A Comparison of Face Detection Algorithms in Visible and Thermal Spectrums

Kristopher Reese[1], Yufeng Zheng[2], Adel Elmaghraby[1]
[1] University of Louisville, Louisville, KY USA
Email: {kwrees02, adel}@louisville.edu
[2] Alcorn State University, Alcorn State, MS USA
Email: zheng@alcorn.edu

Abstract – **Face Detection is the first step of facial recognition algorithms and has been widely researched in the visible spectrum. Current research has shown that thermal facial recognition is as accurate as the visible spectrum recognition algorithms. This paper presents three face detection algorithms in both long-wavelength infrared (LWIR) images and visible spectrum images. The paper compares the Viola-Jones algorithm, Gabor feature extraction and classification using support vector machines, and a Projection Profile Analysis algorithm. The Gabor feature extraction method can detect faces in both spectrums with separate training, but the algorithm is extremely slow. The Project Profile Analysis method can find faces in LWIR images, but is not applicable to visible spectrum images. Our experimental results show that the Viola-Jones algorithm is the most reliable and efficient solution for the implementation of a real-time face detection system using either visible or thermal spectrum images.**

Index Terms – **Face Detection, Object Detection, Thermal Imaging, Image Processing**

I. Introduction

Facial recognition is a biometric technology that could be used for authentication, validation, or in various other security applications. Outside of the use of fingerprints, facial recognition and voice recognition are the most commonly used methods of identification. [1] With security becoming tighter in many parts of the world, the idea of creating a biometric system that can capture information without users being bothered by interactions has become extremely appealing.

Though many other biometric technologies exist and are currently more accurate in comparison to facial recognition [2], the methods require the users to interact with sensing devices. In comparison, a facial recognition algorithm could be run from a security camera, many of which are placed in high security areas already. This would allow people to go about their daily business without having to be bothered by an interaction with other biometric devices.

One of the issues that arises with the use of security cameras is proper lighting conditions. [3] Though some work has shown that the use of Self Quotient Images provides a means for some varying light images to allow facial recognition [4], this does not account for dark rooms or very low light conditions. If camera is not set in a properly lighted room, the camera ceases to be useful. Another solution to the lighting problem has been to use Long-Wavelength Infrared (LWIR) [5] to view the thermal infrared emissions of a body or object.

As has been discussed in other papers on face recognition, there are several steps that have to be completed to have a chance of success. 1) We have to detect the face in an image or frame captured from the imaging device. 2) We have to properly align the face so that we have a chance of success in finding similar faces in the database. 3) clear obstructions, such as glasses in the thermal spectrum, from the image which may cause the recognition algorithm to fail, and finally 4) run the facial recognition algorithm that was chosen. [5]

With the use of LWIR for facial recognition in [6], the feasibility of using current techniques has to be checked. For that reason, this paper focuses on the first step, face detection. Significant limitations were placed on the images to control the experiments. It is assumed that faces are full frontal faces, similar to those that would be found on a passport photo - though distance or resolution are not a factor of concern. It is also assumed that the glasses are allowed in the images. The collected database will be discussed further in section 3.

The Projection Profile Analysis algorithm that was designed for use with thermal facial detection, and was initially proposed for use in thermal face detection in [5]; however there are significant modifications that will be discussed further in section 2. We also discuss the two other algorithms that were tested, the Viola-Jones algorithm [7][8] and Gabor feature extraction [9][10] using Support Vector Machines for classification of images in section 2. The way in which the experiment was designed, and the methods for capturing the Alcorn State University database, is discussed in section 3. A discussion of the results of the experiment can be found in section 4, and the paper concludes in section 5.

II. Algorithms

For this experiment, three algorithms were chosen to run against a number of databases containing visible spectrum images and long-wavelength infrared images (LWIR) of a number of subjects. The databases used for experimentation include the University of Notre Dame database, containing LWIR images and Visible images; the FERET database containing visible images; and the Alcorn State University database, containing LWIR images and visible images. The three algorithms tested were: The Viola-Jones algorithm, Gabor feature extraction and classification using Support

DOI: 02.CSA.2012.01.20

Vector Machines, and a modified version of the algorithm in [5], the Projection Profile Analysis method.

A. Viola-Jones method

The Viola-Jones algorithm has become a very common method of object detection, including face detection. Viola and Jones proposed this algorithm as a machine learning approach for object detection with an emphasis on obtaining results rapidly and with high detection rates. This paper attempts to show its feasibility within the LWIR and visible spectrums of images, though the algorithm has been tested with visible images many times.

The Viola-Jones method of Object Detection uses three important aspects. The first is an image representation structure called integral images by Viola and Jones [7][8]. In this method, features are calculated by taking the sum of pixels within multiple rectangular areas. This is of course an extension of a method by Papageorgiou et al. [11], though this extension proves to be significantly more complex.

The rectangles, shown in figure 1 show the four rectangles that were initially used by Viola and Jones. Leinhart and Maydt later extended the Viola-Jones algorithm to allow for rectangle rotations of up to 45 degrees [12]. Using these methods, the sum of both the white and the shaded regions of each rectangle are calculated independently. The sum of the shaded region is then subtracted from the white region. Further discussion on the mathematics behind this method can be found in [7][8][12].

Viola and Jones admit that the use of rectangles is rather primitive, but they allow for high computational efficiency in calculations [7][8]. This also lends well to the Adaboost algorithm that is used to learn features in the image. This extension of the Adaboost algorithm allows for the system to select a set of features and train the classifiers, a method first discussed by Freund and Schapire [13]. This learning algorithm allows the system to learn the differences between the integral representations of faces to those of the background.

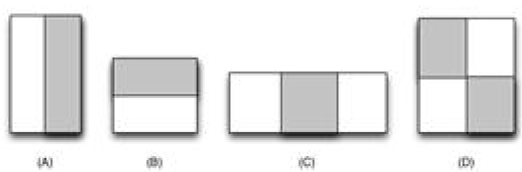

Figure 1. This figure shows four rectangles initially used by Viola and Jones [7][8] which are used to represent images for the learning algorithm. both the light and shaded region are added up separately. The sum of the light is then subtracted from the sum of the shaded regions

The last contribution of the Viola-Jones method is the use of cascading classifiers. In order to significantly speed the system up, Viola and Jones avoid areas that are highly unlikely to contain the object. The image is tested with the first cascade. and the areas that do not conform to the first cascade are rejected and no longer processed. The areas that may contain the object are further processed until all of the classifiers are tested. The areas in the last classifier are likely to contain the object.

B. Gabor Feature Extraction with SVMs

Another method of face detection is to use Gabor filters to extract features of an image. These can then be applied to a neural network (ANN) or support vector machine (SVM) for image classification. Sahoolizadeh et al. [9] and Gupta et al. [10] have used this method in face detection in the visible spectrum. In this method, an image is input into the system. A Gabor wavelet transformation is then run against the image allowing the system to find important feature points on the image. These are then added to the feature vector that is passed to the SVM or ANN, which has been previously trained.

Once these features are extracted, the system uses one of the learning methods (SVM or ANN) to check for faces around the points based on a threshold of the surrounding pixels. After a dilation of the checked regions, the center of the points are found and marked as faces in the image. More information about the mathematics and the system can be found in [9][10]. This discussion of the topic further is beyond the purpose of this paper.

C. Projection Profile Analysis

This last method was designed by the authors in an attempt to solve the problem of finding faces in the thermal spectrum quickly. It is capable of finding faces where only one person exists in the image and using the thermal spectrum. Zheng was the first to use the projection profile analysis algorithm for thermal face detection [5][6].

This method however has been frequently used in the medical imaging community in radiography and magnetic resonance imaging. This algorithm has been used for detecting different views in radiographs [14] and in detecting abnormalities in chest radiographs [15].

The initial algorithm used by Zheng [5][6] uses a region growing segmentation to separate areas of interest from background noise. This region growing segmentation was slow and lowered the accuracy rate slightly. In place of region growing segmentation, histogram threshold segmentation was used for this experiment. This sped up the algorithm by almost 98%, increasing a test of 1,015 images from region growing segmentation time of 13 hours to a histogram segmentation time of 0.15 hours (about 10 minutes). There was also a surprising increase in accuracy of about 13 percent.

This algorithm first works by segmenting the image into 2 segments – background and regions of interest. The background of the thermal database is assuming no thermal noise, and thereby black. Only the heat of the person's body results in becoming a part of the segment of interest. Any pixel that is a part of the segment of interest is marked with a mask of one, as shown in figure 2B. Upon completion of the histogram threshold segmentation, a morphological reconstruction is run against the image to fill any holes in the segments.

DOI: 02.CSA.2012.01. 20

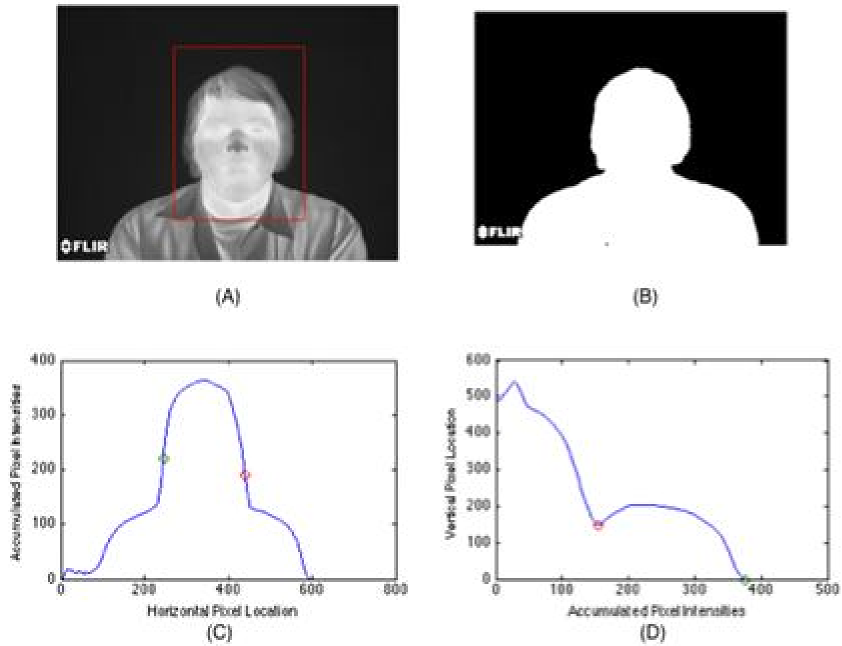

Figure 2. This figure shows some of the steps that were used in detecting the face. (A) shows the original image. It also shows the results of the face detection inside of the box. (B) shows the pixel mask used to calculate the sum of interesting pixels. (C) is the measure of the horizontal pixels used to located the head as opposed to the shoulders or body. (D) shows the vertical pixels used to locate the head as opposed to the shoulders of body

The rows and columns of this mask are then used to determine how many pixels of interest each row or column contains. These summations can then be used to determine the location of the head based on expected changes. Where the head meets the shoulders, a sudden change occurs in vertical pixel intensities. We can take the derivative of this change point and the highest point of the head to measure the height of the head, see figure 2C. Based on the horizontal pixel intensities, see figure 2D, we can measure the distances between the left shoulder change and the right shoulder change to give us the width of the head

Once the height and width are found a box can be drawn around the head. by using the highest point and the leftmost point as a top-left point of the box and drawing the rectangle accordingly. Experimental results show that the Projection Profile Analysis provides a method for capable of determining the location of the faces in the thermal databases tested.

III. Design of experiment

For this experiment, each of the algorithms were tested on both visible and thermal (LWIR) images that are found in the Alcorn State University (ASU) database, the University of Notre Dame (UND) database, and the FERET database. The results are tabulated and compared by determining the percentage of success by going through each image and determining whether an image successfully finds a person's full face. Examples of successful detections can be found in figure 3. Any image that does not closely conform to the examples were included as failures.

The ASU database was captured on the ASU campus and contains images of volunteer students. Each student had images captured by two FLIR SC620 cameras placed side-by-side to capture stereoscopic images. In this experiment, the stereoscopic properties of the database are ignored. Both cameras captured thermal and visible spectrum images at the same time, comprising of 42 total images containing frontal, left and right angled (45 degrees), and left and right profile faces. Glasses were permitted for anyone who wears glasses on the last few images that were captured. Only the frontal faces of the database were included in this experiment.

For this experiment, little preprocessing was done to the images before running the database against the algorithm. The preprocessing included image de-noising using Gaussian filters and adjusting the contrast in the case of the profile projection analysis algorithm. All of the algorithms were implemented using MATLAB to avoid potential influence between differing programming languages.

As the Viola-Jones and the Gabor Feature extraction algorithms require training, a subset of the databases containing images from databases was used to train the algorithm. These images were picked at random from the entire database for training. There were a total of two training sessions. The first session was trained for the LWIR images, containing images from the UND and ASU LWIR databases; the second was for visible spectrum images, containing images from the UND, ASU, and FERET visible spectrum databases. The training set of the algorithms consisted of a subset of 25% of the images from each database chosen at random. This means that approximately 200 of the images from the ASU database, 550 of the UND database, and 757 FERET images were used in training the algorithms. Two independent classifiers were created for Visible and LWIR images.

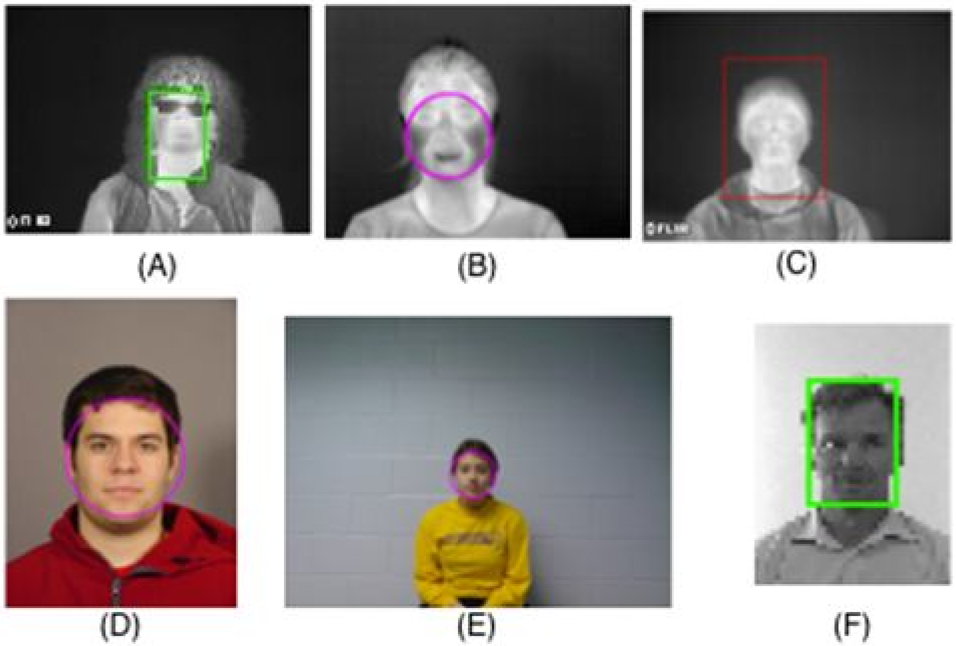

Figure 3. Standard images from each of the tested databases. (A) ASU LWIR database [640 x 480 pixels]. (B) UND LWIR database [312 x 239 pixels]. The purple ellipsoid shows a successful Viola-Jones detection. (C) Another image from the ASU LWIR database. (D) UND Visible Spectrum database [1200 x 1600 pixels]. (E) ASU Visible Spectrum database [2048 x 1536 pixels]. (F) FERET Visible Spectrum gray scale database [256 x 384 pixels]. The thick, green rectangle in (A, F) shows successful Gabor feature extraction detection in LWIR (A) and visible (F) spectrum images. The thin, red line in (C, figure 2A) shows successful Projection Profile Analysis detections. The purple ellipsoid in (B, D, E) show successful Viola-Jones detections in LWIR (B) and visible (D,E) spectrum images

The remaining images were used for testing the algorithms. The training data was used for both the Gabor feature extraction SVM training and the Viola-Jones Boosted classifier training. Standard images for each of the databases can be found in figure 3. Only frontal faces were used in each of the databases. As mentioned, any non-frontal facial positions – including angular or profile – were omitted from the database.

IV. Experimental results

The experiment shows that all three of the methods were capable of determining faces in the LWIR images, and both the Viola-Jones method and the Gabor Feature Extraction were able to find images in the Visible Spectrum, with separate training for both spectrums. Table 1 shows the results of the experiment for each of the implemented algorithms, including the average time to complete a single image in the database and the accuracy against the tested database.

A comparison of the time should be taken cautiously. As mentioned in figure 3, the images have different aspect ratios, which causes the slowdown between different databases. The smaller the image, the faster the algorithms run. For this experiment, no image resizing was done, and therefore the time can be misleading. Based on the aspect ratios of the pictures, the times are fairly constant between databases. One should only compare the time for each algorithm with the same spectrum and database.

The accuracy of the Viola-Jones algorithm remains fairly consistent between each of the databases, approaching 90% for both the Univ. of Notre Dame visible spectrum database and the FERET visible spectrum images. These images were taken under ideal conditions and there is very little background noise that might cause failures in the image. The IR images do show a slight degradation in accuracy from the visible spectrum. The noise in the UND LWIR database is likely the cause of significant degradation in that database. There is only a slight degradation between the ASU LWIR and ASU Visible Spectrum images, which might be accounted for by using a larger training set in future experiments.

The Gabor feature extraction and expansion using SVMs is the slowest of each of the algorithms. The time it takes the algorithm to complete makes the algorithm unable to be implemented in a real-time security system. The accuracy of the algorithm is also no better than those shown in the Viola-Jones or Projection Profile Analysis algorithms. Though this is certainly a means of finding a face in an image, it is perhaps the least ideal of the algorithms for implementation in either LWIR or visible spectrum images.

The modified Projection Profile Analysis algorithm shows the least successes of all of the algorithms. Though the algorithm is fast, it is incapable of finding faces in the visible spectrum. This is due to the intensity of the background and face colors being the same. The algorithm was designed for LWIR, where the face would be significantly brighter than the background. We also see inconsistencies between the two LWIR databases.

The UND LWIR database barely reaches an accuracy of 50%. This is likely due to the noise in the image and the varied light intensities of the background. These variations in the lighting of the background cause the segmentation to misinterpret portions of the background into the face segment. Though the algorithm is capable of reaching the highest

DOI: 02.CSA.2012.01.20

TABLE I. PERFORMANCE COMPARISONS OF FACE DETECTION ALGORITHMS

Algorithm	Method of Measure	ASU		UND		FERET
		IR	VS	IR	VS	VS
Viola-Jones	Time (sec / img)	0.05	4.99	0.01	2.57	0.15
	Accuracy	78.99%	82.27%	70.62%	89.98%	88.91%
Gabor Features w/ SVM	Time (sec / img)	28.27	103.72	2.06	37.77	5.94
	Accuracy	77.64%	74.91%	76.33%	70.03%	86.13%
Projection Profile Analysis	Time (sec / img)	1.40	NA	2.19	NA	NA
	Accuracy	83.84%	NA	52.92%	NA	NA

accuracy of the algorithms for the ASU LWIR, this inconsistency makes the algorithm very unreliable even in the LWIR spectrum.

Overall, the times seem to make little difference between the Viola-Jones and Projection Profile Analysis algorithms. The Gabor feature extraction and image classification using SVMs show significantly slower times in comparison to the other algorithms against the same database. The consistency of the accuracy in the Viola-Jones algorithms makes it more ideal for a face detection system.

V. CONCLUSIONS

Based on the results of the experiment, the Projection Profile Analysis algorithm is the least reliable of the algorithms to find faces in LWIR images. It's inability to find faces in the Visible Spectrum as well make it less than ideal for implementation.

The experiments have also shown that Learning based methods for face detection are capable of working in both LWIR and Visible Spectrum images. The Viola-Jones algorithm is perhaps the best of the algorithms for implementation in a face detection or facial recognition system using the thermal LWIR spectrum.

The accuracy rates leave much to be desired. It would be ideal to find a way to improve the accuracy rate of the Viola-Jones algorithm for LWIR images. A larger training size should improve the algorithm to a small degree, however improvements to the algorithm should be considered, or implementation a tree of features might improve the accuracy. This improvement would be the next process in developing a system that could work as a security system in high-risk areas.

ACKNOWLEDGMENTS

This research was supported by the Department of Homeland Security sponsored Southeast Region Research Initiative (SERRI) at the Department of Energy's Oak Ridge National Laboratory.

REFERENCES

[1] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfield, "Face Recognition: A Literature Survey." ACM Computing Surveys, vol. 35, no. 4, pp. 399–458, 2003

[2] K. Delac and M. Grgic, "A Survey of Biometric Recognition Methods." International Symposium Electronics in Marine, pp. 184–193, 2004

[3] Y. Adini, Y. Moses, S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction." IEEE Trans. Patt. Anal. Mach. Intell. vol. 19, pp. 731-732, 1997

[4] H. Wang, S. Yansheng Wang, "Face recognition under varying lighting conditions using self quotient image." Proceedings of 6th Automatic Face and Gesture Recognition , pp. 819–824, 2004

[5] Y. Zheng, "Face Detection and eyeglasses detection for Thermal Face Recognition" Proceedings of SPIE 8300, 83000C, 2012

[6] Y. Zheng, "A novel thermal face recognition approach using face pattern words." Proceedings from SPIE, Vol. 7667, 766703, 2010

[7] P. Viola and M. Jones, "Rapid Object detection using a boosted cascade of simple features." Proceedings of CVPR, vol. 1 pp. 511–518, 2001

[8] P. Viola and M. Jones, "Robust Real-time Object Detection." International Journal of Computer Vision, Vol. 57, Iss. 2, pp. 137–154, 2001

[9] H. Sahoolizadeh, D. Sarikhanimoghadam, and H. Deghani, "Face Detection using Gabor Wavelets and Neural Networks." World Academy of Science, Engineering and Technology, vol. 45, pp. 552-554, 2008

[10] B. Gupta, S. Gupta, and A. Tiwari, "Face Detection using Gabor Feature Extraction and Artificial Neural Networks." Proceedings from ISCET pp. 18-23, 2010

[11] C. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection. "Sixth International Conference on Computer Vision, pp. 555-562, 1998

[12] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection." Proceedings of the International Conference on Image Processing, vol. 1, pp. 900-903, 2002

[13] Y. Freund and R. Schapire, "A decision theoretic generalization of on-line learning and an application to boosting." Computational Learning Theory: Eurocolt '95, pp 23-37, 1995

[14] E.Kao, C. Lee, T. Jaw,J. Hsu, G. Liu, "Projection Profile Analysis for Identifying different views of chest radiographs". Acad. Rediol. vol. 13, iss. 4, pp. 518-525, 2006

[15] E.Kao, C. Lee, J. Hsu, T. Jaw, G. Liu, "Projection profile analysis for automated detection of abnormalities in chest radiographs", Med. Phys. vol. 33, iss. 1, pp. 118-123, 2006

DOI: 02.CSA.2012.01.20