Study of different techniques for face recognition

Rohan Lade[1] , Harshal Meshram[2], Ankit Kulkarni[3], Prashik Ramteke[4], Aditya Badale[5]
Department of computer science engineering
Jd college of engineering and management
Kalmeshwar road valni Nagpur 441501.

***Abstract*— The most significant part of face recognition is the input representation. This refers to the transformation of the intensity map to a form of input representation that allows easy and effective extraction of highly discriminative features. The next stage is classification. Although this is an important stage, the popular techniques and their strengths are fairly similar to each other. As such, the choice of classification algorithm does not affect the recognition accuracy as much as input representation. Input representation is the major factor that differentiate face recognition algorithms. It can be approached in 2 manners: a geometrical approach that uses spatial configuration of the facial feature, and a more pictorial approach that uses image-based representation. In this paper a set of different face recognition algorithms are reviewed, and the best practices in this domain are studied and verified.**

***Keywords*— Face, detection, machine learning, classification**

I. INTRODUCTION

For face recognition, there are numerous geographical approaches: seminal work in [1]. These are feature-based systems that start by searching for facial features, such as the corners of eyes, corners of mouth, sides of nose, nostrils, and contour along the chin, etc. Feature location algorithms usually based on a heuristic procedure that rely on edges, horizontal and vertical projections of gradient and grey levels, and deformable templates.

The geometry details of a facial feature are captured by feature vectors that include details such as distant, angle, curvature, etc. The implementation method mentioned above uses a feature vector with dimension ranging from 10 to 50. Craw and Cameron's system [2]. uses a feature vector that represents feature geometry by displacement vectors from an "average" arrangement of features. This effectively measures how the particular training face differs from the norm. Once faces are represented in the feature vector, the classification or similarity measurement is usually performed by computing the Euclidean distance or a weighted norm, where dimensions are usually weighted by some measure of variance.

To identify an unknown test face, geometry-based recognizers choose the model closest to the input image in feature space. This approach to face recognition has been limited to frontal views, as the geometrical measurements used in the feature vector changes according to face rotations out of the image plane (x-y plane).The second major type of input representation is the graphical or pictorial approach. This represents faces by filtered images of model faces. In template-based systems, which is the simplest pictorial representation, the faces are represented either by an intensity map of the entire face, or by part-images/sub-images of salient facial feature such as eyes, nose, and mouth. The input to template-based system is usually but not necessarily face images. Some systems use gradient magnitude or gradient vector to take advantage of the better immunity to lighting conditions.

Given a test face, this is then compared to all model templates. The typical way to measure image distant is correlation. In[3] uses normalised correlation on grey level templates. This system motivated the famous template based approach in[4]. that uses normalised correlation on gradient magnitude. In fact, this system has gone so far that Gilbert and Yang [5]. implemented this system in a custom-built VLSI chip to perform real-time face recognition. In[5] uses a hierarchical coarse-to-fine structure to represent and match templates. In[6] uses a template that take advantage of x and y components of the gradient.The next section describes these techniques in details.

II. LITERATURE REVIEW

Principal component analysis (PCA) is used for both recognition and face reconstruction. PCA can be described as an optimised pictorial approach. The dimension size of the image space is reduced to form the face space. Dimension reduction is very significant as only a small percentage of the original number of dimensions is used for classification. The original dimensionality is equal to the number of pixel of the face image. In the face space, the dimension is reduced to the number of eigenfaces [7]. The eigenspace is the face representation framework. To apply PCA, it is assumed that the set of all face images is a linear subspace of all grey level images. The eigenfaces are the basis that span the face space. The eigenfaces are found by applying principal components to a series of faces. In the face space, faces are represented by coefficients that mark the projection onto the corresponding basis.

in[8] were the first ones to apply principal component analysis to face recognition. In[9] first preprocess the face image using Fourier transform magnitudes to remove the affect of translation. In[10] applied PCA to shape-free faces, i.e. faces where the salient features have been moved

to a standardised position. In[11] used PCA on template of major facial features, achieving result comparable to that of correlation but at a fraction of the computational cost. In[12] applied PCA to a series of problems, including recognition of frontal views in a large database of over 3000 subjects, recognition under varying rotation out of the image plane, and detection of facial features using eigen-templates. In[13] have demonstrated that faces can be accurately reconstructed from their face space representation.In addition to PCA, other analysis techniques have been applied to face images to generate a new and more compact representation when compared to the original image space. In[14] represented a face by using autocorrelation on the original grey level images. 25 autocorrelation kernels of up to 2nd order are used, and the subsequent 25-D representation is passed through a traditional Linear Discriminant Analysis classifier. In [15] have applied Singular Value Decomposition (SVD) to the face image where the rows and columns of the image are actually interpreted as a matrix. Cheng et al. used SVD to define a basis set of images for each person, which is similar to face space of that of the PCA. The only difference is that each person's image has its own face space. Hong creates a low dimensional coding for faces by running the singular values from SVD through linear discriminant analysis.

In [16] have used vector quantization to represent faces after the faces are broken down to their important facial features. The face is represented by a combination of indices of best-matching templates from a codebook. The number one issue is how to choose the codebook of feature templates, In[17] have used "isodensity maps" to represent faces. The original grey level histogram of the face is divided up into eight buckets, defining grey level thresholds for isodensity contours in the image. Faces are represented by a set of binary isodensity lines, and face matching is performed using correlation on these binary images.

Connectionist approach to face recognition also use pictorial representations for faces [18]. Since the networks used in connectionist approaches consist only of classifiers, these approaches are similar to the ones described above. In multiplayer networks where simple summating nodes are used, inputs such as grey level images are applied at input layer. The output layer is usually arranged as one node per object and its activity determines the object reported by the network.

The input to the network is a set of training faces. They are fed into the network to train the network using a learning procedure that adjusts the network parameters, usually known as weight. Among connectionist approaches to face recognition, the two most important issues are input representation at the input layer and the overall network architecture. As previously mentioned, the input representations are pixel-based, with [19] using the original grey level images. [20] used directional edge maps, [21] used a threshold binary image, and [22] used Gaussian units applied to the grey level image. A variety of network architectures have been used. A plain multiplayer network trained by back propagation is the most common approach, and was used in [21] and [22]. A rather similar technique, [23] used a radial basis function network with gradient decent. [24] used a recurrent auto-associative memory. It recalls the closest pattern to the applied unit currently in memory. A multiplayer cresceptron is used by [25]. This is a derivative of Fukushima'sNeocognition. The network by [26] works on binarized images, using a network architecture of the sum of set of 4-tuple AND functions.

Hybrid representation methods combine both geometrical and pictorial approaches. In[27] explored a 5D feature vector that stores distances and intensities. This vector is used as a "first cut" filter on the face database. A least squares fit of eye template is used as the final match. In another hybrid approach, In[27] represented faces as elastic graphs of local textural features. This technique is also used by [28]. The graphs' edges capture the feature geometry, which stores information such as distance between two incident features. To represent pictorial information, the graph vertices stores the result of Gabor filters applied to the image at feature locations. The recognition process begins by deforming the input face graph to match model graphs. Then by combining measures of the geometrical deformation and the similarity of the Gabor filter responses, a match confidence is calculated. Although this technique is commonly known as elastic graph matching or in neural net terms as a dynamic link architecture, its mechanism is effectively representing and matching flexible templates.

The meaning of invariance to distortion refers to the ability of the face recognition algorithm to tolerate distortion such as changes in pose, lighting condition, expression etc. Distortion refers to the content of difference between the training face and the test face. Most systems described in the previous section provide some degree of flexibility by

using representations or performing an explicit geometrical normalisation step.

An invariant representation is one that remains constant even if the input has changed in certain respects. A band-pass filter or low-pass filter is able to offer a limited degree of immunity to lighting condition. For example, a Laplacian band pass filter removes low frequency component. This assumes that the image content due to lighting condition such as shadows cast on a face is mainly comprised of low frequency information. As such, higher frequency texture information is preserved. This is true to a certain extent. However, low frequency components contain much of the information that is not due to the lighting condition, especially the face's original content. In other words, the majority of the information content comprises low frequency components.

For translation invariance, some approaches transform the intensity map of face image to frequency domain. Working in the frequency domain, the position of the face will not alter the frequency pattern. A Fourier transform is most widely used for this. However, such transformation does not provide invariance to scale and in-image place rotation. In[29] described a mechanism that is able to tolerate distortion due to a 4 degree of freedom: translation (move along x-axis + y-axis = 2 freedom), scaling and rotation (within the image plane). The first step is to apply Fourier transform to provide translation invariance. Next, the Cartesian image representation is transformed into a complex logarithmic map. This is a new representation where scale and in-image plane rotation become translational parameters in the new space. Applying the Fourier transform again would automatically add invariance to in-image plane rotation and scaling. The problem of out of image plane rotation (rotate about x-axis + y-axis = 2 degree of freedom) is commonly known as the pose problem. So far, nobody has discovered an input representation that is invariant to the pose problem. That is to say there is no input representation that will remain unchanged when the pose problem exists.

If the system is able to determine two points on the face where it knows the standard position and standard distant between them, then the face can be normalised for translation, scale and in-image plane rotation. In geometrical approach, distances in the feature vector are normalised for scale by dividing it by a given distance such as the interocular distance or the length of the nose. In template-based system, faces are often geometrically normalised by rotating and scaling the input image to place the eyes at fixed locations. Out-of-image plane rotation could not be treated by this technique, which is necessary if the system was to handle general pose. Most systems are able to tolerate 4-degree of freedom out of the 6.

By treating the process to allow invariance to the problems mentioned previously as a pre-processing stage, most face recognition systems' core recognition and matching algorithm accept a 2D image as input. As such one could consider most face recognition systems as 2D systems. There have been a few systems that tried to obtain invariance to out-of-image plane rotation. These systems use the technique of storing multiple training views of different poses. In[30] used four slight rotated training faces (up, down, left, right) in addition to the frontal view. In[31] used a database of 5800 faces to construct a Linear Discriminant Analysis classifier. His database consists of 116 subjects with 50 faces per subject. The image source is from a video tape session where each change between each snapshot produces slight pose difference. In[32] used an elastic graph matching technique to match the frontal view to the test face subjected to out of image plane rotation, and to test faces with facial expressions different to that of the training face. All of the techniques and methods described up to this point offer some degree of invariance to pose, lighting condition and facial expression. However, the strength of invariance differs. This thesis presents a new face recognition that offers strong invariance to lighting condition and facial expression.

Although a face recognition algorithm can be very mathematical in nature, the experiment to test its accuracy and robustness is highly empirical, being based on statistical information. As with any experiment that relies on statistical results, the sample size is very important. Sample size refers to the number of training and test faces, both in terms of the number of subjects and the number of views per subject. The difference between the test face set and the training face set is very important as well. The system is considered weak if it performs well only when the test face closely resembles the training face. In other words, the magnitudes of the distorting parameters are minimal. The result then becomes meaningless if the face recognition system is intended only for low level of distortion. The key ability of a robust system is its ability to generalize from a set of training examples.The building of a face recognition system, assuming that the algorithm has been chosen, begins by collecting face images. The set of face images collected is divided into training faces and test

faces. The training faces are processed and transformed into the system's representation format, which is usually a geometry feature vector or a set of templates. For neural network approach, this set of face images is used to train the network.After processing the training faces, recognition begins by feeding the system with a series of test faces. There are two levels of testing, depending on whether the system performs a rejection test on a labelled test face or not.

III. CONCLUSION AND FUTURE SCOPE

The system reviewedin this research achieved 100% recognition rate using a database with 42 subjects. Using 108 faces from outside the database, the resulting false access rate is 0%. Template-based system achieved 100% recognition rate on frontal view of 47 subjects. Other systems used a database of 50 subjects and reached 96% recognition rate. Eigenface system produced a result of 96% being correctly classified using a database of 16 subjects under varying lighting conditions. Some works produced 100% recognition rate using database of 11 subjects. Moreover, others achieved 95% recognition rate with their system using over 3,000 people from the FERET database, produced by the US Army Research Lab, which is by far the largest research-oriented face database ever produced. Otsu and Sato plotted the recognition rate versus false access rate. Their system showed that to achieve recognition rate well over 90%, the false access rate hiked to an unacceptable level.However, the validity to compare the results quoted so far is questionable. This is simply because none of them share the same database, and none used a sufficiently large database for concrete statistical conclusion. The FERET database contains more than 7,500 faces from over 3,000 subjects. The database is constructed to mimic real-world application as close as possible. Different view of the same subject is taken at different time, up to weeks apart.

References

[1] Aitchison A. C., Craw I., "Synthetic images of faces – an approach to model-based face recognition", In Proc. British Machine Vision Conference, 226-232, 2020

[2] Akamatsu S., Sasaki T., Fukamachi H., Masui N., Suenaga Y., "An accurate and robust face identification scheme", In Proc. Int. Conf. on Pattern Recognition, Vol 2, 217-220, 2020

[3] Akimoto T., Suenaga Y., Wallace R. S., "Automatic creation of 3-D facial models", IEEE Computer Graphics and Applications, 13(5):16-22, 2013

[4] Baron R. J., "Mechanisms of human facial recognition", Int. Journal of Man Machine Studies, 15:137-178, 1981

[5] Belhumeur P., Hespanhu J., Kriegman D., "Eigen faces vs. fisherfaces: Recognition using class specific linear projection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 19, July 2017

[6] Beymer D., "Face recognition under varying pose", In Proc. IEEE Conf. On Computer Vision and Pattern Recognition, 756-761, 2014

[7] Beymer D., "Pose-Invariant Face Recognition Using Real and Virtual View", A.I. Technical Report No. 1547, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 2016

[8] Beymer D., Poggio T., "Face recognition from one example view", AI Memo No. 1536, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 2015

[9] Beymer D., Shashua A., Poggio T., "Example based image analysis and synthesis", AI Memo No. 1431, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 2013

[10] Bichsel M., "Strategies of robust object recognition for the automatic identification of human faces", PhD Thesis, ETH, Zurich, 2020

[11] Bichsel M., Pentland A. P., "A simple algorithm for shape from shading", In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 459-465, 2020

[12] Blanz V., Vetter T., "A Morphable Model For The Synthesis Of 3D Faces", Proc. of SIGGRAPH, 2019

[13] Braje W L., Kersten D., Tarr M. J., Troje N. F., "Illumination Effects in Face Recognition", Psychibiology 26, 371-380, 2018

[14] Brotosaputro G., Qi Y., "Finding the estimated position of human facial features by using intensity computation", Electronic Journal of the School of Advanced Technologies Asian Institute of Technology, Vol. 1, Issue 2, 2019

[15] Bruce V., "Changing faces: Visual and non-visual coding processes in face recognition", British Journal of Psychology, 73, 105-116, 1982

[16] Bruce V., Humphreys G., "Recognising objects and faces", Visual Cognition, 1(2/3), 141-180, 2014

[17] Bruce V., Recognizing Faces, Lawrence Erlbaum Assoc., London, 1988

[18] Bruce V., Valentine T, Baddeley A., "The basis of the ¾ view advantage in face recognition", Applied Cognitive Psychology, 1, 109-120, 1987

[19] Brunelli R., "Estimation of pose and illuminant direction for face processing", AI Memo No. 1499, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 2014

[20] Brunelli R., Poggio T., "Face recognition: features versus templates", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 15.10, 1042-1052, 2013

[21] Burt P. J., "Multiresolution techniques for image representation, analysis, and 'smart' transmission", In SPIE Vol. 1201, Visual Communications and Image Processing IV, 2-15, 1989

[22] Cannon S. R., Jones G. W., Campbell R., Morgan N. W., "A computer vision system for identification of individuals", In Proc. IECON, 347-351, 1986

[23] Chellappa R., Wilson C., Sirohey S., "Human and machine recognition of faces: A survey", Proceedings of IEEE, Vol 83, No. 5, 2015

[24] Chen C. W., Huang C. L., "Human face recognition from a single front view", Int. Journal of Pattern Recognition and Artificial Intelligence, 571-593, 2020

[25] Cheng Y. Q., Liu K., Yang J. Y., Wang H. F., "A robust algebraic method for human face recognition", In Proc. In. Conf. on Pattern Recognition, vol 2., 221-224, 2020

[26] Chow G., Li X., "Towards a system for automatic facial feature detection", Pattern Recognition, Vol. 26, No. 12, 1739-1755, 2013

[27] Cootes T. F., Taylor C. J., "Active shape models – smart snakes", In David Hogg and Roger Boyle, editors, Proc. British Machine Vision Conf., 266-275, Springer Verlag, 2020

[28] Craw I., Cameron P., "Face recognition by computer", In David Hogg and Roger Boyle, editors, Proc. British Machine Vision Conf., 498-507, Springer Verlag, 2020

[29] Craw I., Tock D., Bennett A., "Finding face features", Proc. ECCV, Europe, 92-96, 2020

[30] Cyberware, United States, http://www.cyberware.com

[31] Davies G. M., Ellis H. D., Shepherd J. W., "Face recognition accuracy as a function of mode of representation", Journal of Applied Psychology, 63, 180-187, 1978

[32] Desimone R., "Face-selective cells in the temporal cortex of monkeys", Journal of Cognitive Neuroscience, 3(1):1-8, 2020