International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 01,January-2015

Inter-Image Quality Measure Assessments for Face Recognition under Realistic Conditions

Bhagya.B
Assistant Professor
Department of Computer Applications
College of Engineering
Trivandrum

Jisha Babu
Assistant Professor
Department of Computer Applications
College of Engineering,
Trivandrum

Jose T Joseph
Assistant Professor
Department of Computer Applications
College of Engineering, Trivandrum

Shilpa M Thomas
Assistant Professor
Department of Computer Application
College of Engineering,Trivandrum

***Abstract*— Face recognition consists of finding the name of the person (class label) by using the trained model. Face recognition becomes complex due to variation in illumination, pose variation and occlusions in images. They are considered as outliers and removal of these outliers makes the image classification easier. The outliers due to illumination are removed by normalization. Saliency based image quality measures changes with the occlusion and pose variation. So it is used for image classification.**

Keywords—Saliency map, Attention based image quality measure.

I. INTRODUCTION

The past three decades have seen the evolution of numerable face recognition techniques. Out of which, many are very effective. The major problem noticed in most of them is the time consumption. Recent researches in this field focus on optimizing the time and space factors. Initial researches on face recognition have failed because of adopting pixel wise comparison over images. Pixel wise comparison demands waiting time of over a week to identify the efficiency of these methods in single as well as dual core systems. This paper adopts a technique which focuses on the image which gains high attention during perception. So the total processing depends on the pixels in that portion alone. Therefore the pixels on the highly attended region are processed rather than the entire pixels.

.

II. BACKGROUND REVIEW

Existing system performs well under ideal situations. Some face recognition systems (FRS) considers the outliers independently. i.e., some FRS gives good results by using images with only illumination or pose variation or expression or occlusion. No existing FRS can recognize the face properly when there are all the combinations of variability in the images. The challenges in automatic face recognition are to provide robust face recognition under variability in illumination, occlusion, expression, and pose variation.

A. Illumination

Same face appears differently due to the change in lighting. More specifically, the changes induced by illumination could be larger than the differences between individuals, causing systems based on comparing images to misclassify the identity of the input image. For example, the popular eigen-subspace projections used in many systems as features have been analyzed under illumination variation in [5]. Illumination changes cause dramatic changes in the projection coefficient vectors, and hence can seriously degrade the performance

B. Pose variation

Same face appears differently due to changes in viewing condition. Moreover, when illumination variation also appears in the face images, the task of face recognition becomes even more difficult. Using such a model, the difficulty of the pose problem can be assessed and the efficiency of existing methods can be evaluated systematically.

C. Face Occlusion

The presence of elements like beards, glasses or hats introduces high variability. Faces can also be partially covered by objects or other faces. There has been very little work towards explicitly handling facial occlusion. In [6] and [7], it was shown that two "eigen feature" images–the eyes and the nose –could be used for accurate recognition after a change in facial hair. However, no method for selecting the appropriate eigen features was suggested. In the LDA approach described in [8], some promising results were obtained after artificially degrading face images, indicating that LDA might also provide a reasonable solution to handling some degree of decoration, assuming the registration landmarks can still be located.

Pose and Expression

AAM methods [9, 10] have been proposed to handle both varying pose and expression. No recognition results, however, are presented to demonstrate performance.

D. Pose and Light varriations

Methods were presented in [11] and [12] to deal with varying pose and illumination. These methods rely upon generative models that can synthesize a given face under varying illumination from different viewpoints. Although the performance in [13] is quite remarkable, the proposed method employs seven training images for each subject under strictly controlled lighting and does not address expression or decoration.

III. PROBLEM DEFINITION

Face recognition consists of finding the name of the person (class label) by using the trained model where the training dataset (Tr) is used for learning.

Tr = {(X1,Y1),(X2,Y2),.....(Xi,Yi)}

X is pattern which is build up of group of features, and it is denoted as,
X = {a1,a2,.....ai}
where a=feature and Y=label
Te = {(X1,U),(X2,U),.....(Xj,U)}
where U=unknown label
To build a model for face recognition, the following steps are done to each image in the training dataset. Generally, face recognition contains the following phases,

- Detecting part of the image which contains the face
- Normalize the image to remove noises
- Extract features from each image and group it into patterns
- Give a label to images which contains a particular pattern

Face recognition faces some problems during feature extraction. Some of the challenges during feature extraction are the presence of different expressions, pose variations, occlusions and illuminations in the images. Image variability due to these factors makes the feature extraction complex. Dominant features may overridden by these factors. A standard image quality measure is necessary to extract features from images in presence of large image variability. It can correctly classify the images with large possible outliers.

IV. PROPOSED SCHEME DESCRIPTION

The proposed system considers all the possible image variability, and improves the system to an extent with large accuracy. This system is expected to perform well for all types of images with combinations of different natural variability and even with the presence of all possible outliers as a whole.
When the photographic conditions change due to illumination, there can be large differences between images. To eliminate these differences normalize image by adjusting its mean to zero and standard deviation to one. This is the method called normalization. Scatter plots are used to identify the inter image outliers. The pixel values which fall beyond the threshold is remove as an outlier. So image variability due to occlusions and pose variations can be removed. Then calculate the mean of the inter image pixel value which is free from outliers. Based on this mean value, the images are classified. The images having the minimum inter pixel mean value is the best match.

A. *Proposed Algorithm*

1. Find the saliency map image of all images in the database.
2. Calculate the mean of the saliency map and then remove the pixels which are less than the mean value. So that the removal of outlier is done as,
$$\begin{cases} \delta(i,j) & if\ \delta(i,j) \le \theta \\ \varphi & if\ \delta(i,j) > \theta \end{cases}$$
3. Normalize the test and training image. So that effect of natural variability can be removed.
4. Find the difference between normalized test and training images and then calculate distance similarity by subtracting it from unity.
5. Then multiply the distance similarity and binarized saliency of test image.
6. Repeat steps 3 to 5 for all comparisons between a test image and training images.
7. The result obtained in step 4 is ranked and find the accuracy of recognition.

B. *Image Quality Measures*

Generally the image quality measures are classified into objective quality measures and subjective quality measures. Subjective measures are more accurate because it involves the human perception. But it is very time-consuming and it requires human viewers. Objective quality measure is a value which is calculated by using the statistical features of an image. There are varieties of objective image quality measures. The 6 objective quality measures considered here are:

1) Mean Square Error (MSE): Mean Squared Error is the average squared difference between a reference image X and a test X'. Let X(m,n) and X'(m,n) are reference image and test image, where M and N are the number of pixels in row and column. If the images have high similarity then the value of MSE is less. MSE is calculated using equation (1)

$$MSE = \frac{1}{MN}\sum_{M=1}^{M}\sum_{N=1}^{N}\left(X(M,N) - X'(M,N)\right)^2 \quad (1)$$

2) Peak Signal to Noise Ratio(PSNR): Peak Signal-to-Noise Ratio is the ratio between the maximum possible power of an image and a power of corrupting noise, given in decibels. PSNR is calculated using equation (2)

$$PSNR = 10\log_{10}\left(\frac{max^2}{MSE}\right) \quad (2)$$

3) Average Difference(AD): Average difference is the average of the difference between reference image X and text image X'. AD is calculated using equation (3)

$$AD = \frac{X(m,n) - X'(m,n)}{MN} \quad (3)$$

If the two images have high degree of similarity then the value of AD is less.

4) Maximum Difference(MD): Maximum difference is the maximum of difference between two images. MD is calculated using equation (4)

$$MD = Max\left[X(m,n) - X'(m,n)\right] \quad (4)$$

The value of MD is less when the two images have high degree of similarity.

5) Normalized Absolute Error(NAE): Normalized absolute error is the ratio of normalized difference of reference image and test image to the normalized reference image. NAE is calculated using equation (5) . The value of NAE is less when the two images have high degree of similarity.

$$NAE = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N}|X(m,n) - X'(m,n)|}{\sum_{m=1}^{M}\sum_{n=1}^{N}|X(m,n)|} \quad (5)$$

*6) Structured content (SC):*Structural content is the ratio of the squares of reference image to that of test image. SC is calculated using equation (6)

$$SC = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N}X(m,n)^2}{\sum_{m=1}^{M}\sum_{n=1}^{N}X'(m,n)^2} \quad (6)$$

The value of SC is close to 1 when the two images have high degree of similarity.

Fig.1 and Fig.2 are used in the simulation and results are shown in Table 1.

Fig. 1. Original image

Fig. 2. Distorted image

TABLE I. COMPARISON OF ORIGINAL IMAGE AND DISTORTED IMAGE

Image Quality Measures	Best match	For original image and distorted image
MSE	0	792.7585
PSNR	INF	19.1394
AD	0	-1.5020
MD	0	108
NAE	0	0.1831
SC	1	0.9159

One of the images is the original image of Lena and other is the noisy image of Lena. MSE shows large dissimilarity between those images, because it takes square of the differences. So the small presence of noise will make high dissimilarity. For the best match, MSE will be 0.

PSNR and MD also show high dissimilarity. For the best match, PSNR will be Infinity and MD will be 0.
AD and NAE will be 0 for the best match. But AD=-1.5020 and NAE=0.1831, when comparing the Lena with and without noise. It is not a large variation. But SC shows a comparatively close match with each other. Because SC=1 for the best match, but here 0.0841is the only variation from 1when comparing lena with noise and Lena without noise.

The above mentioned image quality measures are not capable of giving proper similarity measure. Because image variability masks the prominent features present in the image. So a standard image quality measure is needed, which can outperform on an image with high variability. An image quality measure, which can extract the perception or attention features, can only find the correct degree of similarity. This attention or perception based image quality measure contains the element of subjective quality measure called perception
.

7) Attention Based Image Quality Measures: When we observe a scene, human eye filters the data in the scene, and locate the attention to a particular region based on its visual properties. Visual attention analysis is divided into two categories: top-down and bottom-up. In top-down approach, there is a prior-knowledge about the image, or goal such as, finding a particular object or some information about the context. It is called goal-oriented approach. In the bottom-up approach, there is no prior knowledge of the image.
Attention is driven by the visual properties of the scene. Visual properties of the scene are varying according to the color, contrast and orientation. Viewers usually pay more attention to the region where some information or object is more prominent than others. This region is called salient region. When we look into a scene, high salient region is attended first, and then the next high salient region and so on.

V. FACE DATABASE

AR database: Images of 100 people (over 2600 color images) are included in the AR database. Different facial expressions, illumination conditions and occlusions added during the photography session. Two sessions per person (2 different days) are performed. This face database was created by Aleix Martinez and Robert Benavente in the Computer Vision Center (CVC) at the U.A.B. The database contains 26 different variant images of each person. Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). The pictures were taken at the CVC under strictly controlled conditions. No restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to participants. Each person participated in two sessions, separated by two weeks (14 days) time. The same pictures were taken in both sessions. This face database is publicly available. It is free for academic use.

VI. EXPERIMENT RESULT

The algorithm works well with most of the illumination conditions. But some of the facial expressions and combination of expression with occlusions decreases the accuracy of algorithm. This is due to the fact that prominent features get masked due to occlusion and pose variation. Table II shows the percentage of recognition accuracy when comparing class 1of the AR database as the training image with other classes in the AR database as test images. The overall accuracy obtained is 60%.

TABLE II. PERCENTAGE OF RECOGNITION ACCURACY OBTAINED BY USING AR DATABASE

Comparison of training image (1)with other classes	Accuracy(%)
2	85
3	76
4	31
5	91
6	86
7	73
8	77
9	67
10	58
11	67
12	56
13	43
14	77
15	59
16	65
17	13
18	82
19	69
20	57
21	52
22	41
23	29
24	51
25	41
26	28

VII. RELATED WORK

The saliency map is a two dimensional map whose activity topographically represents visual saliency. That is, an active location in the saliency map encodes the fact that this location is salient. The idea behind saliency is first proposed by Koch and Ullman in 1985. Thereafter, many researchers are proposed variety of saliency maps by considering various visual features.

There are two popular computational models for visual attention.

- Itti's model
- Winner-take-all approach

Itti's model analyses the image looking for the most relevant visual properties: color, intensity and orientation. The algorithm creates a map corresponding to these 3 properties and combines these maps to predict the saliency of the regions. The predicted fixations are selected as the most salient points in the image using winner-take-all approach. Winner-take-all network detects the higher saliency at any given time. In order to combine the attention information with the objective quality metrics, a saliency is needed, instead of the individual (ordered) fixation points given by itti's algorithm. To obtain such a map, set of predicted fixation points are filtered out using Gaussian filter. For this, inhibition of return process suppresses the last attended location from the saliency map, so that attention can focus on the next most salient location.

In [1] and [2], the concept of saliency map is discussed in which, made a saliency map based on an assumption; the salient region has very different low level features such as color, intensity, and orientation than the others. The bottom-up model of saliency is introduced in which multi scale low level feature extraction is performed on the input image. The low level features are converted to feature maps and then it is normalized within the range [0, M], where M is the global maximum, which is used to define the size of the image. Then feature maps are integrated to form the 'saliency map' or 'master map'.

A natural indicator of occlusion is the T-junction, a photometric profile shaped like a "T", which is formed where the edge of an object occludes a change in intensity in the background. T junctions and line terminators are not considered in this model. In [1] and [2], performance varies with certain noises and performance falls when complex scene is considered. It shows that if the noise has the same features of salient objects, then the system shows incorrect salient regions.

Consider the Fig.3, and its saliency map in Fig.4. Fig.5 is the saliency map overlayed on Fig.3 using Itti-koch approach.

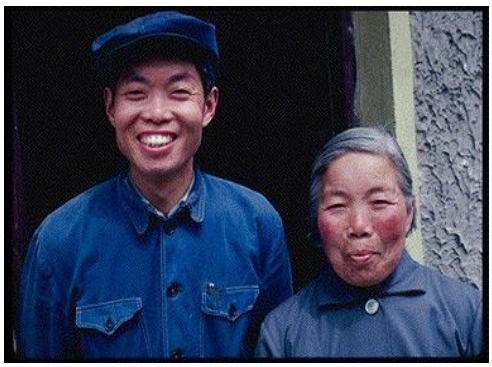

Fig. 3. Faces image

Fig. 4. Saliency map of faces image using Itti.Koch method

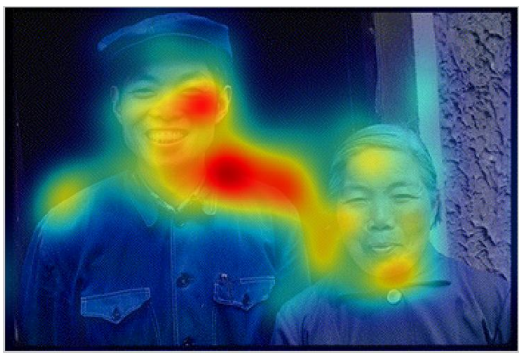

Fig. 5. Saliency map(Fig.4)overlayed on Fig.3

The saliency map in Fig.4 could not reflect the proper salient regions. Because if we look into an image containing faces, the region which contain faces gain more attention. But saliency map shows part of faces with a region containing background as salient.

In [3], a graph based approach is proposed. It is a bottom-up saliency model. Bottom-up saliency model means, it considers only the instantaneous sensory input and it does not consider the internal state of the organism, for finding the salient region. It is an improved saliency map than the itti-koch saliency map. The graph based saliency map is obtained by the following steps:

1. Extract feature vectors at locations over the image plane.
2. Form activation maps using feature vectors.
3. Normalize the activation map.

Fig.6 is the saliency map using graph based approach [3]. Fig.7 is the saliency map(Fig.6) overlayed on Fig.3Graph based saliency map shows more part of the faces as salient as compared to the itti-koch saliency map.

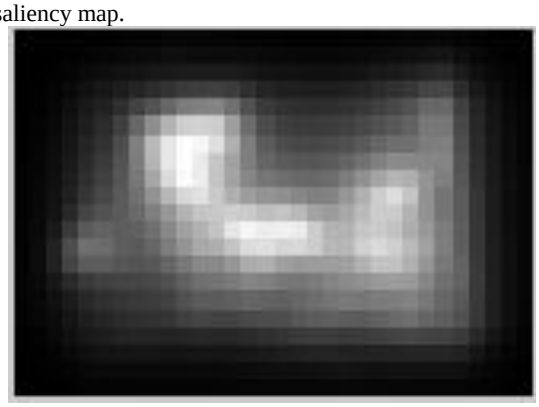

Fig. 6. Saliency map of faces image (Fig.3) using Graph based method

Fig. 7. Saliency map(Fig.6)overlayed on Fig.3

In [4], saliency map based on image signature is introduced.

Image signature (image) =sign (DCT (image))

Image signature is an approach to figure ground separation problem. Figure-ground separation problem means, the difficulty in distinguishing objects from its surroundings. Color input is resized to a coarse 64x48 pixel representation. Then for each color channel, the saliency map is formed from the image reconstructed from the image signature. It uses only 3 color channels at a single spatial task. So it has less computational complexity. This descriptor can be used to approximate the spatial location of a sparse foreground hidden in a spectrally sparse background.

Fig.8 is the saliency map using image signature approach [4]. Fig.9 is the saliency map overlayed on Fig.3

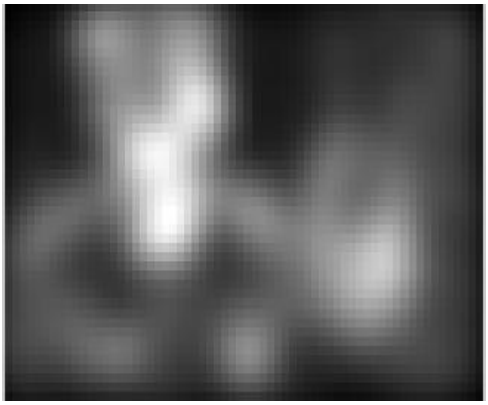

Fig. 8. Saliency map of faces image (Fig.3) using image-signature method.

Fig. 9. Saliency map (Fig.8) overlayed on Fig.3

Graph based saliency is more efficient than others. Because it shows more parts of the face as salient compared to other methods. So graph based saliency is used for experiments.

VIII. CONCLUSION

The visual attention based image quality measures can correctly recognize the faces with large outliers. This measure is used as a local measure for pixel-wise comparison. The illuminated pixels can be removed by normalization. So visual attention based quality measures can perform well in face recognition. So this can be useful for biometric scanners, object searching, and fingerprint verification. The drawback of existing system is expected to be improved to an extent.

REFERENCES

[1] Laurent Itti, Christof Koch." Computational modelling of visual attention" Nature reviews neuroscience, 2001 - papers.klab.caltech.edu

[2] L. Itti, C. Koch, and E. Niebur. "A model of saliency-based visual attention for rapid scene analysis." IEEE Trans. on Pattern Analysis and Machine intelligence, 20(11):1254–1259, 1998.

[3] Harel, Koch, and Perona, P. 2007. "Graph-based visual saliency". In *Proc. NIPS 19*, 545–552.

[4] Xiaodi Hou, Jonathan Harel, Christof Koch: "Image Signature: Highlighting Sparse Salient Regions". IEEE Trans. Pattern Anal. Mach. Intell. 34(1): 194-201 (2012)

[5] B. Moghaddam and A. Pentland. "Probabilistic visual learning for object representation". IEEE Transactions on Pattern Analysis and Machine Intelligence,19(7):696–710, July 1997. face recognition.

[6] A. Pentland, B. Moghaddam, and T. Starner. "View-based and modular eigen spaces for face recognition". In Proceedings of the

IEEE Computer Society Conference on Pattern Recognition, pages 84–91, 1994.

[7] T. Cootes, K. Walker, and C. Taylor. “View-based active appearance models”. Automatic Face and Gesture Recognition, 2000. Proceedings.

[8] T. Cootes, G. Wheeler, K. Walker, and C. Taylor. “Coupled view active appearance models”. In Proceedings of the British Machine Vision Conference, volume 1, pages 52–61, 2000.

[9] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. “From few to many:Illumination cone models for face recognition under variable lighting and pose”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(6):643– 660, June 2001.

[10] W. Y. Zhao and R. Chellappa. “SFS based view synthesis for robust face recognition”. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, pages 285–292, 2000.

[11] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(6):643– 660, June 2001.