ISSN 2321-6905

International Journal of Science Engineering and Advance Technology

Challenging Issues in Face Recognition

[1]Saurabh Gupta
Research Scholar
Department of IT,
MIT, Muzaffarpur, India
saurabhg.mit@gmail.com

[2]*Dr. R. K. Singh*
Head- Department of Information Technology, MIT, Muzaffarpur, India

[3]*Dr. Kumar Balwant Singh,*
Head – Department of Physics G. P. Darbhanga, India
kbsphysics@yahoo.co.in

Abstract: ***This paper provides a concise & up-to-date review of research efforts in face recognition techniques based on 2D and 3D images. Recent research has also demonstrated that the fusion of different imaging modalities and spectral components can improve the overall performance of face recognition. The majority of face recognition methods have been developed by scientists with a very technical background such as biometry, pattern recognition and computer vision methods using 3D shape applied to data embodying facial expressions. Different strategies to deal with expressions are presented. The underlying concepts and practical issues relating to the application of each strategy are given, without going into technical details. The discussion clearly articulates the justification to establish archival, reference databases to compare and evaluate different strategies.***

Introduction

Anthropometry has existed for many years and has evolved with each advent of new technology and computing power [1]. Face recognition is a natural human ability and a widely accepted identification and authentication method. It is a rapidly growing research area due to increasing demands for security in commercial and law enforcement applications. As a result of this, face recognition methodology has shifted from a purely 2D image-based approach to the use of 3D facial shape.[2] However, one of the main challenges still remaining is the non-rigid structure of the face, which can change permanently over varying time-scales and briefly with facial expressions.

Comparative study of various techniques including algorithms use in face Recognition:

In 2005 Kyong I. Chang et. al [3] report on the largest experimental study to date in multimodal 2D+3D face recognition, involving 198 persons in the gallery and either 198 or 670 time-lapse probe images. PCA-based methods are used separately for each modality and match scores in the separate face spaces are combined for multimodal recognition. Major conclusions are: 1) 2D and 3D have similar recognition performance when considered individually, 2) combining 2D and 3D results using a simple weighting scheme outperforms either 2D or 3D alone, 3) combining results from two or more 2D images using a similar weighting scheme also outperforms a single 2D image, and 4) combined 2D+3D outperforms the multiimage 2D result. This is the first (so far, only) work to present such an experimental control to substantiate multimodal performance improvement.

Gaps:

The topic of 3D face recognition has been only lightly explored; also the topic of multiimage representations of a person for face recognition is even less well explored. We should note that the results reported in this paper are obtained using manually marked eye locations. Thus, these are in a sense "best possible" results since an automatic eye-finding procedure is almost certain to introduce errors. Algorithms for automatically locating landmark points on the face are another area in which more research is needed. Currently, 3D scanners do not operate with the same flexibility of conditions of lighting, depth of field, and timing as normal 2D cameras. Thus, 3D face imaging requires greater cooperation on the part of the subject. Also, some 3D sensing technologies, such as the Minolta, are "invasive" in the sense that they project light of some type onto the subject. Clearly, another important area of future research in 3D face recognition is the development of better 3D sensing technology.

After this in year 2006 Kevin W. Bowyer et. al [4] examines face recognition using normal intensity images, infrared images, three-dimensional shape, and combinations of these. He compares the performance improvement obtained by combining three-dimensional or infrared with normal intensity images (a multimodal approach) to the performance improvement obtained by using multiple intensity images (a multisample approach). Combining results from different types of imagery gives significantly higher recognition rates than are obtained by using a single intensity image. However, significantly higher recognition rates are also obtained by combining results from multiple intensity images. Overall, initial results indicate that, using an Beigenface recognition algorithm and weighted score fusion, multisample techniques can result in a performance increase comparable to that of multimodal techniques.

Gaps:

This is the only work to compare face recognition using normal 2-D images, range images representing 3-D shape, and infrared images, and also the only work to evaluate the multimodal combination of the three types of images. Experimental results are based on an image dataset that has images of the same persons in each of the three modalities. For a given person at a given acquisition session, the 2-D, 3-D, and IR images are acquired over a time interval of just a few minutes. This should allow the training, gallery, and probe sets for each modality to contain comparable images in the different modalities.

He used the same PCA-based recognition engine, with the face space tuned individually for each modality, and all landmark points marked manually for each modality & found that 3-D resulted in a higher rank-one recognition rate than 2-D, but that the difference was not statistically significant. He also found that the rank-one recognition rate for IR imagery was statistically significantly lower than that for 2-D or 3-D. However, the range of lighting conditions used in the image acquisition was typical of indoor office environments and this may not show off the strength of IR sensing.

We also compared the performance of individual modalities with multiple modalities. We found that each of the multimodal performances improved over all of the individual modalities, and that the multimodal 2-D 3-D & IR technique performed best of all. The differences between the various multimodal performances were found not to be statistically significant. However, all of the multimodal performances were quite high, making it difficult to reliably detect differences. Additional investigation using a larger and more challenging dataset might reveal performance differences that were not detected here.

Also, for the experiments described in this paper, which use the same basic Eigen-face algorithm and score-level fusion in comparing multisample versus multimodal, the multisample approach with intensity images achieves performance equivalent to the multimodal approach. However, using four identical intensity images will result in the same performance as using one image.
It is worth noting once more that the images used in this study were all approximately frontal view and acquired under reasonably good lighting conditions. In conditions of very low light, infrared images can produce results where normal intensity images could not. And in conditions of extreme non-frontal pose, 3-D face shape may be able to produce useful results where normal intensity images could not.

Lastly, to achieve the maximum possible performance, it seems reasonable that the eventual solution could be some combination of multisample and multimodal. Either multisample alone or multimodal alone could be expected to reach a plateau in performance at some number of samples or modalities. Achieving performance greater than this may required a combination of multiple samples of multiple modalities.

Again In 2006 Xiaoguang Lu[5], et.al examines the performance of face recognition systems that use two-dimensional images depends on factors such as lighting and subject's pose and developed a face recognition system that utilizes three-dimensional shape information to make the system more robust to arbitrary pose and lighting. For each subject, a 3D face model is constructed by integrating several 2.5D face scans which are captured from different views. 2.5D is a simplified 3D (x, y, z) surface representation that contains at most one depth value (z direction) for every point in the (x, y) plane. Two different modalities provided by the facial scan, namely, shape and texture, are utilized and integrated for face matching. The recognition engine consists of two components, surface matching and appearance-based matching. The surface matching component is based on a modified Iterative Closest Point (ICP) algorithm. The candidate list from the gallery used for appearance matching is dynamically generated based on the output of the surface matching component, which reduces the complexity of the appearance-based matching stage. Three-dimensional models in the gallery are used to synthesize new appearance samples with pose and illumination variations and the synthesized face images are used in discriminate subspace analysis. The weighted sum rule is applied to combine the scores given by the two matching components. Experimental results are given for matching a database of 200 3D face models with 598 2.5D independent test scans acquired under different pose and some lighting and expression changes. These results show the feasibility of the proposed matching scheme.

Gaps:
He proposed a combination scheme, which integrates surface (shape) matching and a constrained appearance-based method for face matching that complement each other. The surface matching is achieved by a hybrid ICP scheme. The registered 3D model is utilized to synthesize training samples with facial appearance variations, which are used for discriminate subspace analysis. The matching distances obtained by the two matching components are combined using the weighted sum rule to make the final decision. Regardless of the pose, lighting, and expression, given the feature points, the entire matching scheme is fully automatic, including surface registration/ matching, dynamic candidate list selection, 3D synthesis, sample image cropping, LDA, and appearance- based matching.

This research was an encouraging first step in designing a system that is capable of recognizing faces with arbitrary pose. Non rigid deformation such as expression is a challenge to the current system. More sophisticated surface matching schemes are being pursued to improve the surface matching accuracy and speed. We are exploring3D models that can be deformed to deal with non-rigid variations

In the year 2007 Ajmal S. Mian et al.[6] Presents a fully automatic face recognition algorithm and demonstrate its performance on the FRGC [Face recognition Grand challenge] v2.0 data. Algorithm is multimodal (2D and 3D) and performs hybrid (feature based and holistic) matching in order to achieve efficiency and robustness to facial expressions. The pose of a 3D face along with its texture is automatically corrected using a novel approach based on a single automatically detected point and the Hotelling transform. A novel 3D Spherical Face Representation (SFR) is used in conjunction with the Scale-Invariant Feature Transform (SIFT) [7] descriptor to form a rejection classifier, which quickly eliminates a large number of candidate faces at an early stage for efficient recognition in case of large galleries. The remaining faces are then verified using a novel region-based matching approach, which is robust to facial expressions. This approach automatically segments the eyes, forehead and the nose regions, which are relatively less sensitive to expressions and matches them separately using a modified Iterative Closest Point (ICP) algorithm. The results of all the matching engines are fused at the metric level to achieve higher accuracy. He uses the FRGC benchmark to compare the results to other algorithms that used the same database. His multimodal hybrid algorithm performed better than others by achieving 99.74 percent and 98.31 percent verification rates at a 0.001 false acceptance rate (FAR) and identification rates of 99.02 percent and 95.37 percent for probes with a neutral and a non neutral expression, respectively.

Results: Several novelties were presented while addressing major problems in the area of 3D and multimodal face recognition. His contributions include:
1. A fully automatic pose correction algorithm,
2. AN SFR for 3D faces,
3. A novel SFR-SIFT-based rejection classifier, and
4. A region-based matching algorithm for expression robust face recognition.

Although these algorithms have been applied to 3D face recognition, they can easily be generalized to other 3Dshapes. In addition to these novelties, we, for the first time in the literature, successfully used the 3D nose as an independent biometric. He addressed three major challenges, namely, automation, efficiency, and robustness to facial expressions.

The performance of algorithms was demonstrated on the largest publicly available Corpus of 3D faces data. The performance of the SFR-SIFT rejection classifier was 0.036, which amounts to 27.78 times improvement in recognition time. Our multimodal hybrid recognition algorithms achieve 99.74 percent and 98.31 percent verification rates at 0.001 FAR for faces with a neutral and a non neutral expression, respectively. The identification rates for the same were 99.02 percent and 95.37 percent. In terms of accuracy, these results are slightly better than any previously published results. This is quite understandable as there was little room for improvement. The individual verification rate at 0.001 FAR of 3D region-based matching algorithm alone is 98.5 percent, which is a strong indicator of the potential of 3D face recognition. He shows that the eyes-forehead and nose regions of a face contain the maximum discriminating features important for the expression of robust face recognition.

Mohammad H. Mahoor et. al [8] in 2008 presented a fully automated multimodal (3-D and 2-D) face recognition system. For the 3-D modality, authors model the facial image as a 3-D binary ridge image that contains the ridge lines on the face. He use the principal curvature max to extract the locations of the ridge lines around the important facial regions on the range image (i.e., the eyes, the nose, and the mouth.) For matching, he utilize a fast variant of the iterative closest point to match the ridge image of a given probe image to the archived ridge images in the database. The main advantage of this approach is reducing the computational complexity by two orders of magnitude by relying on the ridge lines. For the 2-D modality, he models the face by an attributed relational graph (ARG), where each node of the graph corresponds to a facial feature point. At each facial feature point, a set of attributes is extracted by applying Gabor wavelets to the 2-D image and assigned to the node of the graph. The edges of the graph are defined based on Delaunay triangulation and a set of geometrical features that defines the mutual relations between the edges is extracted from the Delaunay trianglesand stored in the ARG model. The similarity measure between the ARG models that represent the probe and gallery images is used for 2-D face recognition. Finally, he fuse the matching results of the 3-D and the 2-D modalities at the score level to improve the overall performance of the system. Different techniques for fusion, such as the Dempster–Shafer theory of evidence and weighted sum of scores are employed and tested using the facial images in the third experiment dataset of the Face Recognition Grand Challenge version 2.0.

GAPS:
For the multimodal system, He fused the matching results of the 2-D and 3-D modalities at the score level. He compared the DS theory of evidence with the weighted sum technique for fusion. Although both techniques produced comparable results, the DS theory of evidence has the advantage that it does not require the calculation of any parameters for fusion. In particular, if the number of modalities increases, finding the optimum weights becomes hard (finding a global optimum point in a high-dimensional space is a challenging problem).

The findings were on improving the performance of the ridge modelling technique for the recognition of 3-D facial images with expressions on a large database, such as FRGC V2.0. In addition, he should extend the ARG modelling technique to both the 2-D and 3-D data. This approach will integrate both the intensity and the shape information of the facial images in one model. Based on the availability of data from each modality (i.e., 2-D and 3-D), the recognition task using each individual modality alone or by fusing data from the two modalities can be handled by this approach in a unified manner using an integrated ARG model.

Continuing the preceding works in FR in 2009 Martin D. Levine et al [9] proposed 3D facial reconstruction systems attempt to reconstruct 3D facial models of individuals from their 2D photographic images or video sequences. Currently published face recognition systems, which exhibit well-known deficiencies, are largely based on 2D facial images, although 3D image capture systems can better encapsulate the 3D geometry of the human face. Accordingly, face recognition research is gradually shifting from the legacy 2D domain to the more sophisticated 2D to 3D or 2D/3D hybrid domain.

Currently there exist four methods for 3D facial reconstruction. These are: Stochastic Newton Optimization method (SNO) ICIA + 3DMM, and LiST can be classified as ‘‘analysis-by-synthesis” techniques and SAIMC can be separately classified as a ‘‘3D supported 2D model”. In this paper, He introduce, discuss and analyze the difference between these two frameworks. He begin by presenting the 3D morphable model which forms the foundation of all four of the reconstruction techniques. This is followed by a review of the basic ‘‘analysis-by-synthesis” framework and a comparison of the three methods that employ this approach. He next review the ‘‘3D supported 2D model” framework and introduce the SAIMC method, comparing it to the other three. The characteristics of all four methods are summarized in a table that should facilitate further research on this topic.

Conclusion: Although both major frameworks involve shape parameter estimation, the procedures are significantly different. To begin with, the ‘‘3D supported 2D models” framework requires several feature points on the input image.The step that achieves shape parameter estimation, referred to as ‘‘shape alignment” in the ‘‘3D supported 2D model” approach, fits the 3D morphable model (3DMM) to these feature points only, while non-feature points are not considered by the ‘‘shape alignment” procedure. This is appreciably different from the ‘‘analysis-by-synthesis” framework that updates shape parameters globally (all pixels in 2D images will be taken into consideration by the fitting procedure). This is why the ‘‘3D supported 2D models” framework does not require a gradient descent technique, the most time-consuming step, to minimize a non-linear objective function as do SNO, ICIA + 3DMM, and LiST. But the ensuing superior efficiency of ‘‘3D supported 2D models” more or less achieved by sacrificing the accuracy of shape reconstruction. In SAIMC, a shape alignment is used to reconstruct the shape based on the manually selected 84 feature points. Non-feature points do not contribute to the actual reconstruction but the correction does improve the accuracy of the (x, y) coordinates of the overall estimated shape. Unfortunately, the z (depth) information, which is dominated by the alpha parameters obtained in the earlier ‘‘shape alignment” process, cannot be fixed. This is the main predicament associated with the ‘‘correction”. Moreover, the relevant shape transformation parameters (e.g., rotation matrix, scale, focal length of the camera, and translation vector) are tainted in the ‘‘3D supported 2D models” case because only one frontal 2D image is used as the input.

As was explained above, the ‘‘analysis-by-synthesis” and ‘‘3D supported 2D models” frameworks are almost completely different from each other except for the fact that both utilize a 3D morphable model (3DMM). 3DMM is the only intersecting aspect of these two reconstruction frameworks. The efficiency of the algorithms is estimated as
_ SNO: 4.5 min
_ ICIA + 3DMM: on average 30 s.
_ LiST: 54 s (rough estimate).
_ SAMIC: 10 s and ‘‘fifteen times faster than LiST”
Table 2 summarizes the qualitative differences between the four methods.

Finally, yet importantly, in Table 2, we observe that no quantitative or comparative analysis of the reconstruction accuracy for any of the four methods has appeared in the literature. This remains to be done in the future.

The methodologies of the general face recognition algorithms in each stage of face recognition.

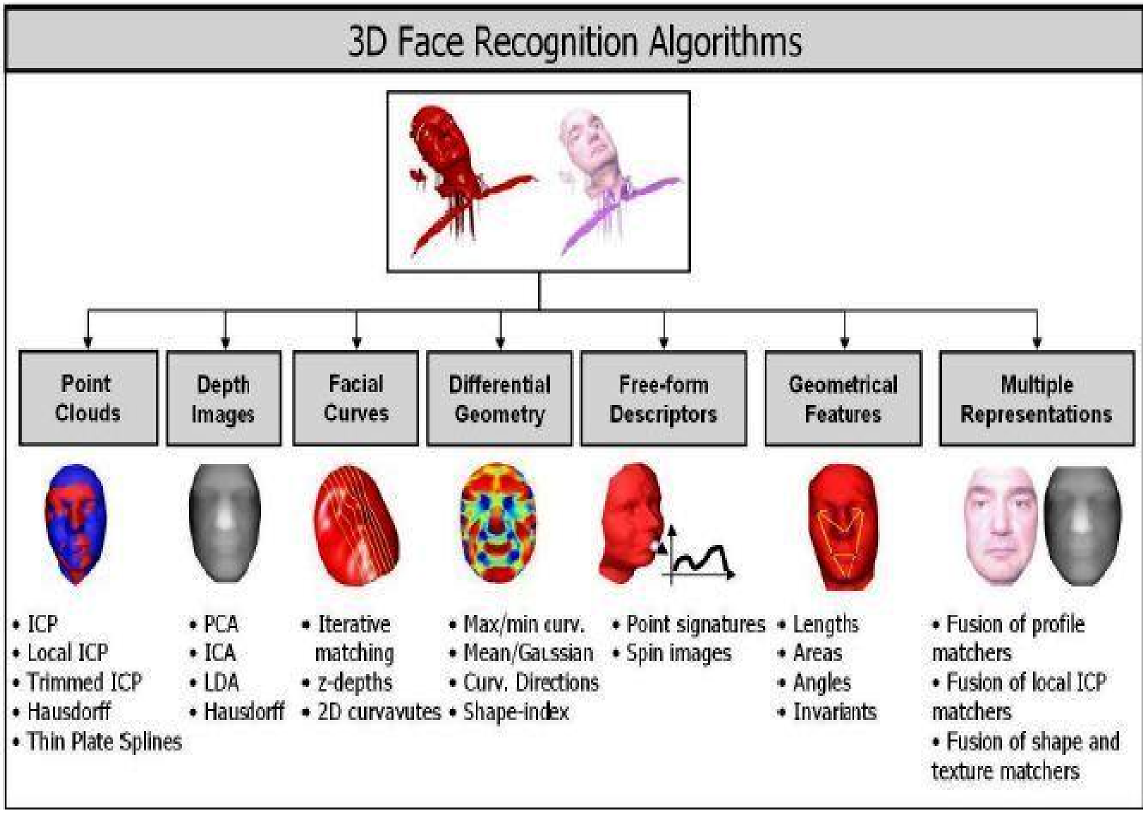
Approach	FR stage			
	Region-based representation	Feature extraction	Dimension reduction	Classification
Eigenfaces [74,75]	Holistic	Pixel intensity	Principal component analysis	Nearest neighbour
Fisherfaces [7]	Holistic	Pixel intensity	Linear discriminative analysis	Nearest neighbour
SOM+CN [47]	Evenly distributed image patches	Pixel intensity	Self organising map	Convolutional network
LEM [27]	Holistic	Line edge map	Line segment Hausdorff distance	Nearest neighbour
DCP [28]	Holistic	Local directional corner points	Minimum warping cost	Nearest neighbour
Template matching [16]	Patches around eyes, nose, and mouth	Pixel intensity	None	Normalised correlation
Modular PCA [61]	Regions around eyes, nose, and mouth	Pixel intensity	Principal component analysis	Nearest neighbour
EBGM [79]	Regions around 31 facial component points	Gabor wavelet	Normalised correlation	Averaging
LBP [2]	Evenly distributed image patches	Local gradient binary codes	Histogram	Weighted Chi square

Filareti Tsalakanidou et. al [10] in 2010 suggests in his paper a completely automated facial action and facial expression recognition system using 2D 3D images recorded in real time by a structured light sensor. It is based on local feature tracking and rule based classification of geometric, appearance and surface curvature measurements. Several experiments conducted under relatively non controlled conditions demonstrate the accuracy and robustness of the approach.

Conclusion: A fully automated system for facial action unit detection and facial expression recognition in sequences of 2D and 3D images was presented in this paper. The proposed system is based on a novel real-time model-based face tracker and a set of special local feature detectors, which effectively combine 3D face geometry and 2D appearance data. The use of 3D information facilitates detection of surface deformations even in case of subtle facial muscle movements. Facial action is represented by a set of geometric, appearance-based and surface-based measurements, which are effectively classified into emotional related expressions using a rule-based approach. A method for detecting temporal events related to action unit or facial expression activation periods was also proposed. The proposed techniques were evaluated in a large database with more than 50 subjects and 800 sequences demonstrating increased accuracy and robustness under pose variations. Future work will exploit the dynamics of facial measurements towards automatic decoding of all action units and their combinations. 3D information will be further exploited for facial feature tracking and facial action unit recognition. More specifically, the standard ASM fitting technique, which is based on image gradient profiles derived from 2D images, will be extended to include local 3D surface information in the form of curvature or surface gradient descriptors. Such an approach is expected to offer increased localization accuracy as well as increased robustness against pose and illumination variations. In addition, new curvature measurements will be proposed for detecting action units related to the mouth and chin. Finally, the proposed techniques will be extended to cope with large head poses.

Further enhancing this work the year 2011 Utsav Prabhu et al [11] formulates and analysed the Classical face recognition techniques that have been successful at operating under well-controlled conditions; however, authors have difficulty in robustly performing recognition in uncontrolled real-world scenarios where variations in pose, illumination, and expression are encountered. In this paper, He proposes a new method for real-world unconstrained pose-invariant face recognition. He first construct a 3D model for each subject in our database using only a single 2D image by applying the 3D Generic Elastic Model (3D GEM) approach. These 3D models comprise an intermediate gallery database from which novel 2D pose views are synthesized for matching. Before matching, an initial estimate of the pose of the test query is obtained using a linear regression approach based on automatic facial landmark annotation. Each 3D model is subsequently rendered at different poses within a limited search space about the estimated pose, and the resulting images are matched against the test query. Finally, He compute the distances between the synthesized images and test query by using a simple normalized correlation matcher to show the effectiveness of our pose synthesis method to real-world data. He presented very convincing results on challenging data sets and video sequences demonstrating high recognition accuracy under controlled as well as unseen, uncontrolled real-world scenarios using a fast implementation.

Taxonomy of 3D face recognition systems:



Challenge for 3D face recognition: Improved methology:

One barrier to experimental validation and comparison of 3D face recognition is lack of appropriate datasets. Desirable properties of such a dataset include: (1) a large number and demographic variety of people represented, (2) images of a given person taken at repeated intervals of time, (3) images of a given person that represent substantial variation in facial expression, (4) high-spatial resolution, for example, depth resolution of 1 mm or better, and (5) low frequency of sensor-specific artefacts in the data. Expanded use of common datasets and baseline algorithms in the research community will facilitate the assessment of the state of the art in this area.
It would also improve the interpretation of research results if the Statistical significance, or lack thereof, was reported . Another aspect of improved methodology would be the use, where applicable, of explicit and distinct training, validation, and test sets.

A more subtle methodological point is involved in the comparison of multi-modal results to results from a single modality.[12] Multi-modal 3D + 2D performance is always observed to be greater than the performance of 2D alone. However, as explained earlier, this comparison is generally biased in favour of the multi-modal result. A more appropriate comparison would be to a 2D recognition system that uses two images of a person both for enrolment and for recognition. When this sort of controlled comparison is done, the differences observed for multi-modal 3D + 2D compared to ‘‘multi-sample’’ 2D are smaller than those for a comparison to simple 2D research issue of how to select the best set of multiple samples of a given modality is one that could be important in the future.

References:

[1] Gupta, S.; Markey, M. K.; Bovik, A. C. (2010). "Anthropometric 3D Face Recognition". *International Journal of Computer Vision (IJCV)* 90 (3): 331–349. doi:10.1007/s11263-010-0360-8.

[2] Daniel Riccio, Jean-LucDugelay “Geometric invariants for 2D/3D face recognition”*Pattern Recognition Letters*, *Volume 28, Issue 14*, *15 October 2007*, *Pages 1907-1914*

[3] Chang, K.I.; Bowyer, K.W.; Flynn, P.J.; “An evaluation of multimodal 2D+3D face biometrics” Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume: 27 , Issue: 4 Digital Object Identifier: 10.1109/TPAMI.2005.70 Publication Year: 2005 , Page(s): 619 - 624 Cited by: 70

[4] Bowyer, K.W.; Chang, K.I.; Flynn, P.J.; Xin Chen; “Face Recognition Using 2-D, 3-D, and Infrared: Is Multimodal Better Than Multisample? ” Proceedings of the IEEE Volume: 94 , Issue: 11 Digitalobjectidentifier:10.1109/JPROC.2006.885134 Publication Year: 2006, Page(s): 2000-2012 Cited by: 5

[5] Xiaoguang Lu; Jain, A.K.; Colbry, D.; “Matching 2.5D face scans to 3D models” Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume: 28 , Issue: 1 Digital Object Identifier: 10.1109/TPAMI.2006.15 Publication Year: 2006 , Page(s): 31 - 43 Cited by: 58

[6] Mian, A.S.; Bennamoun, M.; Owens, R.; “An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition “Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume: 29 ,Issue: 11 Digital Object Identifier:10.1109/TPAMI.2007.1105 Publication Year: 2007 , Page(s): 1927 - 1943 Cited by: 40

[7] Mohamed Aly ,”Face Recognition using SIFT Features” International journal of computer vision, 60, 2006.

[8] Mahoor, M.H.; Abdel-Mottaleb, M.; “A Multimodal Approach for Face Modeling and Recognition” Information Forensics and Security, IEEE Transactions on Volume: 3 , Issue: 3 Digital Object Identifier: 10.1109/TIFS.2008.924597 Publication Year: 2008 , Page(s): 431 - 440 Cited by: 4

[9] Martin D. Levine, Yingfeng (Chris) Yu “State-of-the-art of 3Dfacial reconstruction methods for face recognition based on a single 2D training image per person” *Pattern Recognition Letters*, *Volume 30, Issue 10*, *15 July 2009*, *Pages 908-913*

[10] Filareti Tsalakanidou, Sotiris Malassiotis “Real-time 2D+3D facial action and expression recognition “ Original Research Article *Pattern Recognition*, *Volume 43, Issue 5*, *May 2010*, *Pages 1763-1775*

[11] Prabhu, U.; Jingu Heo; Savvides, M.; “Unconstrained Pose-Invariant Face Recognition Using 3D Generic Elastic Models Pattern Analysis and Machine Intelligence,”IEEE Transactions on Volume:33 , Issue: 10 Digital Object Identifier:10.1109/TPAMI.2011.123Publication Year: 2011 , Page(s): 1952 - 1961

IEEE Journals

[12] Bronstein, A. M.; Bronstein, M. M., and Kimmel, R. (2006). "Three dimensional face recognition" *International Journal of Computer Vision (IJCV)* 64 (1): 5–30. doi:10.1007/s11263-005-1085-y