two other angles determining the pose of the face (θ and γ) were left to their average value (0). 2D translation and size parameters, t_{2d} and f , were estimated using the 2D positions of the two eyes and the mouth also provided by the maker of the database. As explained in [13], the 2D locations were automatically recovered using FaceIt's face finding module and manually verified.

On average, the algorithm converges in 8 iterations. The time required per iteration is less than a minute on a Pentium III 800 MHz. More than the half of this time is consumed in the optical flow computation. We anticipate an important speed up of the algorithm if we apply it in a coarse to fine approach, however the impact of this implementation on the algorithm performances must be carefully evaluated.

Figure 3 shows five fitting examples. The first line depicts the initial values of the parameters, the second represents the fitted model image and the last shows the original input images. Due to legal reasons, we are allowed to show images of very few individuals only. Unfortunately, this subset of individual is not representative of the whole database as many persons wore glasses and have facial hair.



Fig. 3. Examples of Fitting. **Top row:** Initial parameters, **Middle row:** Fitting result, **Bottom row:** Input image. The fourth column is an example of bad fitting.

7.2 Identification

The fitting procedure provides shape and texture coefficient which, as explained in Section 2, depends of the identity only. Identification of a novel face image is performed by matching the coefficients recovered from the image with the coefficients recovered from the image of a known individual. The matching criterion used is a simple nearest neighbour classification rule using the following correlation-based similarity measure:

$$d = \frac{\langle \alpha, \alpha' \rangle}{\|\alpha\| \|\alpha'\|} + \frac{\langle \beta, \beta' \rangle}{\|\beta\| \|\beta'\|} \quad (25)$$

where the primed symbols pertain to the gallery image (known individuals) and the non-primed to the probe image (novel images). We use the *closed universe* model whereby every probe image shows an individual that can be found in the gallery. To enable past and future algorithm assessment, we made publicly available our identification results on our web page¹: for each probe image we provide the ordered list of gallery images along with the distance d between the gallery images and the probe image.

¹ <http://graphics.informatik.uni-freiburg.de/publications/list>

Pose variations. Here, the performances of the fitting algorithm are evaluated with respect to identification across 13 different poses with ambient light. The gallery contains a single image per individual at the same pose. Each pose is used in turn in the gallery. Table 1 shows the complete confusion matrix. Each of the 13×13 cells refer to an identification experiment of 68 probe images using a gallery of 68 individuals. The last column of the table shows the percentage of correct identification for a particular gallery view averaged on the entire probe set. The only reported results of this set of experiments (performed on the same images using the same initial parameters (eyes and mouth coordinates)) is in [13] using Visionics FaceIt, one of the most successful commercial face recognition system. FaceIt finished as the top performer in the Face Recognition Vendor Test 2000 [4]. Visionics claim that FaceIt can handle pose variation up to 35 degrees in all direction. Our identification results outperform FaceIt in all 169 cells. For front gallery and probe views, our algorithm outperforms FaceIt by a few percents and the improvement of our algorithm compared to FaceIt increases as the probe and gallery pose increase, averaging to more than 60 % correct identification augmentation. **These results clearly show that, for pose-invariant face identification, a 3D representation is valuable.**

Table 1. Confusion table for pose variation. Each row of the confusion table shows the recognition percentage on each of the probe poses given a particular gallery pose. The camera position, indicated by its azimuth and altitude angles is shown in Figure 2. The table layout is courtesy of [13].

azimuth altitude Probe Pose	-66 3	-47 13	-46 2	-32 2	-17 2	0 15	0 2	0 1.9	16 2	31 2	44 2	44 13	62 3	
Gallery Pose	34	31	14	11	29	09	27	07	05	37	25	02	22	mean
34	-	81	84	78	60	56	54	41	43	44	50	53	53	61
31	91	-	96	93	87	82	79	78	69	65	63	57	47	77
14	88	97	-	100	93	78	85	82	68	62	54	62	51	79
11	76	94	99	-	94	85	94	91	74	65	50	53	44	78
29	60	84	90	96	-	94	96	93	78	68	59	54	40	78
09	53	79	81	88	94	-	97	90	87	84	76	69	43	80
27	49	78	87	93	97	99	-	99	99	94	75	76	50	84
07	35	75	81	96	93	90	99	-	93	82	62	62	38	77
05	46	78	85	88	90	93	99	94	-	93	91	85	63	85
37	47	81	78	82	82	85	93	79	94	-	91	90	78	83
25	50	71	72	60	71	84	85	65	82	85	-	88	84	77
02	49	56	65	60	62	71	78	68	75	90	91	-	81	73
22	56	43	51	44	43	49	47	50	57	71	74	84	-	59

Combined pose & light variations. Here, we used the portion of the PIE database containing 4488 images of faces of 68 individuals viewed from 3 poses (front (camera 27), side (cam. 5) and profile (cam. 22)) and illuminated from 21 different directions.

Again only a single image per individual is used in the gallery. Each pose is used in turn in the gallery. We chose light number 12 for the gallery as the fitting is generally fair at that condition. Table 2 shows the confusion matrix for each gallery broken down into probe views and illuminations. It appears that the best generalisation performances are obtained with a side-view gallery. As expected, the poorest performances are obtained by a profile-view. This is probably due to the lack of features at that view: only one eye visible, no nostril edges, only half a mouth visible. The results for generalisation across light variation for similar or near-by poses (front and side) are promising. The decrease in identification performance for profile views is similar to that experienced with ambient light only (see previous paragraph). Therefore we conjecture that this performance drop is due more to pose variation than to light variation. To augment our understanding of the effects of combined pose-illumination variations on identification accuracy, an image database with more poses should be used.

Table 2. Identification results in presence of combined pose and illumination variation. The first column lists the directed light number. The three groups of triple columns show the results obtained for three different galleries: Front, side and profile. For each gallery, the percentage of correct identification for a pose-illumination condition is displayed. See Figure 2 for flash light positions.

light	Front Gallery			Side Gallery			Profile Gallery		
	Front	Side	Profile	Front	Side	Profile	Front	Side	Profile
ambient	97	84	48	88	91	49	57	72	87
2	71	60	22	50	69	31	29	49	51
3	95	78	28	84	87	26	63	54	54
4	98	83	45	85	96	47	54	51	66
5	98	91	65	94	97	65	51	53	81
6	98	89	65	94	100	65	65	62	84
7	98	92	65	96	99	71	57	65	87
8	98	94	48	99	100	71	63	78	96
9	100	97	57	100	100	75	69	88	100
10	98	89	58	93	99	71	60	60	78
11	100	97	72	97	100	82	66	75	93
12	-	98	78	99	-	91	75	78	-
13	98	97	77	100	100	91	82	85	100
14	98	98	83	99	100	91	82	90	100
15	98	97	80	97	100	93	81	90	100
16	95	94	71	99	99	84	87	93	100
17	97	89	75	91	94	87	76	85	100
18	97	85	58	88	96	60	43	47	78
19	97	86	54	88	99	57	51	50	81
20	100	97	72	99	100	75	59	71	85
21	100	98	58	100	100	81	74	79	97
22	97	97	78	97	100	90	74	90	100
mean	97	91	60	93	96	71	65	71	86

8 Conclusion

This paper presented LiST, a novel algorithm for fitting a 3D Morphable Model to face images. LiST exploits the linearity of the model in the recovery of shape and texture parameters leading to a robust and efficient algorithm. The shape parameters are recovered using linear equations from a shape error estimated by optical flow (using the input image and the current texture). The texture parameters are also computed using linear equations from a texture error estimated using the current shape. LiST was evaluated on the difficult problem of combined pose-illumination variation face identification. State of the art results were obtained. However they should be confirmed on a database containing more individuals. LiST opens up avenues for further research: **Initial parameters robustness:** Now that the usefulness of our algorithm has been shown on face identification applications, we want to evaluate the performance of our algorithm with respect to noisy initial parameter estimation.

Efficiency: An efficient way of speeding up the algorithm is to implement it in a coarse to fine fashion. We are planning to evaluate the impact of a coarse to fine implementation on the algorithm performances.

Illumination Recovery: Recently, Basri *et al.* [1] and Ramamoorthi *et al.* [19] have shown that the diffuse part of the illumination was well approximated by a 9D linear space. We want to leverage these findings in our algorithm and recover the light parameters also using linear equations. Additionally, this method would alleviate the constraints on the initial light direction requirements.

Acknowledgement. We thank Simon Baker and Ralph Gross for providing us with PIE, a great face image database and for the support they gave us in the use of it. We also thank Nikos Canterakis for interesting discussions. This material is based upon work supported by the European Research Office of the US Army under contract No. N68171-01-C-9000.

References

1. Ronen Basri and David Jacobs. Lambertian reflectance and linear subspace. In *8th International Conference on Computer Vision*, volume 2, pages 383–390, 2001.
2. J.R. Bergen and R. Hingorani. Hierarchical motion-based frame rate conversion. Technical report, David Sarnoff Research Center Princeton NJ 08540, 1990.
3. D. Beymer and T. Poggio. Image representation for visual learning. *Science*, 272, 1996.
4. D. M. Blackburn, M. Bone, and P. J. Phillips. Face recognition vendor test 2000: Evaluation report. Technical report, DoD Counterdrug Technology Development Program, 2000.
5. V. Blanz, S. Romdhani, and T. Vetter. Face identification across different poses and illuminations with a 3d morphable model. In *Proc. of the 5th Int. Conf. on AFGR*, 2002.
6. V. Blanz and T. Vetter. A morphable model for the synthesis of 3D-faces. In *SIGGRAPH 99 Conference Proceedings*, Los Angeles, 1999. Addison Wesley.
7. Volker Blanz. *Automatische Rekonstruktion der dreidimensionalen Form von Gesichtern aus einem Einzelbild*. PhD thesis, Universität Tübingen, Germany, 2001.
8. T. Cootes, G. Edwards, and C. Taylor. Active appearance model. In *ECCV*, 1998.
9. T. Cootes, K. Walker, and C. Taylor. View-based active appearance models, 2000.

10. T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 1995.
11. J.D. Foley and A. van Dam. *Fundamentals of interactive computer graphics*. The systems programming series. Addison-Wesley, Reading, Ma, 1984.
12. M. Gleicher. Projective registration with difference decomposition, 1997.
13. Ralph Gross, Jianbo Shi, and Jeff Cohn. Quo vadis face recognition? In *Third Workshop on Empirical Evaluation Methods in Computer Vision*, 2001.
14. R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
15. M. Jones and T. Poggio. Multidimensional morphable models: A framework for representing and matching object classes. In *Proceedings of the Sixth ICCV*, 1998.
16. P. J. Phillips, P. Rauss, and S. Der. Feret (face recognition technology) recognition algorithm development and test report. ARL-TR 995, U.S. Army Research Laboratory, 1996.
17. F. Pighin, R. Szeliski, and D.H. Salesin. Resynthesizing facial animation through 3d model-based tracking. In *Proceedings of the 7th ICCV*, pages 143–150, 1999.
18. Vetterling Press, Teukolsky and Flannery. *Numerical recipes in C : the art of scientific computing*. Cambridge University Press, Cambridge, 1992.
19. Ravi Ramamoorthi and Pat Hanrahan. A signal-processing framework for inverse rendering. In *SIGGRAPH 2001 Conference Proceedings*, pages 117–128, 2001.
20. S. Sclaroff and J. Isidoro. Active blobs. In *6th ICCV*, 1998.
21. T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination and expression (pie) database of human faces. Technical Report CMU-RI-TR-01-02, CMU, 2000.