Analyzing Impact of Image Scaling Algorithms on Viola-Jones Face Detection Framework

Himanshu Sharma
Advanced Electronics System
Academy of Scientific and Industrial Research
CSIR-Central Electronics Research Institute, Pilani, India
himanshusharma102@gmail.com

Sumeet Saurav, Sanjay Singh, Anil K Saini, Ravi Saini
IC Design Group
CSIR-Central Electronics Engineering Research Institute
Pilani, Rajasthan, India
sumeetssaurav@gmail.com

***Abstract*— In today's world of automation, real time face detection with high performance is becoming necessary for a wide number of computer vision and image processing applications. Existing software based system for face detection uses the state of the art Viola and Jones face detection framework. This detector makes use of image scaling approach to detect faces of different dimensions and thus, performance of image scalar plays an important role in enhancing the accuracy of this detector. A low quality image scaling algorithm results in loss of features which directly affects the performance of the detector. Therefore, in this paper we have analyzed the effect of different image scaling algorithms existing in literature on the performance of the Viola and Jones face detection framework and have tried to find out the optimal algorithm significant in performance. The algorithms which will be analyzed are: Nearest Neighbor, Bilinear, Bicubic, Extended Linear and Piece-wise Extended Linear. All these algorithms have been integrated with the Viola and Jones face detection code available with OpenCV library and has been tested with different well know databases containing frontal faces.**

Keywords*— *Image Scaling; Viola and Jones face detection framework; Haar cascade classifiers; OpenCV.

I. Introduction

Facial expression is one of the most powerful and immediate means for humans to communicate their emotions, cognitive states, intensions and opinions to each other [1]. However, before going to analyze facial expression of an individual, the first step required is face detection on which further processing is performed. Basically, face detection is a process of finding the location and number of faces present in an arbitrary image or video sequence. Detecting faces in a still image or live video stream is a complex task. A number of factors accounts for the complexity. Some of them are occlusion, lighting conditions, variability in scale, location, orientation and pose [2]. All these factors change the overall appearance of faces present in an image. Therefore, to overcome these limitations a number of algorithms have been developed over the time for face detection each differing from each other either in terms of computational complexity, detection accuracy and robustness against varying lighting condition. However, design of an efficient face detector is still an open challenge among the researchers.

Different techniques available for face detection can be broadly classified into four categories: Knowledge based, feature invariant based, template based and appearance based [2]. Knowledge based methods encode human knowledge of what constituents a typical face. A hierarchical knowledge-based method to detect faces has been reported in [3]. Although, this approach significantly reduced the computation time but failed to achieve a high detection rate. On the other hand feature based approach utilizes structural features to find faces in an image. The major disadvantage of feature based methods is their lack of robustness against illumination, noise, and occlusion. There are a number of variants of feature invariant approach based face detection algorithm have been reported in literature. These are either based on Grouping of edges [4]-[5], Space Gray-Level Dependence matrix (SGLD) of face pattern [6], Mixture of Gaussian [7]-[8], Integration of skin color, size and shape [9] and many more. As the name indicates, in template matching based face detection algorithm, a predefined face pattern is used. Different techniques for face detection developed over the time using template matching approach has been discussed in [2]. Finally, in appearance based methods, templates for faces are learned from examples in images using statistical analysis and machine learning instead of using predefined templates as done in template matching based algorithm. The most well known Viola and Jones face detection algorithm [10], also comes under the category of appearance based face detection algorithm. It is the first face detection framework capable of achieving real-time performance.

II. Viola-Jones Face Detection Framework

Viola and Jones provided the fast and efficient way to detect a face in a given image. It is based on Haar-like features and cascade AdaBoost classifier. It is the first face detection framework which is capable of providing real time performance. Thus, many image processing applications which require faces as their input are build using this algorithm and it is the most commonly used face detection algorithm. This algorithm mainly consists of three basic steps. In the first step, a Haar classifier is trained using thousands of positive and negative images which acts as templates to detect faces. Haar classifier training is followed by cascade classifier training which allows for efficient face detection during runtime. Finally, a window is swept across image and its scaled version to detect faces at different locations and sizes.

In the Viola-Jones face detection framework, a window slides over the entire image from left to right and then from top to bottom. Haar features are calculated within each window using the already trained cascade classifier and decision is taken whether to reject the window or pass it to next stage of the cascade. To detect faces of different sizes there are two possibilities, which are either to scale down the image by some fixed scale keeping the size of the scanning window fixed or to scale up the size of the scanning window itself, keeping the original image as it is. These two methods have been discussed below to a more detail.

1) *Window Scaling Approach*

Window scaling is the easy and fast way to detect the face in an image. In this method the window of size 24x24 slides over the entire image from left to right and then from top to bottom. On completion of one complete scanning the size of this window is increased by certain factor (preferably 1.2) and this process continues till the size of the window become equal to the input image size. As discussed above this approach is used to detect faces of different sizes in an image.

2) *Image Scaling Approach*

The current OpenCV implementation of the Viola and Jones face detection algorithm uses image scaling approach to detect faces of different sizes in an image. However, scaling down an image often results in loss of important details which may ultimately result in loss of features and can affect the detection rate. Thus, in order to ensure minimal loss of features and detection rate, we need to have a good quality image scaler.

III. IMAGE SCALING TECHNIQUES

Image scaling algorithms are widely used in the many fields, ranging from consumer electronics, i.e. mobile phones, display devices. The challenging task in image interpolation is to preserve the visual content of the image while changing its resolution.

For our analysis we have used five most commonly used image scaling algorithms. These are Nearest Neighbor, Bi-Linear, Bi-Cubic, Extended Linear, and Piece-Wise Extended Linear image scaling algorithm. These algorithms have been briefly discussed below.

A. Nearest Neighbor Interpolation

Nearest neighbor is the simplest interpolation from a computational standpoint, where each interpolated output pixel is assigned the value of the nearest sample point in the input image [11].

B. Bi-Linear Interpolation

Bi-linear algorithm is slightly more advanced and complicated than the nearest neighbor algorithms [12]. In Bi-linear interpolation each output image pixel is computed by two linear interpolations done on 2x2 image window in order to find out the target pixel location. It reduces the blocking and blurring effect which is common in nearest neighbor, but these effects are still visible in interpolated images.

C. Bi-Cubic Interpolation

Bi-cubic image interpolation is little more complex than nearest neighbor and Bi-linear interpolation [13]. In Bi-cubic interpolation the target pixel is computed by 2-D interpolation using 4x4 neighboring pixels.

D. Extended Linear Interpolation

Extended Linear interpolation is almost same as Bicubic interpolation algorithm in term of quality of resized image [14], but it significantly reduces the computational overhead compared to Bicubic. Extended linear use 16 weighted coefficients which are generated from 16 neighboring pixels input image to calculate the value of the target pixel.

E. Piece-Wise Extended Linear Interpolation

Piece-Wise Extended Linear Interpolation is the modified version of extended linear interpolation method discussed above [15]. It uses 4 piecewise linear function to calculate the value of the target pixel. Piece-wise extended linear is computationally less complex and fast in processing in comparison of extended linear (8 multiplication, 12 addition) and Bi-cubic interpolation (20 multiplication and 20 addition).

IV. DATABASE PREPARATION

In order to evaluate the performance of different image scaling algorithms on Viola and Jones face detection framework, we have used four different face databases available in the literature. A brief discussion of different databases used in our experiments has been given below.

A. Adience Database

The Adience database [16], consists of 26,580 images, portraying 2,284 individuals, classified for 8 age groups, gender and including subject labels (identity). For our experimentation purpose we selected 1000 random frontal face images which have been resized to 250x250 using irfanView software.

B. Labeled Faces in the Wild

The Labeld Faces in the Wild (LFW) [17], database contains more than 13,000 images of faces collected from the web. Here we have also selected 1000 random frontal face images which have been resized to 250 x 250 using irfanView software.

C. Cohn-Kanade AU-Coded Expression Database

The well known Cohn-Kanade database consists of different facial images of 100 university students [18]. For our experiment purpose we have selected 1000 random frontal face images from this database.

D. JAFFE Database

This database contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models [19]. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects. For the experiment purpose we selected 213 image from the JAFFE database.

E. Indian Face Database (IFD-CEERI)

The Indian Face Database – CEERI (IFD-CEERI) has been constructed at the Perception and Cognition Engineering Lab in the IC Design Group of CEERI-Pilani. The size of each image is of 720 x576. For our experiments we have randomly chosen 1000 frontal face images from the database.

V. RESULTS AND DISCUSSION

All image scaling algorithms has been coded using C++ and integrated with OpenCV using Visual Studio 2010, IDE. In order to analyze the accuracy of different image scaling algorithms we have used Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) between the original and rescaled images. The PSNR and SSIM values found are the mean of the results from five standard test images. These values were generated by resizing each source image up/down (from original size of 512x512) to size of 256x256, 300x300, 400x400, 500x500, 600x600 and 700x700, after which each image was scaled back down/up to size 512x512 using algorithm being tested. The average value of the PSNR and SSIM calculated using all the above image manipulation steps are used as the final PSNR and SSIM value of the image scaling algorithm. Different test images used in our experiments has been shown in the fig. 3. The average PSNR and SSIM value of five test images shown in fig. 3, using different scaling algorithms has been shown in table II and table III respectively.

TABLE I. PSNR COMPARISION OF DIFFERENT IMAGE SCALING ALGORITHMS

Algorithm	Average PSNR (dB)
Nearest Neighbor	28.8364
Bi-linear	33.9458
Bi- Cubic	35.4615
Extended Linear	34.6496
Piece-Wise Extended Linear	35.4015

TABLE II. SSIM COMPARISION OF DIFFERENT IMAGE SCALING ALGORITHMS

Algorithm	Average SSIM
Nearest Neighbor	0.9325
Bi-Linear	0.9735
Bi- Cubic	0.9787
Extended Linear	0.9776
Piece-Wise Extended Linear	0.9788

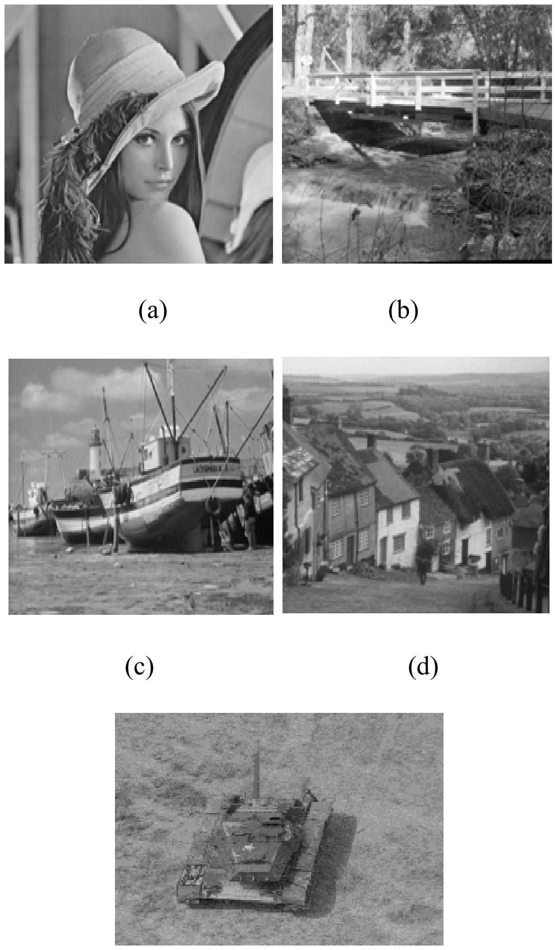

Fig. 3. Test Images used in our experiments: (a) Lena (b) Bridge (c) Boat (d) Gold Hill and (e) Tank.

To analyze the impact of image scaling and window scaling algorithms on face detection we have used three performance metrics: True Positive (TP), False Positive (FP) and False Negative (FN). The performance of different image scaling algorithms on the four databases has been shown in the table IV to table VIII and that of the window scaling has been shown in table IX.

TABLE III. PERFORMANCE METRICS OF NEAREST NEIGHBOR

Database	Nearest Neighbor		
	True Positive	*False Positive*	*False Negative*
Adience	980	20	1
LFW-a	985	15	3
Cohn Kanade	1000	0	0
JAFFE	213	0	0
CEERI-Dataset	966	39	6

TABLE IV. PERFORMANCE METRICS OF BILINEAR

Database	Bilinear		
	True Positive	*False Positive*	*False Negative*
Adience	983	17	1
LFW-a	986	14	3
Cohn Kanade	1000	0	5
JAFFE	213	0	0
CEERI-Dataset	977	28	8

TABLE V. PERFORMANCE METRICS OF BICUBIC

Database	Bicubic		
	True Positive	*False Positive*	*False Negative*
Adience	982	18	2
LFW-a	989	11	1
Cohn Kanade	1000	0	3
JAFFE	213	0	0
CEERI-Dataset	986	19	15

TABLE VI. PERFORMANCE METRICS OF EXTENDED LINEAR

Database	Extended Linear		
	True Positive	*False Positive*	*False Negative*
Adience	977	23	2
LFW-a	986	14	2
Cohn Kanade	1000	0	1
JAFFE	213	0	0
CEERI-Dataset	970	35	7

TABLE VII. PERFORMANCE METRICS OF PIECE-WISE EXTENDED LINEAR

Database	Piece-Wise Extended Linear		
	True Positive	*False Positive*	*False Negative*
Adience	978	22	2
LFW-a	987	13	2
Cohn Kanade	1000	0	1
JAFFE	213	0	0
CEERI-Dataset	988	17	20

TABLE VIII. PERFORMANCE METRICS OF WINDOW SCALING APPROACH

Database	Window Scaling		
	True Positive	*False Positive*	*False Negative*
Adience	990	10	3
LFW-a	993	7	13
Cohn Kanade	1000	0	11
JAFFE	213	0	1
CEERI-Dataset	976	24	60

VI. CONCLUSION

In this paper, we investigated the impact of different images scaling techniques on the Viola and Jones face detection framework. The quality of scaled images using all the image scaling techniques has been analyzed using their SSIM and PSNR values. From our analysis result we found that the window scaling based method performed relatively better compared to the image scaling techniques. Moreover, in the image scaling based approach as expected Bicubic algorithm performed well, which is followed by Piece-wise extended linear, Bilinear, Extended linear and Nearest neighbor. For our own database, we found that Piece-wise extended linear image scaling technique gave the best result compared to all other algorithms.

ACKNOWLEDGMENT

The authors express their deep sense of gratitude to Dr. Chandra Shekhar, Director CSIR-CEERI for encouraging research and development activities. The authors, would like to thanks Dr. A.S Mandal, Chief-Scientist and Prof. Raj Singh, Chief Scientist and Group Leader, IC Design Group, CSIR-CEERI, for their constructive suggestions. The authors also like to thanks Mr. Sanjeev Kumar, Technical Officer, CSIR-CEERI, for providing technical support. The financial support of DeitY/MCIT is also gratefully acknowledged.

REFERENCES

[1] P. Ekman, Emotion in the Human Face, Cambridge University Press, New York, 2nd edition, 1982.

[2] Yang, Ming-Hsuan, D.J. Kriegman, and N. Ahuja,"Detecting faces in images: A survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, 2002.

[3] Yang, Guangzheng, and T. S. Huang, "Human Face Detection in Complex Background," Pattern Recognition, vol. 27, no. 1, pp. 53-63, 1994.

[4] T.K. Leung, M.C. Burl, and P. Perona, "Finding faces in cluttered scenes using random labeled graph matching," in Proceedings of the Fifth IEEE International Conference on Computer Vision, pp. 637-644, 1995.

[5] K.C. Yow and R. Cipolla, "Feature-Based Human Face Detection," Image and Vision Computing, vol. 15, no. 9, pp. 713-735, 1997.

[6] Y. Dai and Y. Nakano, "Face-Texture Model Based on SGLD and its application in face detection in a color scene," Pattern Recognition, vol. 29, no. 6, pp. 1007-1017, 1996.

[7] J. Yang and A. Waibel, "A real-time face tracker," in Proceedings of the Third IEEE Workshop on Applications of computer Vision, pp. 142-147, 1996.

[8] S. McKenna, S. Gong, and Y. Raja, "Modelling Facial Colour and Identity with Gaussian Mixtures," Pattern Recognition, vol. 31, no. 12, pp. 1883-1892, 1998.

[9] R. Kjeldsen and J. Kender, "Finding Skin in Color Images," in Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, pp. 312- 317, 1996.

[10] Paul Viola and Michel J. Jones," Robust Real-Time Face Detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, 2004.

[11] J. A. Parker, R. V. Kenyon, and D.E. Troxel," Comparison of interpolation methods for image resampling," IEEE Transaction on Medical Imaging, vol. MI-2, no. 1, 1983.

[12] T. M. Lehamann, C. Gonner and K. Spitzer," Survey: interpolation methods in medical image processing," IEEE transaction on medical imaging, pp 1049- 1075, 1999.

[13] H. S. Hou and H. C. Andrews, "Cubic splines for image interpolation and digital filtering," IEEE Transaction on Acoustic, Speech, Signal Processing, vol. ASSP-26, pp. 508–517, 1978.

[14] C. Lin, M. Sheu, H. Chiang, W. Tsai, and Z. Wu, "Real-Time FPGA architecture of extended linear convolution for digital image scaling," in Proceedings of the IEEE International Conference on Field-Programmable Technology, pp. 381–384, 2008.

[15] P. N. Gour, S. Narumanchi, S. Saurav, and S. Singh,"Hardware accelerator for real- time image resizing," in Proceedings of the 18th IEEE International Symposium on VLSI Design and Test, pp 1-6, 2014.

[16] E. Eidinger, R. Enbar, and T. Hassner,"Age and gender estimation of unfiltered faces," IEEE Transactions on Information Forensics and Security, vol. 9, no. 12, pp. 2170-2179, 2013.

[17] G.B Huang,M. Ramesh,T. Berg, et .al, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments,", Technical Report 07-49, University of Massachusetts, vol. 1, no. 2, Amherst, 2007.

[18] T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 46-53, 2000.

[19] L. Michael, S. Akamatsu, M. Kamachi, et al," Coding facial expressions with gabor wavelets," in Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205, 1998.