THE FORTH INTERNATIONAL CONFERENCE ON COMMUNICATIONS AND ELECTRONICS (ICCE 2012)

1st – 3rd August 2012, Hue Royal City – Vietnam

WELCOME

Welcome to the Fourth International Conference on Communications and Electronics on USB. ICCE is becoming a reputable biennial international conference series in the scientific community on the areas of Electronics and Communications recently. Following the past successful conferences, the fourth ICCE (ICCE 2012) looks for significant contributions to various topics in communication engineering, networking, microwave engineering, signal processing and electronics engineering. The conference will also include tutorials, workshops, and technology panels given by world-class speakers.

This edition is designed so that you may locate papers by session or authors. Papers are originated as electronic files and were converted to Adobe Acrobat PDF file format for crossplatform access. Even though the viewing quality on your monitor may vary, all papers print clearly.

Thank you and Enjoy!

IEEE Catalog Number: CFP1216B-CDR
ISBN: 978-1-4673-2491-5

Organized by

Hanoi University of Science and Technology

UNIVERSITY OF TECHNOLOGY SYDNEY

Co-organized by

VIỆN ĐIỆN TỬ - VIỄN THÔNG
School of Electronics and Telecommunications

TRƯỜNG ĐẠI HỌC BÁCH KHOA
Danang University of Technology

Technically sponsored by

IEEE

KICS
Korea Information and Communications Society

REV

IEEE COMMUNICATION SOCIETY Vietnam Chapter

IEEE photonics SOCIETY
Transforming science into technology

IEEE COMMUNICATIONS SOCIETY Japan Council Chapter

ECO net

Trend

Co-operatively sponsored by

EiC

EiC Communications Society

TABLE OF CONTENTS

PART A: FULL REGULAR PAPERS

PART B: POSTER PAPERS

A method for hand detection based on Internal Haar-like features and Cascaded AdaBoost Classifier

Van-Toi NGUYEN[1,2] Thi-Lan LE[1] Thi-Thanh-Hai TRAN[1] Rémy MULLOT[2] Vincent COURBOULAY[2]

[1]International Research Institute MICA
HUST - CNRS/UMI-2954 - GRENOBLE INP
HANOI UNIVERSITY of SCIENCE and TECHNOLOGY
VIET NAM

[2]L3i Laboratory
LA ROCHELLE UNIVERSITY
FRANCE

Abstract—**Hand detection is the first step in almost hand posture recognition systems. This paper presents a hand detection method based on Viola-Jones detector [1]. The main contribution of this work is a new approach for hand detection that detects internal region of the hand without background based on local features of this internal region. We call the set of these features as Internal Features. In case of Haar-like features, we call them as Internal Haar-like features. We also propose a framework for hand detection that combines several individual hand posture detectors. Experimental result shows that the proposed method outperforms the conventional method based on Viola-Jones detector with the same computation time. The proposed method is reliable for hand detection in the hand posture recognition system.**

Keywords- Hand detection; Hand posture recognition; Haar-like features; Cascaded AdaBoost Classifier.

I. INTRODUCTION

In recent years, hand gesture recognition has become a very active research topic because of its potential use in human–computer interaction, image/video coding, and content-based image/video retrieval, camera surveillance, just name a few. Figure 1 shows one of the general framework of hand gesture recognition.

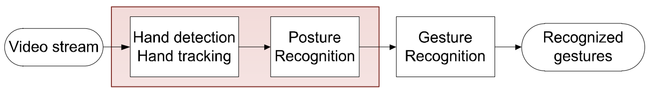


Figure 1. The frame work of hand gesture recognition.

The first step of this framework is hand detection. Accurate detection of hands in still images or video still remains a challenging problem due to the variability of hand appearance. Many previous methods for hand detection based on skin color [2][3][4][5]. In the complex background, there are many regions having the similar color of skin color. In addition, these methods often have to do some another process to decide which region contains the hand. The methods based on skin color also depend on lighting condition. Some other methods use the existing detectors such as Viola-Jones to detect hand [6][7][8][9]. These approaches have a problem with complex and variant background that is the confusion between the object and the background. For example in the Viola-Jones detector, the leaner records characteristics of the hand and it also records characteristics of the background. In this paper, we propose a hand detection method based on Viola-Jones detector that has properties of Viola-Jones detector but avoids the bad effect of the background.

The contributions of our work are:

- Firstly, we propose a frame work of hand detection in the hand postures recognition.
- Secondly, we introduce two new concepts that are Internal Features and Internal Haar-likes features.
- Finally, in overall also, we propose a method for hand detection based on Internal Haar-likes features and Cascaded AdaBoost Classifier.

The remaining of this paper is organized as follow. In section II, we present the related works. In section III, we introduce our hand detection method. The experimental results are analyzed in section IV. Section V gives our conclusion and future work.

II. RELATED WORKS

Several approaches have been proposed for hand detection in recent years. There are two main categories [10]. Approaches in the first category (named appearance based approach) rely on the appearance of the hand itself. Approaches in the second category utilize the context supplied by surrounding parts.

In the first category, the appearance based methods exploit features from only the hands, without using information from other body parts. Zhu et al. [3] generate a hand color model and a background color model for a given image, and then use these models to classify each pixel in the image into hand pixel and background pixel. Ong et al. [11] present a novel, unsupervised approach to train an efficient and robust detector which not only detects the presence of human hands within an image but also classifies the hand shape. In their paper, a tree structure of boosted cascades is constructed. The head of the tree provides a general hand detector while the individual branches of the tree classify a valid shape as belong to one of the predetermined clusters exemplified by an indicative hand shape. Kolsch et al. [9] presented a view-specific hand posture detection with an object recognition method recently proposed by Viola and Jones. Tran et al. [6] proposed a hand posture classification method consisting of 2 steps. The first step aims at detecting skin regions using a very fast algorithm of color segmentation based on thresholding technique. This segmentation is robust to lighting condition thank to a step of color normalization using neural network. In the second step, each skin region is classified into one of hand posture classes using Cascaded Adaboost technique. Yuan et al. [4] proposed to use motion residue as a feature for hand detection, and use dynamic programming to identify an optimal sequence of hand locations in videos.

In second category, the methods represent the shape of hands as a deformable shape, and uses shape matching algorithms to find hands. Athitsos et al. [12] proposed Hidden State Shape Models to represent shapes of variable structure. In this method, hand shapes are modeled as shapes of variable structure. The detection algorithm aims at finding instances of such shapes in images with large amounts of clutter by finding globally optimal correspondences between image features and shape models. Thayananthan et al. [13] compare two methods for object localization from contours: shape context [14] and chamfer matching of templates [15]. Coughlan et al. [16] presented a novel deformable template which detects the boundary of an open hand in a grayscale image without initialization by the user. Zhang et al. [10] evaluated four features for hand detection: color, temporal motion, gradient norm, and motion residue, and they explored the potential of these features for building a reliable hand detector. At first, they used these four features separately to identify where the hands are in each frame of their gesture videos. Then they evaluated different combinations of such features using weighted linear combination, so to build a more accurate hand detector.

III. PROPOSED APPROACH

As analyzed in the section II, the methods for hand detection based on skin color depend on lighting condition and skin color. The object detection method of Viola-Jones has been proved that it can work well on the face detection because it can also avoid the bad effect of lighting and skin color variation [1]. Therefore, we can apply this method into hand detection or posture classification. All of the previous works that apply Viola-Jones detector into hand detection or hand posture recognition use training positive samples which are approximate circumscribed rectangle's areas of the hand [6][8][9]. An ***approximate circumscribed rectangle's area of the hand*** (ACRH) is an upright rectangle which contains whole hand but contains as little background as possible, Figure 2a.

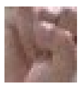
(a)  (b) 

Figure 2. Two kind of positive sample of a hand posture: (a) A traditional positive sample which is an ACRH; (b) A positive sample used in our system which is an AIRH.

Different from the face, these hand positive samples often include a lot of background (see Figure 3). In this case, Viola-Jones detector is affected badly by this part of background. In our approach, the main idea is using the training positive sample which is the inscribed rectangle's area in the hand without background. An ***approximate inscribed rectangle area in the hand*** (AIRH) is an upright rectangle area which has maximum area but does not include background (see Figure 2b). Therefore, our method can avoid the bad effect of background.

(a)  (b) 

Figure 3. (a) a positive face sample use in [1]; (b) a positive hand sample use in [9].

A. *Framework of hand detector*

Figure 4 describes our proposed framework for hand detection. In this framework, n is number of hand postures.

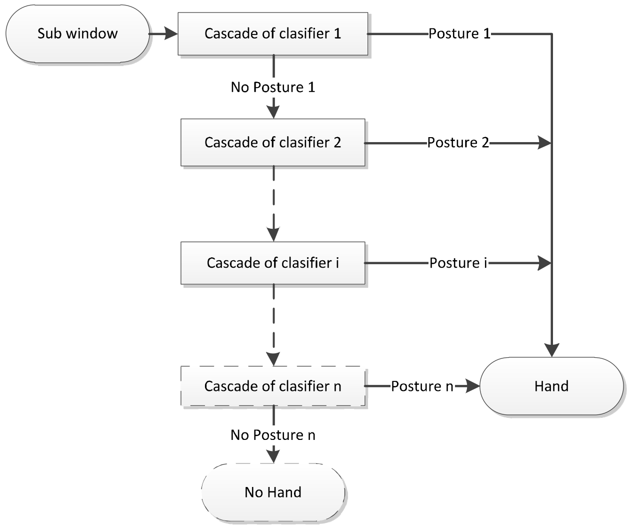


Figure 4. The proposed framework for hand detection.

The value of n depends on specific application system. Cascade of classifiers i is the Viola-Jones detector for detecting posture i. Cascade of classifiers i decides whether a sub-window contains a posture i or not. A sub-window is classified into hand category if there is a cascade of classifiers i classifies this sub-window into *ith* category. A specific sub-window where is position of a hand posture is able to be accepted by a cascades of classifiers that made in order to detect another hand posture. This is a challenge in hand postures classification that is the confusion between different postures. Our study aims at detection hand in the given set of hand postures. The hand posture classification is beyond the scope of this paper. All the sub-windows of the input frame are passed through the system to detect all regions where a hand posture appears. Those sub-windows which are rejected by the initial cascade of classifiers will be processed by a sequence of cascades of classifiers. If any cascade of classifier accepts the sub-window, no further processing is performed; the sub-window is classified as hand region.

The structure of a cascade of classifiers is essentially that of a degenerate decision tree [7]. A cascade of classifiers is described in Figure 5.

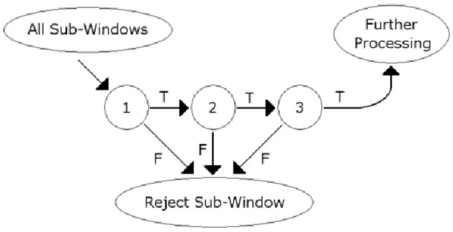


Figure 5. A cascade of classifiers, [1].

Stages in the cascade are strong classifiers constructed by training classifiers using AdaBoost and then adjusting the threshold to minimize false negatives. A variant of AdaBoost is used both to select the features and to train the classifier. The AdaBoost learning algorithm is used to boost the classification performance of a simple learning algorithm. It does this by combining a collection of weak classification functions to form a stronger classifier. For example the perceptron learning algorithm searches over the set of possible perceptrons and returns the perceptron with the lowest classification error. After the first round of learning, the examples are re-weighted in order to emphasize those which were incorrectly classified by the previous weak classifier. Weak classifiers are only required to be slightly better than chance. At each round of boosting, the feature-based classifier is added that best classifies the weighted training samples. The final strong classifier takes the form of a perceptron, a weighted combination of weak classifiers followed by a threshold. In traditional system based on Viola-Jones method, the set of weak classifiers are all classifiers which use one feature from the Haar-like features in combination with a simple binary thresholding decision. In order to avoid the bad effect of background, we propose a concept called Internal Haar-like features. These features are introduced in sub section *B* of section *III*.

The evaluation of the strong classifiers generated by the learning process can be done quickly, but it is not fast enough to run in real-time. For this reason, the strong classifiers are arranged in a cascade in order of complexity, where each successive classifier is trained only on those selected samples which pass through the preceding classifiers. If at any stage in the cascade a classifier rejects the sub-window under inspection, no further processing is performed and continues on searching the next sub-window (see Figure 5).

The cascade architecture has interesting implications for the performance of the individual classifiers. Because the activation of each classifier depends entirely on the behavior of its predecessor, the false positive rate for an entire cascade is

$$F = \prod_{i=1}^{K} f_i.$$

where F is the false positive rate of the cascaded classifier, K is the number of classifiers, and f_i is the false positive rate of the *ith* classifier on the examples that get through to it. The detection rate is

$$D = \prod_{i=1}^{K} d_i.$$

where D is the detection rate of the cascaded classifier, K is the number of classifiers, and d_i is the detection rate of the *ith* classifier on the examples that get through to it.

Thus, to match the false positive rates typically achieved by other detectors, each classifier can get away with having surprisingly poor performance. For example, for a 32-stage cascade to achieve a false positive rate of 10^{-6}, each classifier need only achieve a false positive rate of about 65%. At the same time, however, each classifier needs to be exceptionally capable if it is to achieve adequate detection rates. For example, to achieve a detection rate of about 90%, each classifier in the aforementioned cascade needs to achieve a detection rate of approximately 99.7%.

B. Internal Haar-like features

1) Haar-like features

A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in these regions and calculates the difference between them. Viola et al [1] proposed a very fast Haar-like features computation scheme by using the Integral Image. A simple rectangular Haar-like feature can be defined as the difference of the sum of pixels of areas inside the rectangle, which can be at any position and scale within the original image. Therefore, the exhaustive set of rectangle features is relatively large. Given that the base resolution of the detector is 24 × 24, the exhaustive set of rectangle features is 160,000. This modified feature set is called *2-rectangle feature*. Viola and Jones also defined 3-rectangle features and 4-rectangle features, Figure 6.

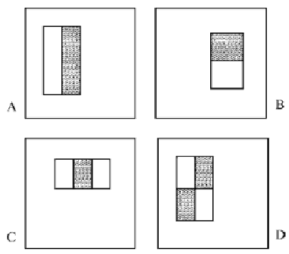


Figure 6. Haar-like features. Example rectangle features shown relative to the enclosing detection window. The sum of the pixels which lie within the white rectangles are subtracted from the sum of pixels in the grey rectangles. Two-rectangle features are shown in (A) and (B). (C) shows a three-rectangle feature and (D) a four-rectangle feature, [1].

The values indicate certain characteristics of a particular area of the image. Each feature type can indicate the existence (or absence) of certain characteristics in the image, such as edges or changes in texture. For example, a 2-rectangle feature can indicate where the border lies between a dark region and a light region.

During training process, the weak learning algorithm selects the single rectangle feature which best separates the positive and negative examples. In the detection phase of the Viola–Jones object detection framework, a window of the target size is moved over the input image, and for each subsection of the image the Haar-like feature is calculated. This difference is then compared to a learned threshold that separates non-objects from objects. Because such a Haar-like feature is only a weak learner or classifier, a large number of Haar-like features are necessary to describe an object with sufficient accuracy. In the Viola–Jones object detection framework, the Haar-like features are therefore organized in a cascade of classifier to form a strong learner or classifier.

Lienhart *et al.* [17] extended the Haar-like features by adding 45° rotated features and proposed 45° rotated features computation scheme by using the Integral Image also, Figure 7. The features that we use in our system are the Extended Haar-like Features.

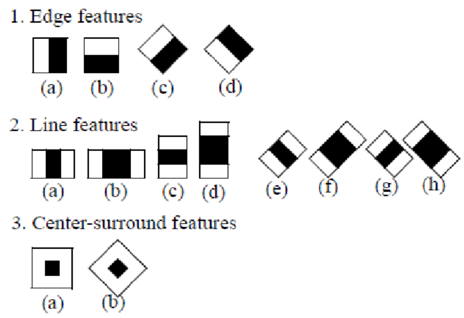


Figure 7. Extended Haar-like Features, [17].

2) *Internal Haar-like features*

All of previous works, for example [6][9], that applied Viola-Jones detector into hand detection or hand posture recognition use positive samples as the rectangle region contains whole hand so the positive sample includes a lot of background. Figure 8 shows Haar-like features for 3 types of postures used in [6].

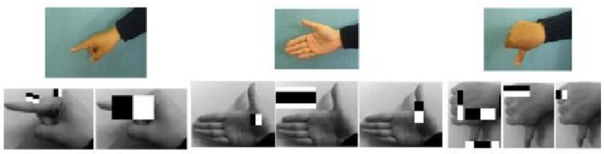


Figure 8. Examples of Haar like features for 3 types of postures used in [6].

As we can see, there are many rectangle features found on background area. During training process, some of them are chosen to build weak classifier. This means the weak classifier keeps characteristics of the background in its knowledge. So, when detect hand, the detector may accept a background sub-window that has characteristics of background kept by the weak classifier. The detector also may reject foreground sub-window that has not these characteristics.

In order to avoid the bad effect of the background on the detector, we propose a technique for leading the trainer only focusing on and choosing from features inside the hand region. In this technique, the positive sample is the AIRH (Figure 2b). The Haar-like features used in this technique are ***Internal Haar-like features***. Generally speaking, an ***Internal Features*** is a feature that reflects only own characteristics of the interested object but does not include characteristics of others components such as background.

In the following, we give the definition for Internal Feature, Internal Haar-like Feature and its property:

Definition 1: Internal Features

An internal feature is a feature extracted on a region inside the object of interest (without background).

Definition 2: Internal Haar-like Features

An internal Haar-like feature is a Haar-like feature extracted on a region inside the object of interest (without background).

Property of Internal feature: As computed on a region inside the object, Internal Feature is invariant to background changing.

In our method, we exploit the invariant to background changing property of Internal Feature to avoid the bad effect of background on the hand detector. When we use AIRH positive samples to train Viola-Jones detector, the learner only remembers Internal Haar-like features. Therefore, the learner only remembers own characteristics of the hand postures which helps detector separating non-hand postures from hand postures. Therefore, the detector will not make confusion of the hand posture with the background. Of course, if there is a background sub-window that is very similar to the positive hand posture then the detector will make confusion of this background sub-window with a hand posture. There is a problem that may occur in case of a specific hand posture has an internal region with poor characteristics, there are many background regions similar to the internal region of this hand posture. In this case, the detector will make a lot of confusion between the hand and the background. This problem can be resolved in the further processing steps of the hand postures recognition system. According to the experimental results, in overall, the detector uses Internal Haar-like features is better than the detector uses traditional Haar-like features.

IV. EXPERIMENTS

A. *Set of Hand postures*

To validate our idea, we test on a set of 15 hand postures (Figure 9). The number of postures in this posture set is higher than previous works. Tran et al. [6] used 4 postures. Kölsch et al. [9] used 8 postures. Our set of postures is diversified. There are finger closed postures. There are also open finger postures with difference number of open fingers. Posture 2 faces the back side of the hand to the camera. Posture 3 faces the back side of all the fingers to the camera. Another else faces front side of the hand to camera. The set of postures used in this paper is representative enough to evaluate a method of hand detection. Therefore the experiment in this paper is reliable.

B. *Training databases preparation*

We use a Hercules Deluxe Optical Glass webcam with 320x240 resolution and 30 frames per second to capture the videos in order to collect the training databases. The videos are captured in a real environment inside of the office of L3i laboratory with neon lighting, complex background. Twenty two people were asked to act all of the 15 postures twice at different positions in the same office. While acting postures, actors were asked to move the hand in order to change the background including case of the hand overlapping on the face. For each hand posture, we collected two positive sample databases from the same original videos. The numbers of samples of the databases are equal that is 4400. The first one is a set of ACRHs. The second one is a set of AIRHs (see Figure 2). Figure 10 shows examples of ACRHs and AIRHs of relevant postures.

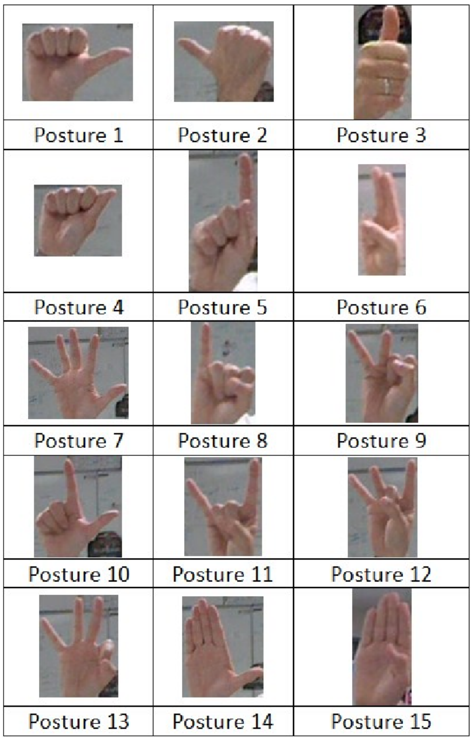


Figure 9. Set of hand postures.

C. Experimental result

Two detectors were built to compare with each other. The first one is the Traditional Viola-Jones detector which uses Haar-like feature with traditional way. We use ACRH positive samples for training Traditional Viola-Jones detector. The second one is our detector which uses Internal Haar-like features. We use AIRH positive samples for training our detector. Both of two detectors are tested on the same video captured all of 15 hand postures that acted by a person in similar environment to training video capturing. The total number of frames is 2355. Each frame includes a hand posture. These frames are evenly distributed throughout the postures. This testing set is enough large to archive a reliable result.

We counted manually the result. If there is much positive detection with an approximate coordinates then we count only one time. We compared the results through Precision, Recall and F-measure. This comparison is shown in Table I and Table II. In overall, the performance of our hand detector is better than the traditional Viola-Jones detector. Examples in Figure 11 and Figure 12 illustrate the bad effect of background on the traditional detector. These examples also illustrate our improvement in avoiding this bad effect.

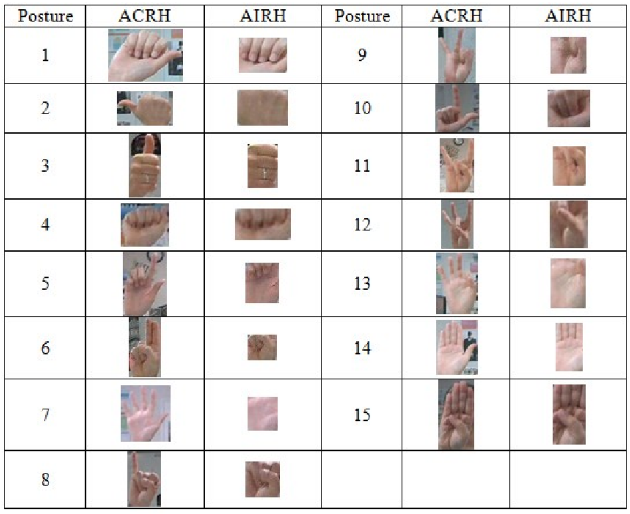


Figure 10. Examples of ACRHs and AIRHs of relevant postures.

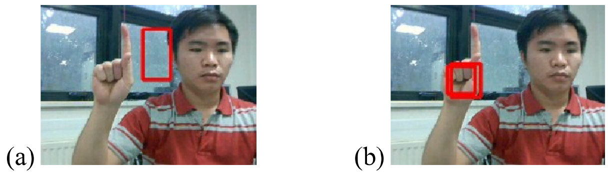


Figure 11. An example: Hand detection results of (a) traditional detector; (b) our detector

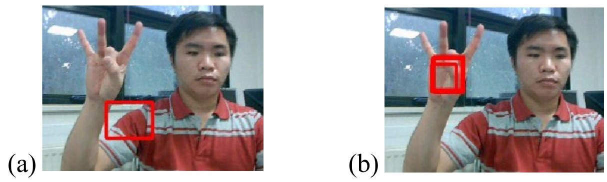
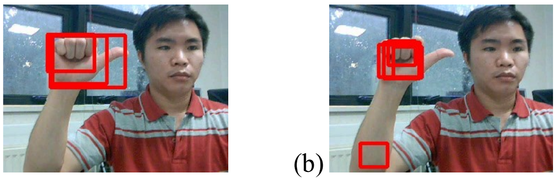


Figure 12. An example: Hand detection results of (a) traditional detector; (b) our detector

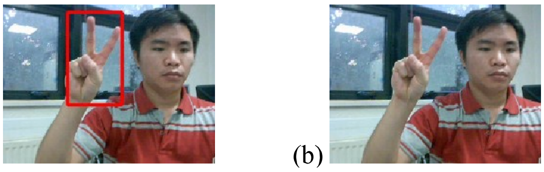
However, there are some specific cases in which the Viola-Jones detector works better than ours. In Figure 13, our detector gives a false positive that is a sub-window in the arm area because this sub-window is similar to AIRH of that posture. But the traditional Viola-Jones does not make that confusion. The sub-window is not similar to the ACRH because the ACRH contains a lot of information of background. In Figure 14, our detector does not detect the hand while the traditional Viola-Jones works well. This because the AIRH of this posture has little characteristic which helps the detector distinguish the hand from the background while the ACRH contains information of both the background and the hand that helps the detector distinguish the ACRH from another regions. And addition, the background in this frame similar to the background in some ACRH positive samples.

In the case that the working environment of hand detection system is defined before training the detector, we can use traditional Viola-Jones detector to get better result. In this case, the background should be remembered during training. In case when we do not know the environment of the system, we can improve the performance of our detector by using higher resolution camera in order to get more detail information of the hand.



(a) (b)

Figure 13. An example: Hand detection results of (a) traditional detector; (b) our detector



(a) (b)

Figure 14. An example: Hand detection results of (a) Traditional detector; (b) Our detector

TABLE I. THE RESULTS OF TWO DETECTORS

Performance measurements	**Traditional Viola-Jones detetor**	**Our detector**
Number of detected positives	2094	2371
Number of true positive	1863	2315
Number of false positive	231	56

TABLE II. COMPARISON BETWEEN TWO DETECTORS

Performance measurements	**Traditional Viola-Jones detetor**	**Our detector**
Precision	89%	98%
Recall	79%	98%
F-measure	84%	98%

V. CONCLUSION AND FUTURE WORK

In this paper, we introduced two new concepts that are Internal Features and Internal Haar-likes features and proposed a framework of hand detection based on Viola-Jones face detector [1] and Internal Haar-like feature. The experimental result shows that this method is promising for detecting hand while does not take background into account. In the near future, we will validate this method with bigger databases and use this detection result for hand gesture recognition. We will also try to combine this method with some another methods in order to do further steps of hand gestures recognition system.

ACKNOWLEDGMENT

The research leading to this paper was supported by the National Project B2011-01-05 "*Study and develop an object detection and recognition system in smart and perceptive environments*". We would like to thank the project and people involved in this project.

REFERENCES

[1] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 2001, pp. 511-518.

[2] L. Dung and M. Mizukawa, “Fast Fingertips Positioning Based on Distance-based Feature Pixels,” in *Third International Conference on Communications and Electronics (ICCE)*, 2010, pp. 184-189.

[3] X. Zhu, J. Yang, and AlexWaibel, “Segmenting hands of arbitrary color,” *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, pp. 446-453, 2000.

[4] Q. Yuan, S. Sclaroff, and V. Athitsos, “Automatic 2D Hand Tracking in Video Sequences,” in *Seventh IEEE Workshops on Application of Computer Vision*, 2005, no. 2004, pp. 250 - 256.

[5] J. Choi and B.-kuk Seo, “Robust Hand Detection for Augmented Reality Interface,” in *Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry*, 2009, vol. 1, no. 212, pp. 319-322.

[6] T.-T.-H. Tran and T.-T.-M. Nguyen, “Invariant Lighting Hand Posture Classification,” in *IEEE International Conference on Progress in Informatics and Computing (PIC)*, 2010, pp. 827 - 831.

[7] P. Viola and M. J. Jones, “Robust Real-Time Face Detection,” *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, May 2004.

[8] K. Mathias and M. Turk, “Analysis of Rotational Robustness of Hand Detection with a Viola-Jones Detector,” in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, 2004, pp. 07 - 110.

[9] M. Kölsch and M. Turk, “Robust Hand Detection,” in *The Sixth IEEE international conference on Automatic face and gesture recognition*, 2004, pp. 614-619.

[10] Z. Zhang, R. Alonzo, and V. Athitsos, “Experiments with computer vision methods for hand detection,” in *Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '11*, 2011, pp. 1-6.

[11] E.-J. Ong and R. Bowden, “A Boosted Classifier Tree for Hand Shape Detection,” in *Face and Gesture Recognition*, 2004, pp. 889 - 894.

[12] V. Athitsos, J. Wang, S. Sclaroff, and M. Betke, “Detecting Instances of Shape Classes That Exhibit Variable Structure,” in *Work*, pp. 1-15.

[13] A. Thayananthan, B. Stenger, P. H. S. Torr, and R. Cipolla, “Shape context and chamfer matching in cluttered scenes,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003, pp. 1063-6919.

[14] S. Belongie, J. Malik, and J. Puzicha, “Shape Matching and Object Recognition Using Shape Contexts,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 24, pp. 509-522, 2002.

[15] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf, “Parametric correspondence and chamfer matching: Two new techniques for image matching,” in *International Joint Conference on Artificial Intelligence*, 1977, pp. 659-663.

[16] J. Coughlan, A. Yuille, C. English, and D. Snow, “Efficient Deformable Template Detection and Localization without User Initialization,” *Computer Vision and Image Understanding*, vol. 78, no. 3, pp. 303-319, Jun. 2000.

[17] R. Lienhart and J. Maydt, “An extended set of Haar-like features for rapid object detection,” in *Proceedings. International Conference on Image Processing*, 2002, pp. 900-903.