Fast Template Matching Method Based Optimized Sum of Absolute Difference Algorithm for Face Localization

Nadir Nourain Dawoud
Dept. Electrical and Electronic
Universiti Teknologi PETRONAS
Tronoh, Malaysia

Brahim Belhaouari Samir
Dept. Fundamental and Applied Science
Universiti Teknologi PETRONAS
Tronoh, Malaysia

Josefina Janier
Dept. Fundamental and Applied Science
Universiti Teknologi PETRONAS
Tronoh, Malaysia

ABSTRACT
Recently, Template matching approach has been widely used for face localization problem. Normalized Cross-correlation (NCC) is a measurement method normally utilized to compute the similarity matching between the stored faces templates and the rectangular blocks of the input image to locate the face position. However, there is always an error on locating the face due to some non-face blocks seem more to be the face position than correct blocks because of variation either in illumination or image with clutter background. In this paper we proposed a fast template matching technique based Optimized Sum of Absolute Difference (OSAD) instead of using NCC to reduce the effects of such variation problems. During the experiments a number of similarity measurements tested to prove the high performance of OSAD compared with other measurements. Two sets of faces namely Yale Dataset and MIT-CBCL Dataset were used to evaluate our technique with success localization accuracy up to 100%.

General Terms
Face localization, Template matching.

Keywords
Similarity measurements, Sum of Absolute Difference.

1. INTRODUCTION
Face localization is the first step in any automatic recognition system where it is a spatial case of face detection. In face localization problem there is already an existing face in the input image and the goal to determine the location of this face. However, face localization from input image is a challenging task due to variation in scale, pose, occlusion, illumination, facial expressions and clutter background. Despite of numerous methods have been proposed to detect the faces in input image, there is still a need to improve the performance of localization and detection methods. A survey on face recognition and some detection techniques can be found in [1]. The more recent survey mainly on face detection was written by Yang et al. [2]. They classify face detection methods into four main categories as follows: Knowledge-based [3], Feature invariant method [4]Template matching [5]Appearance-based method [6]. Template matching approaches have been widely used to locate human faces in the input images. In such approaches, first human face samples (mainly frontal face) are predefined and stored in the system database. Later, correlation between the input image blocks and stored face samples are computed to locate the face. The advantages of these methods that they are robust to noise, simple to implement and it does not take a long time to locate the candidate faces from input images [7]. However, it is not sufficient for detecting faces in images with high variation in background and illumination due to concern on face features shapes which are the effects of these variations. One of the simplest methods of template matching methods is the average face obtained from set of face samples then stored in the database. Then the rectangular blocks in the input image with high similarity correlation score is propose to be the face position. This method can be called filter match method where the input image is convolving with flipped version of the average face as filter. Statistically, filter matching assumes additive white Gaussian noise (AWGN) which is very bad for image variation such as clutter back ground, illumination and expressions [7]. To reduce the effect of high face variation problem, Eigenfaces approach is adopted to enhance matched filter performance [8] which makes linear combination for Eigenfaces of the average face and it assumes that each face should be closed to this linear combination. However, Eigenfaces approach has its own problem where it reflects the variation in the face and in the noise as well [9]. Due to this problem, there is always some localization error where non-face blocks may give high matching similarity to the linear combination of average face and its Eigenfaces more than the face blocks. Therefore, Eigenfaces methods can give good detection rate when the noise is white noised clutter. Meng et al. [7] proposed a new method to localize the human faces using linear Discriminant from gray scale image. To minimize the Bayesian error they developed an optimal Discriminant template by modeling faces and non-faces as Gaussian distribution. In addition, they compared their result with the matched filter and Eigenfaces methods and it was 92.7% using University of Michigan face database. One of the methods which are widely used to compute the correlation between average face template and rectangular blocks of the input images is similarity measurements such as Normalized Cross Correlation (NCC) [10, 11]. However, NCC is affected by illumination and clutter background problems because sometimes there are non-face blocks that have almost the same value of the average face template matrix. This problem can be solved by using Sum of Absolute Differences algorithm (SAD) [12] which is widely

used for image compressing and object tracking but still SAD needs more optimization to give more accurate positions for face in the input image. Moreover, SAD can give high localization rate for facial where the image is with high illumination variation but it may be affected by variation in background.

In this paper, we propose a fast face localization technique based on OSAD instead of using NCC to reduce the effects of such variation problems. The rest of the paper is organized as follows: in the next section we introduce our proposed technique. In section 3, experimental results and similarity measurements comparison are presented and final conclusion with future work is introduced in section 4.

2. PROPOSED METHOD

In our proposed technique, first the average face will be obtained from a set of training sample (see Figure 1). Then an input image will be divided into rectangular blocks. After the face template was computed it is now called average face. Then the OSAD algorithm will be used to show which image window will be either equal to zero or near to zero. Then the location of this window will be determined and extracted from the image to give the face location. To understand the methodology of that we need to define the meaning of the difference in the different spaces, and then explain OSAD algorithm. Moreover, we need to explain the optimum image window in order to minimize the error rate and be able to increase the accuracy of image window determination. To explain the meaning of the difference, a number of mathematical definitions based on the space of the representation can be given and the following spaces will present that:

1. The difference between two points in one dimension can be formulated as follow:

$$d(A,B)=|x_1-x_2| \quad (1)$$

2. The difference between two points in two dimensions can be formulated as follow:

$$d(A,B)=\sqrt{(x_1-x_2)^2+(y_1 \_ y_2)^2} \quad (2)$$

3. The difference between two functions f(x) and g(x), can be formulated as follow:

$$d(f,g)=\int|f(x)-g(x)|\,dx \quad (3)$$

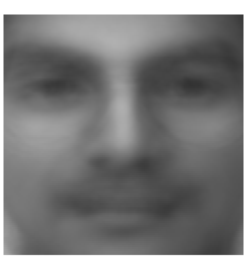

Fig 1: Average image of all persons.

4. The difference between two matrices A and B and it can be formulated as follow:

$$d(A,B)=A-B \quad (4)$$

Now that the meaning of difference in different spaces is clear, we can give a definition for the SAD algorithm.

Actually, this algorithm is widely used; it's simple, and very easy to implement in order to find a relationship between the image windows. This algorithm is based on calculating the difference between each point in the image window and the corresponding point in the template window will be used for comparison. Then, these differences will be added together to measure the similarity between two images. There are many applications for SAD such as motion estimation, object recognition and video compression. (See Figure 2) can give an example of SAD method and the subtraction will produce a new matrix:

1	2	3
4	5	6
7	8	9

A

6	3	0
2	5	1
8	7	1

B

Fig 2: Difference between two matrixes.

-5	-1	3
2	0	5
-1	1	8

Fig 3: Resulting matrix.

In this matrix (see Figure 3) there are some negative values. Therefore we will take the absolute value of all matrix elements and then sum up these elements. The result of this summation gives SAD between the image window and template window. SAD can be computed by using the equation:

$$d(A,B)=\sum_i\sum_j|A(i,j)-B(i,j)| \quad (5)$$

SAD =5+1+3+2+0+5+1+1+8=26

In contrast with the other common correlation based similarity methods namely Sum of Squared Difference (SSD), Normalized Cross Correlation (NCC) and Sum of Hamming Distances (SHD), SAD is simple, more accurate and less complex. While, sometimes it is not accurate if two image windows have the face and they have almost the same SAD. So to improve this algorithm we need to normalize Eq. (5) to find the optimum image window that contained the face. The following equation gives Optimized SAD (OSAD):

$$d(A,B) = \sum_i \sum_j \frac{|A(i,j) - B(i,j)|}{max(A(i,j), B(i,j))} \quad (6)$$

Now, the proposed method can be summarized as follow:

- Compute the template image of the face (average image) by adding a number of the face templates together.
- Divide the original image into number of square windows as same size of the face template.
- Match the template image with the original image windows and calculate SAD of each window by using Eq. (6).
- Save this result in new matrix with each value corresponds to the original window.
- Select the minimum value in the stored matrix and determine the location of the corresponding image window.
- Remove the other image window and save the new image which will represent the face only.

3. EXPERMENTS AND RESULT

Two sets of faces namely Yale Dataset and MIT-CBCL Dataset were used to evaluate our technique.

3.1 Yale Dataset

The dataset established by Yale University [13]. We have taken 11 images of 15 persons with total of 165 images for our experiment. The images of each person are with different facial expression or configuration such as: center-light, with/without glass, happy, sad, left-light, with/without glass, normal, right-light, sad, sleepy, surprised, and wink. Few examples of these images are shown (see Figure 4).

3.2 MIT-CBCL Dataset

This dataset is established by Center for Biological and Computational Learning (CBCL) in Massachusetts Institute of Technology (MIT) [14]. It has 10 subjects with 200 images per subject. The images of each subject are with different light conditions, clutter background, scale and different poses. Few examples of these images are shown (see Figure 5).

In template matching approach, the correlation between the reference image (template) and the target image (input) can be calculated by a number of similarity measurements. Table 1, shows the result of locating the face by using ten different correlation measurements, and it demonstrates the increase in accuracy by using the OSAD against the other measurements

Fig 4: Samples from Yale dataset

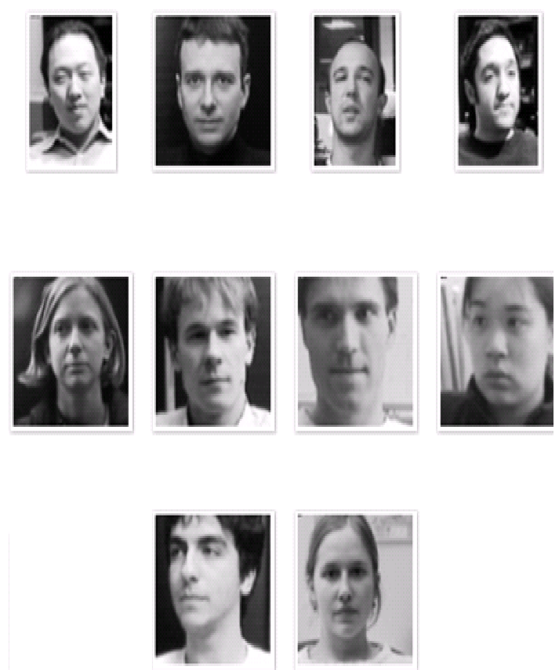

Fig 5: Samples from MIT-CBCL dataset

From the table above, the accuracy of the face localization by using OSAD is 100% and that is due to the selected input image window with small value of OSAD and this window should be the face location whether there is a shade in the image or not (Figure 6) shows example of face localization by OSAD. In addition, the OSAD between the template and shade window will not be smaller due to the high difference between the values of window pixels and the template pixels and it's the same for the window with high light. It means that, the window with the included face will represent the face. For the SAD, the accuracy is 98% which is acceptable in comparison with other measurements. This localization error is due to the shifting of the template image over the input image where the two windows have almost the same SAD or a very small difference but the two windows have the face. This case will produce an error percentage locating the face, but it is only a small error location. As for the SSD, the accuracy is acceptable but its complexity higher than OSAD and SAD and that it maximize the error rate. In addition, if there are two windows with pixels values close to each other SSD is not useful to determine which one is similar to the template. In case of NCC, there is a significant increase in the error rate and that is due to the shade input image. The pixels values of the shade parts are smaller than the parts of the image and this case has a high percentage of NCC between the template image and input image which will determine a wrong location. In contrast of the SDD error, the error in NCC gives a complete wrong location of the face. SHD gives a poor localization rate because it calculates the distance between two strings and not between matrices. Therefore, SHD is not useful for face localization or detection but it can be used to calculate the difference between the signals.

Table II shows the comparison between the proposed technique and the other measurements on the MIT-CBCL database. In this test, the concern on the high poses variation from 30 degree left the 30 degree right and clutter background as well. From the table, Template Matching and OSAD are not affected by the variation in poses but instead are affected by clutter background because of the existence of some objects in the background. Due to that, some windows in input image will have the same or near pixels values to template image, and this problem increases the error rate in general.

Table 1: Comparison between similarity measurements and OSAD on Yale dataset

Similarity Measure	Accuracy (%)
Optimized Sum of Absolute Difference (OSAD)	100
Sum of Absolute Differences (SAD)	98
Zero-mean Sum of Absolute Differences (ZSAD)	98
Locally scaled Sum of Absolute Differences (LSAD)	98
Sum of Squared Differences (SSD)	95
Zero-mean Sum of Squared Differences (ZSSD)	95
Locally scaled Sum of Squared Differences (LSSD)	95
Normalized Cross Correlation (NCC)	80
Zero-mean Normalized Cross Correlation (ZNCC)	80
Sum of Hamming Distances (SHD)	43

Table 2: Comparison between the proposed method and other similarity measurements on MIT-CBCL, the first set contains faces with different poses and the second set contains faces with clutter background

Method	Number of images	Pose	Clutter Background
OSAD	100	100%	96%
SAD	100	100%	94%
ZSAD	100	100%	94%
LSAD	100	100%	94%
SSD	100	100%	89%
ZSSD	100	100%	89%
LSSD	100	100%	89%
NCC	100	95%	73%
ZNCC	100	95%	73%
SHD	100	40%	40%

Fig 6: Example of localization by OSAD.

4. CONCLUSION

In this paper we proposed a fast template matching technique based on OSAD. OSAD proved superiority compared with the other similarity measurements spatially NCC. OSAD is not affected by variation in illumination while it is affected by variation in image background. Due to this problem future work will focus on developing a new similarity measure to locate the faces from clutter background.

5. REFERENCES

[1] R. Chellappa, C. L. Wilson and S. Sirohey, "Human and machine recognition of faces: a survey," *Proceedings of the IEEE,* vol. 83, pp. 705-741, 1995.

[2] Ming-Hsuan Yang, D. J. Kriegman and N. Ahuja, "Detecting faces in images: a survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 24, pp. 34-58, 2002.

[3] Zhao Fei and Qiao Qiang, "Face detection based on rectangular knowledge rule and face structure," in *Information Science and Engineering (ICISE), 2009 1st International Conference on,* 2009, pp. 1235-1239.

[4] S. Jeng, H. Y. M. Liao, C. C. Han, M. Y. Chern and Y. T. Liu, "Facial feature detection using geometrical face model: An efficient approach," *Pattern Recognit,* vol. 31, pp. 273-282, 3, 1998.

[5] A. K. Jain, Y. Zhong and M. Dubuisson-Jolly, "Deformable template models: A review," *Signal Process,* vol. 71, pp. 109-129, 12/15, 1998.

[6] M. -. Yang, N. Ahuja and D. Kriegman, "Face recognition using kernel eigenfaces," in *Image Processing, 2000. Proceedings. 2000 International Conference on,* 2000, pp. 37-40 vol.1.

[7] Lingmin Meng and T. Q. Nguyen, "Frontal face localization using linear discriminant," in *Signals, Systems, and Computers, 1999. Conference Record of the Thirty-Third Asilomar Conference on,* 1999, pp. 745-749 vol.1.

[8] A. Pentland, B. Moghaddam and T. Starner, "View-based and modular eigenspaces for face recognition," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on,* 1994, pp. 84-91.

[9] Lingmin.Meng and Truong.Q.Nguyen, "Frontal face detection using multi-segment and wavelet," in 1999.

[10] D. Tsai and C. Lin, "Fast normalized cross correlation for defect detection," *Pattern Recog. Lett.,* vol. 24, pp. 2625-2631, 11, 2003.

[11] Shou-Der Wei and Shang-Hong Lai, "Fast Template Matching Based on Normalized Cross Correlation With Adaptive Multilevel Winner Update," *Image Processing, IEEE Transactions on,* vol. 17, pp. 2227-2235, 2008.

[12] M. J. Atallah, "Faster image template matching in the sum of the absolute value of differences measure," *Image Processing, IEEE Transactions on,* vol. 10, pp. 659-663, 2001.

[13] "Yale Face Database " vol. 2010,

[14] "CBCL FACE RECOGNITION DATABASE " vol. 2010,