International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 5, Sep-Oct-2025, ISSN (Online): 2395-566X

Monocular Vision-based Real-Time Pallet Detection in Harsh Industrial Environment

Chung Hyok Pak, Un Sim Ri, Se Hyon Kim
Faculty of Automation Engineering, Kim Chaek University of Technology, Pyongyang, Democratic People's Republic of Korea

Abstract- It is the important global trend to use the unmanned production lines in order to meet the demand of customers and improve the efficiency for industrial processes. The automated storage and delivery system (ASDS) is one of the main components of the unmanned production line. It consists of many shields and several automata and complex control systems for loading and unloading, so its cost is so high. For the tradeoff of the cost and performance of cargo handling, forklift is a best alternative to the lack of financial ability enterprises/factories. In this paper, we propose a pallet detection method to allow forklifts to engage the pallet autonomously using only a monocular vision on the forklift in the harsh industrial environment. To reduce the number of features and increases the detection efficiency, we describe the pallet features by combining the Haar-like features and multi-block local binary pattern (MBLBP). 8 sets of Haar-type encoding models make the LBP feature better to encode the local structure. Adaboost classifier that use distribution information of features in training set, allows to detect pallet candidates with high accuracy and efficiency in harsh industrial environments. In particular, improved feature to maximize the margin when pattern classes are projected onto the classification hyperplane is used to enhance the discriminate ability of classifier and reduce the computational cost. The analysis of the geometric features of the pallets using integral-sum-difference (ISD) excludes the wrong candidates with high efficiency. The experimental results demonstrate that our proposed algorithm could detect the pallet with average rate of more than 98% and is robust to environmental changes.

Keywords – Pallet detection, MBLBP, haar-like feature, wavelet and Adaboost.

I. INTRODUCTION

As the science and technology progresses, automatic machine is replaced with human's activity for production to overcome the lack of the labor and skills, and improve the productivity.
This trend has led to the development of unmanned technologies in the material transportation processes.
The most of modern factories introduced the unmanned processes so it has resulted in an increase in the level of intelligent manufacturing and AGV plays an important role.
The AGV is widely used in automatic transportation systems transporting products and components, because it can speed up the transportation and dramatically save its cost.

Generally, it is the basic and common task of autonomous forklift to lift or unload the pallet from the rack and carry out the pallet from one position to another.
Therefore, it is prior condition for performing the task well enough to detect the pallet and determine the relative position of the forklift using sensors.
In last several decades, many methods have been proposed for the detection and pose estimation of pallets.
In the early stage, RFID and infrared sensors are used to detect the pallets and measure the distance between forklift and pallets.1,2
Sensor-based methods have low computational cost in the detection and recognition of pallets and are less affected by environment change, so it can perform tasks with high accuracy.
However, it is so expensive and time-consuming to build the operating condition for forklifts and often has the disadvantages of being maintained.
Thus, image-based methods have been proposed in order to overcome these disadvantages and they are already mainstream in the research and implementation of AGVs. Guang-zhao Cui. et al3 proposed a methodology that detect pallets and determine its position from foreground images obtained by 2D cameras mounted on autonomous forklift.
The developed system has significantly reduced costs, but it has affected by the influence of environmental conditions, i.e., illumination.

In most autonomous forklifts, it is another useful option for the detection and pose estimation of the pallets to use RGB-D camera4,5 which form the depth images. These kinds of sensors can detect and estimate pose of pallets with high accuracy

© 2025 IJSRET

because it gives depth images. On the other hand, it is too expensive to compute and the sensor which serves good quality of measurement data is also expensive.
Also, ranging Lidar6 was used for pallet detection.
Some researchers have installed Lidar on the forklift and used it to scan panoramic views periodically.
It determines the pallet's position with high accuracy from collected 3D point cloud data.
But the laser scanner is too expensive and has high computational complexity.
In addition, various sensor fusion methods have been proposed to enhance the accuracy of pallet detection and its pose estimation.

As mentioned above, the pallet detection system in autonomous forklift has to be cheap, has a high detection efficiency, and satisfy the requirement of real-time performance, so that it can support rapid and efficient decision making. 2D cameras are very cheaper than laser scanners and 3D cameras and are very convenient for maintenance.
From these conditions, we can detect pallet with high accuracy in harsh industrial environments using only 2D camera, independent of different environmental change. This approach will make the system more simple and lead to the extension to other application areas.

Literature review
Pallet detection in forklift is the most important and key technique.
Researchers mainly have used image sensors to achieve this detection goal.
Object detection is the fundamental task of the computer vision, which consists of feature extraction and detection process.

In the image-based approach, we capture the foreground or depth images containing pallets by using the image sensor and separate the pallets from the background using various visual and geometric features of the pallets.
Simple ways to detect pallets using features are to fit to the pallets with artificial labels7, 8.
These ways are not so cheap since all pallets may have to be customized, and when the visual labels may be masked, the detection capability is degraded, therefore additional algorithms considering it should be added.
For avoiding these problems, methods that detect the features of the pallet itself such as color, spatial and geometric features have been proposed.
First, more details about the detection and pose estimation of the pallets using color features are proposed.3, 9
The approach presented in [3] consists of offline subsystem and online search subsystem.

The offline subsystem gets the single threshold using the color features of the pallets to support the detecting condition for online subsystem.
The online subsystem gets the pallet region based on color similarity and then estimate the pallet pose by applying certain operations to the image such as Morphology, Sobel operation, Hough transform.
This way may not be used in complicated environments.
Because it requires camera calibration and needs constraint condition on the pose of the pallet.
And Also, there is no way to detect multiple pallets and need to continuously change the offline according to the illumination changes.

To overcome lowering the detection efficiency in offline, the authors9 applied the Otsu algorithm based on the fact that there is a clear difference between the pallet color and the background color, and obtained an adaptive threshold, which is used to segment the image.
J. Pagès et al.14 obtained the distribution-based features of objects based on the color features of pallets with trainsets and then classified the pallets using Fisher’s linear discriminant function (LDF).

Distribution-based features may improve detection efficiency in a simple background, but may be less efficient in sundry background.
Thus, there has been a lot of researches on using spatial and geometric features because color features cannot be safely extracted from image features for illumination changes and environmental changes.
Haar-like feature is one of the most popular feature descriptors for face recognition, which reflect the texture features of objects well.10, 11
In [17], a novel pallet detection system using a monocular vision system was proposed.

This paper consists of two stages of detection and tracking. We detect pallet by using Directional Weighted Overlap ratio (DWO) and pallet structural feature and then reduce the most of the dynamic background in the tracking step and improved the processing efficiency.
Authors used Haar-like features using integral images as features reflecting the pallets, and detect pallets by AdaBoost algorithm to search candidate regions of pallets and determine the adaptive threshold of color change using pallet’s adaptive texture features to avoid mismatch candidates.
This study improves the detection rate of the pallet and reduces the cost of constructing the system.
It is inevitable that the feature order increases significantly as the resolution of the image increases in order to satisfy the scale invariance in object detection.

Because the face recognition task does not require high real-time performance, the Haar feature has been successfully applied, but the detection task of pallet requires real-time performance.

© 2025 IJSRET

International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 5, Sep-Oct-2025, ISSN (Online): 2395-566X

Also, the Haar-like feature is insensitive to variations of illumination and may reduce the detection efficiency.
Under these situations, a local binary pattern (LBP) feature12 was proposed, and the LBP is the main property of the local image texture and it shows that the LBP histogram is a very powerful texture feature descriptor.
In [13], a classifier is designed by Boosting decision tree using the edge features of pallets and the LBP histogram features to detect pallets. This feature also does not satisfy the requirement of real-time performance.
There are many approaches using geometric features to detect pallets.

In [5], They obtained RGB and point cloud data using RGB-D camera and eliminated undesirable points, such as wall and ground using the RANSAC algorithm.
Then, they have got the center of the three support logs from the pallet's geometric feature and detect the pallet by the distance between the centers, and determine the relative position and pose of the forklift.
In [15], authors used Lidar to detect pallets.
Single-shot detector (SSD) algorithm allows to detect pallet and measure relative distance between forklift and it.
This way can significantly improve the detection rate, as they are independent of image distortion and illumination changes, unlike using cameras as image sensors.
On the other hand, these ways can make you detect more correctly, compared to the detection method of pallet using 2D images, because 3D point cloud data can exploit many information such as color, texture, geometric features and spatial distribution.

Unfortunately, because this way is highly affected by the quality of point cloud data and the pallet's geometric type, and the detection rate of the pallet is also greatly limited, these are unsuitable for the real-time detection system and it is very expensive.
Some researchers have tried to improve the recognition rate of the pallet with machine learning methods.19, 20
They trained SSD with RGB images and get a training model and then detect pallets and determine its position by combining it with depth data.
But the performance of the machine learning methods is mostly determined by the training data set and may not be able to satisfy the real-time performance due to the high computational cost.

It is shown that the image-based method provides more information about the foreground than the physical sensor, so it becomes the mainstream of sensors that recognize the external environment in an intelligent system.
But It still have many difficulties to detect pallets and determine its pose by image processing-based method due to the large of computational amount.
This study allows to detect multiple pallets using a monocular vision system mounted on a forklift.
Experimental results show that the proposed approach is more accurate and can detect multiple pallets in real industrial environment.

Contribution
In response to the above problems, we propose a novel pallet detection method which is robust to environment, low-cost and real-time with only monocular vision.
The main contribution of this paper are as follows.
(1) We propose HMBLBP feature descriptor to that is relatively low in computation and efficient in pallet detection.
This approach uses MBLBP to take advantage of Haar-like features and LBP while reducing the dimension of features, so allows to ensure the real-time for detection.
Also, it is robust to illumination changes, and further improve the detection efficiency against changes in view angle.
(2) We design the Adaboost classifier that provides the high accuracy for pallet candidate detection.
For purpose, we design the Adaboost classifier with distribution-based feature.
It allows to improve the detection rate of the pallet candidate and detect multiple pallets.
(3) We apply Haar wavelet transform to detected pallet candidate region and then detect the pallet correctly from the geometric feature of the pallet.

The remainder of paper is organized as follows.
In Section 2, system architecture of proposed pallet detection is described.
In Section 3, the proposed pallet candidate detection method based on feature descriptor combined with Haar-like feature, MBLBP, distribution-based feature is described.
Also, pallet detection way based on the geometric feature of pallet is proposed.
Section 4 outlines experimental comparison analysis for verifying the proposed way in applications.
This work may be contributed to reduce the differences between the theoretical and field application research.

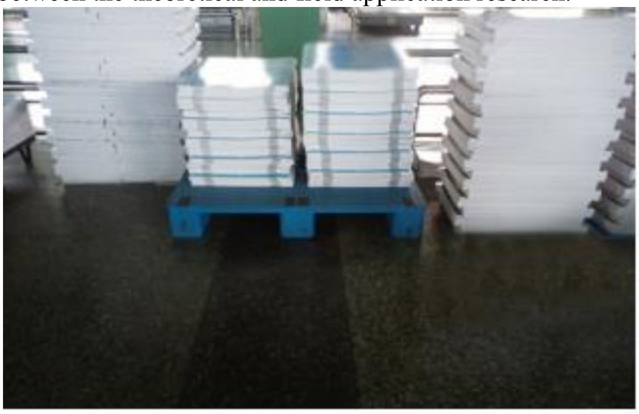


Figure 1: Pallet(1200(W)*152(H)*1080(D)mm)

© 2025 IJSRET

International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 5, Sep-Oct-2025, ISSN (Online): 2395-566X

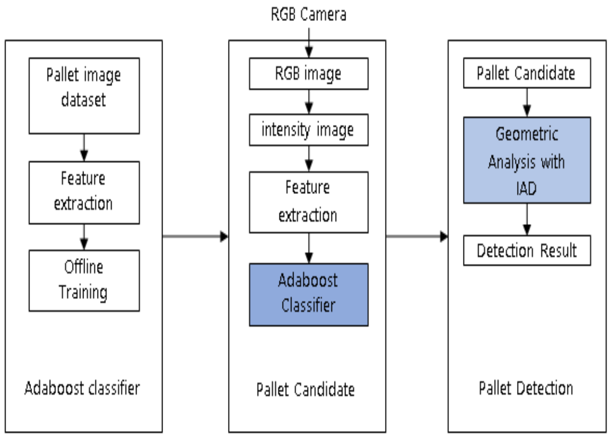


Figure 2: System Architecture

II. SYSTEM ARCHITECTURE

In the publishing industry, pallets are transported by AGVs. (Fig 1)
The pallets are shipped by a production line from the automated warehouse and the forklift transfer these pallets to the destination and load the complete publication onto the automated warehouse entry line.
The diagram of proposed method shows in Fig 2.
First, Proposed way for detecting and positioning of the pallet is divided into pallet detection classifier modeling (offline subsystem) and detection process (online subsystem).
In designing of the pallet detection classifier, we extract the features of the pallet from the given training data and train the Adaboost classifier by it.

This training is done offline.
Also, in the online part we detect the pallet candidate from the foreground image obtained by the trained classifier.
Second, we detect the pallet from its candidate using the geometric features.

III. PROPOSED PALLET DETECTION METHOD

feature extraction of pallet

Feature extraction is one of the important steps in the object detection and the feature description directly gives an impact on the effectiveness of object detection.
As mentioned in Introduction, the pallet has relatively simple geometric structure, therefore there is no need for real-time pallet detection to use feature descriptors such as SIFT or ORB that are complex and have high computational costs.
In [17], Haar-like feature with Adaboost classifier were used for pallet detection.

As shown in Fig 3, Haar-like features are defined as the difference between the total brightness sum of white and black regions, it is simple, fast to compute, and can derive rectangular features.
But it is very uncomfortable to detect using Haar-like features under the conditions of complex background information and varying illumination.

This is because the Haar-like feature is relatively sensitive to the shape and illumination changes of the target and ignores the spatial distribution information between image pixels.
LBP is a local texture rendering descriptor, which is simple to compute, insensitive to illumination changes, and can represent the texture information of an image well.
Therefore, we try to achieve better detection performance using feature descriptor combined with Haar-like features and LBP.

HMBLBP feature

We extract the features of the pallet using Haar-type LBP feature descriptor since Haar-like feature descriptor may not effectively represent the texture information embedded in the image.

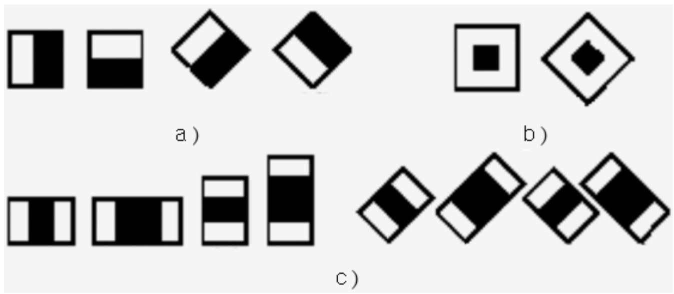


Figure 3. Haar-like features
(a-contour feature, b-central feature, c-line feature)

Ojala et al.12 proposed a local binary pattern, which is a way to describe texture, and has been used for various object detection, including face recognition.
This operator represents the neighbor pixel sequence as the binary pattern, generated by comparison between center pixel value and its neighbor for each pixel in the image.
It is presented as Eq. (1), which characterizes the local spatial structure of the image.

$$f_{R,N} = \sum_{i=0}^{N-1} s(p_i - p_c)2^i, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad b \quad (1)$$

Where pi is one of N pixels around the center pixel pc in a circle of radius R or square.
The LBP feature descriptor is a binary one, known as a relatively simple and robust to illumination variations.

Recently, Zhang et al.21 proposed a multi-block LBP (MBLBP) feature, which extends the single-pixel-centered feature to block-based features.
The MBLBP feature consists of 3x3 blocks and each cell contains several pixels. (Fig 4.)
The 8-bit encoding sequence is generated along a circular pattern from 1 to 8, in Eq. (1) the intensity of the center pixel is replaced by the average intensity of the center block, and the intensity around the center pixel is replaced by the average intensity of the surrounding block.

The multi-scale version of MBLBP is shown in Fig 4.
The MBLBP feature constructs a feature descriptor with a bit sequence that reflects whether the average intensity in each block is greater or less than the average intensity of the center block.
It is insensitive to intensity variations, and the number of features is very smaller than LBP, and the object detection rate does not differ significantly.

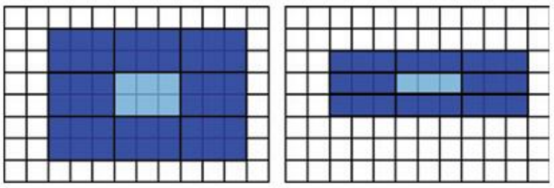


Figure 4. MBLBP feature

If we use the average intensity of the blocks, we may be able to lose some of structural information, and the Haar-type feature is used to calculate the average intensity for more accuracy.
Qiu Qin-jun et al.25 successfully detected vehicle by using NHLBP (new Haar-type LBP) feature combined Haar-like feature with LBP feature.
Authors constructed a 5x5 small window, suggested 12 sets of encoding models and then finally built a new NBLBP operator.

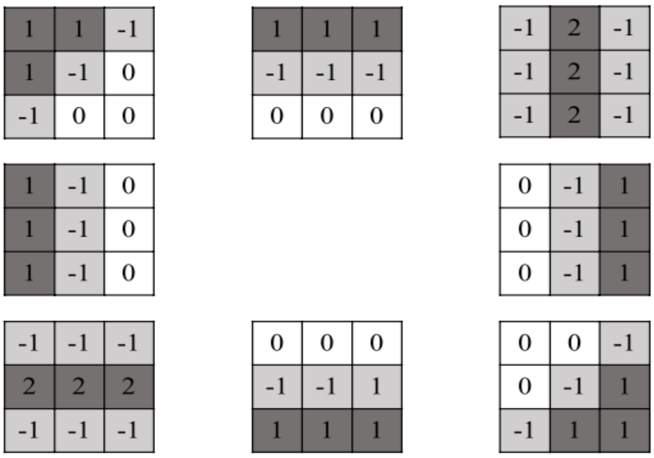


Figure 5. 8 sets of encoding models

Here, we construct a 3×3 small window and propose 8 sets of encoding models that reflect edge features, line features, and diagonal features, encoding models are shown in Fig 5.
In the encoding model, the weight of the dark regions, light regions, empty regions are 1, -1, 0 respectively, so that it correctly reflects the change information of the Haar-type feature texture.
The encoding model is given in a symmetric form, as can be seen in the Fig 5, the dark and bright regions involves the edge features, diagonal features, and corner features of Haar-like features.
In common, the Haar-like feature involves line feature, but this is excluded because almost all the pallets have a single color, and the line feature can We calculate the HMBLBP feature using encoding model by the following operation equation:

$$f_{HMNLBP,k} = \sum_{i=0}^{N-1} s\left(p_{h_k,i} - p_{h_k,c}\right)2^i, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{2}$$

$$p_{h_k,i} = M_k \cdot W_i(x,y), \quad p_{h_k,c} = M_k \cdot W_c(x,y) \quad k = 1, \cdots, L$$

Where Mk is the encoding model, L is the number of encoding models, · is point multiplication operation,   is the intensity of neighboring and center blocks of multiple blocks, and Wi(x, y) is the window area to operate with 3×3 small window, each of which constitutes a multi-block LBP feature.
As shown in Eq. (2), the number of MBLBP features is equal to the encoding models for a single multi-block sliding window.

Haar-like feature is a set of features obtained by a kind of image filter filtered by rectangular features.
Thus, we can better describe the spatial characteristics of the pallet by calculating the intensity of the central block of Eq. (2) and the nearing block using the parent modeling,
The experimental results show that we detect more correctly by HMBLBP feature than Haar-like feature.
This feature can be used effectively not only for pallet detection but also for detection of certain geometric objects.

Combination of HMBLBP and distribution-based features
Although we describe the features of the pallets using the HMBLBP feature and detect a pallet with these features, it is insensitive to changes in pallet viewing angle.
This is why the features of the object is described differently with the changes in pallet viewing angle whatever features we used and the classifier needs better performance.
Pavani et al.22 proposed a new optimized weighting factor selection method to improve the recognition ability in object detection.

$$w_{opt} = \arg\max_{w} \frac{w^T S_b w}{w^T S_w w} \tag{3}$$

$$S_b = \left(\mu_p - \mu_n\right)\left(\mu_p - \mu_n\right)^T$$

$$S_w = S_p + S_n$$

© 2025 IJSRET

International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 5, Sep-Oct-2025, ISSN (Online): 2395-566X

$$S_p = \sum_{x_j \in x_p} (x_j - \mu_p)(x_j - \mu_p)^T$$

$$S_n = \sum_{x_j \in x_n} (x_j - \mu_n)(x_j - \mu_n)^T$$

,

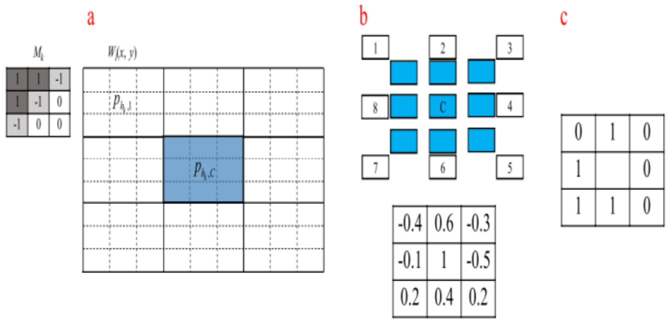


Figure 6. D-HMBLBP feature

a) combination MBLBP with Haar-like feature, b) LBP with distribution-based optimal weight, c) result

Here wopt is the optimal projection direction which maximize the distance between- the positive and the negative training dataset and minimized the distance between in training datasets, µp, µn are the mean vectors of the positive and negative training samples, Sp and Sn are the within-pattern scatter matrix of positive and negative samples, Sb and Sw are the between-patterns and within-pattern scatter matrix, respectively.

Fig 7 shows the 2-D distribution of the positive and negative training samples projected to a specific direction and the optimal projection direction.
Fig 7a shows the projection to the correct direction and Fig 7b shows the projection to incorrect direction.
Therefore, we combine the HMBLBP feature with the distribution-based feature, and then determine the optimal projection direction to maximize the margin between the positive and negative sample feature vectors using the distribution information of the HMBLBP features extracted from the positive and negative samples.
D-HMBLBP is defined as follows:

$$f_{D-HMBLBP,k} = \sum_{i=0}^{N-1} g\left(p_{k_j,i}, p_{k_j,c}, w_i, w_c\right) 2^i \qquad (4)$$

Here Pc is the mean intensity of the central cell, Pi is the mean intensity of the ith directional cell, wc, wi is the combined weight coefficient of the central and adjacent cells, respectively.
The function g reflects the relationship between Pc and Pi and it is defined as following:

$$g(x) = \begin{cases} 1, & w_i p_{k_j,i} \ge w_c p_{k_j,c} \\ 0, & w_i p_{k_j,i} < w_c p_{k_j,c} \end{cases} \qquad (5)$$

We calculate wi by Eq. (3).
Fig 6 shows the detail process which generate the D-HMBLBP feature.

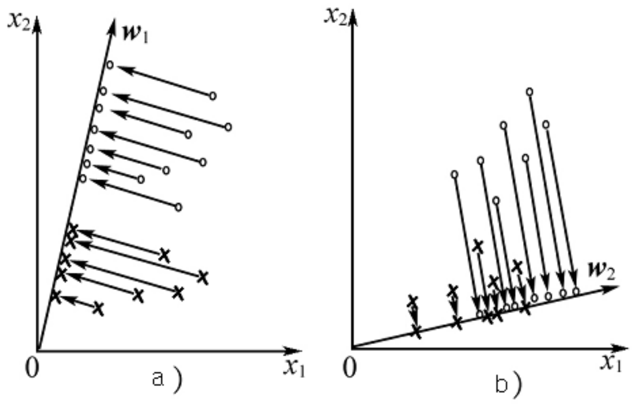


Figure 7. a) projection into the true direction,
b) projection into the false direction

Fig 6a shows the combination process of Haar-like features and MBLBP features, where each block consists of 3x3 cells.
Each block is separated into eight center-neighbor features, and the optimal weight of each cell is calculated by Eq. (3) as shown in Fig 6b.
Distribution-based feature allows to improve the object recognition ability more than LBP.
Finally, we calculate the D-HMBLBP feature by Eq. (4).
D-HMBLBP is a special case of MBLMP and the difference is that all blocks/cells around the rectangular region have different weights.
We can improve the recognition ability of the classifier and reduces the extraction time to exclude the negative samples in extracting the object features by using the distribution-based information of the training samples.

Classification and candidate detection

- **Classifier design**

It is the most common approach in recent research, to extract the essential object features and then detect objects from the foreground using extracted features.
In [18], they designed classifier for pallet detection by using Single Shot Multi boxes Detector (SSD) and finally detected a pallet offline using both positive and negative sample training data.
They designed a classifier using the cascade classifier proposed by Viola et al.24 to detect a pallet and use the D-HMBLBP feature and AdaBoost algorithm.
They used a cascade classifier designing method, which is a combination of weak classifiers to recognize face.
The D-HMBLBP feature is a weak classifier, and combination of weak classifiers can improve the recognition speed of pallet and leads to a lower incorrect detection rate.

© 2025 IJSRET

International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 5, Sep-Oct-2025, ISSN (Online): 2395-566X

In this paper, we used AdaBoost learning algorithm to detect a candidate of pallet by a strong classifier consisting of N weak classifiers.
Adaboost learning algorithm gives more weight to misclassified patterns when each weak classifier is designed during training and calculates the optimal threshold and weight, and iterates above process by a certain number of times to reduce the classification error.
We need a large number of positive and negative samples to train the strong classifier for designing the strong classifier by using the boosting algorithm.
Although the strong classifier is well designed, it is impossible to detect multiple pallets with 100% accuracy under the different environmental changes.
If a detected pallet candidate region is not a pallet, the candidate is discarded and the candidate region that remains at the end is considered as a pallet.

- **Pallet candidate detection**

Once a strong classifier is designed, it can detect the candidate of pallet in real time.
One of the main goals we require in pallet detection is the real time performance.
Therefore, in this case the online image processing should be as fast as possible.
First, we convert the foreground image entered from the camera into an intensity image.
We give weight to each component values of the RGB image and convert it to the intensity image by following equation:

$$I_i = \{I(i,\ j),\ 1 \le i \le ry \quad \text{or} \quad 1 \le j \le rx \tag{6}$$

$$I(i,\ j) = 0.299\, r(i,\ j) + 0.587\, g(i,\ j) + 0.114\, b(i,\ j) \tag{7}$$

Here, the weight coefficients of the R, G, and B components are determined by imitating the human visual characteristics.
Next, we filter acquired intensity image.
As mentioned in the following section, we need to filter to avoid the effect of image noise on the pallet's geometric detection.

There are several ways for filtering images, including linear filtering, nonlinear filtering26, wavelet filtering and etc.
In linear filtering, we give weights to all pixels that the operation window contains and it can result in image blurring.
Linear filter is easy to design and it is suitable for frequency response analysis, however nonlinear combination of neighboring pixels can achieve better performance.

Therefore, we use a filter that uses α-mean median, a type of nonlinear filter, which selects the mean value of all pixels except the maximum and minimum values from the neighborhood of each pixel as the median value. Since the impulsive noise value is usually in outside of the neighboring pixels, the median filter can filter such bad pixels and preserve the edge feature of the image well.
Moreover, this guarantees the real-time performance of the filtering algorithm.
Finally, we use AdaBoost classifier to detect the candidate pallet in the image.

Fig 8 shows the result of the pallet candidate detection.
As shown in Fig 8, four pallet candidates were detected, only one of which is the real pallet.
Three false-detected pallets were detected due to their similarity in shape to the pallets.
The pallet candidates detected by a above procedure are then checked to identify whether they are pallets or not.

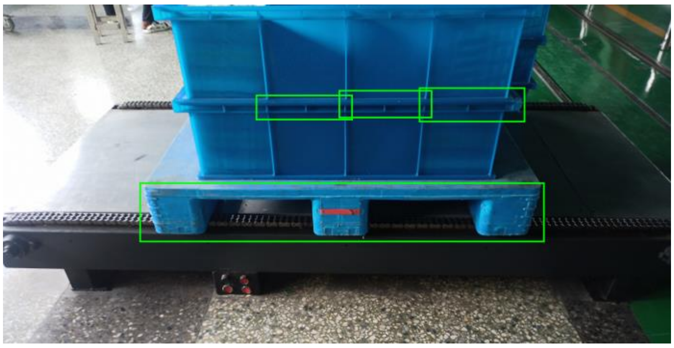


Figure 8. the detection result using D-HMBLBP feature descriptor and Adaboost classifier

Pallet detection based on geometrical feature

The geometric feature of the pallet we want to detect is very straightforward.
The pallet has three support logs, a top plate on the support logs, and a support logs are aligned at equal spacing.
There are also two holes in the middle of the pallet, and this hole's color is much deeper than the support log's color.
Due to this structural feature of the pallet, the pallet's appearance has a dramatic shading area.

Therefore, there are two regions, high color change region and low color change region at the gray scale from each other in the pallet support log's region.
It is very important to obtain an index that reflects the pallet's basic structure by color changes in the support area for the correct pallet detection.
As mentioned in Introduction, in [17], the authors calculate an adaptive threshold from a percentage of the pallet front's gray value for a binary image to achieve a feature index that reflects the pallet structure from the image and analyzed the structural features of the pallet candidate using binary images and sliding search windows.

On the other hand, the pallet detection should be started when the forklift moves towards the pallet and enters the foreground, then the forklift can match with the pallet accurately by only detecting a pallet and its attitude.

© 2025 IJSRET

International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 5, Sep-Oct-2025, ISSN (Online): 2395-566X

The forklift has to go to front of the pallet to match with it accurately and it is necessary for that to detect the pallet correctly, calculate the relative position and attitude of the forklift even if the pallet viewing angle is not frontal, as shown in Fig. 9b and then repeat the action modifying path based on it.

Thus, we propose a way to analyze the structure of the pallet by using the difference of integral sums, which is calculated as following:

$$f(i)=\begin{cases}\sum_{b}\sum_{j}\left|g(j,i-b)-g(j,i+b)\right|, & \forall i\in(d,c-b]\\ 0, & else\end{cases} \qquad (8)$$

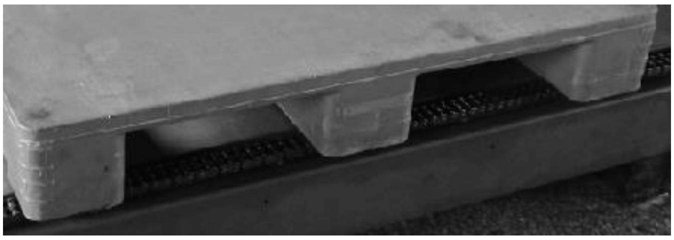
Where g(i, j) is the gray scale value of the pixel (i, j) in the image, i is variable for the column of the image, j is variable for the row of the image, b is the bandwidth variable, and c means the column number of the image.
The algorithm calculates the integral sum difference for all columns of the image from left to right of the candidate region. This approach can show clearly pallet structure in the dramatic shaded region, especially the longitudinal contour feature of the pallet candidate image (Fig. 10) relatively. The most important key to this operation is how to define the bandwidth b.
The smaller the bandwidth b is, we can get more exact high-frequency component of the image, and the larger the bandwidth b is, we can get more exact low-frequency characteristics.
In fact, this operation is a kind of Haar wavelet transform with fixed scale variables.

a) b)
Figure 9. Main Structural image of pallet a) Front image of pallet, b) image of pallet with Varying

Wavelet transform is a useful mathematical analysis technique that is widely used in image compression and multi-resolution analysis and it converts time-space signals into frequency-space signals by scale variables and shift variables. We are going to get the pallet's geometric feature that reflect the change of the dramatic shading region correctly and this feature corresponds to the high-frequency characteristic.

But if we choose the too low bandwidth b to emphasize only the high-frequency characteristics, there will be much local peaks by image noise. This is very unsuitable for geometric feature analysis. Also, we can't detect the geometric features accurately for too large bandwidth and there is a drawback that the computational cost increases.
It is important to determine suitable bandwidth b.

Here, we determined the optimal bandwidth from the pallet's size.



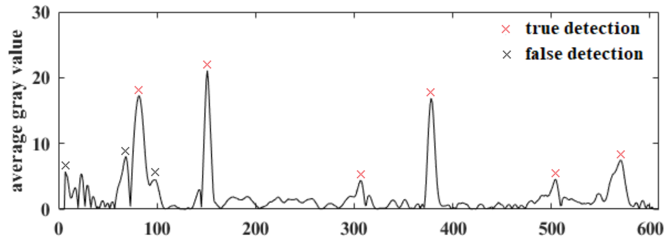


Figure 10. Difference between absolute sum

Fig 11 shows the geometric size of the pallet we use.
The whole length of the pallet from the front is 1200 mm, the support log is 150 mm, and there is no bottom support.
We have to detect the gray scale change in the both sides of pallet's middle log, bandwidth b and the support width d satisfy the following equation:

$$d-2b > 2b \Rightarrow b < d/4 \qquad (9)$$

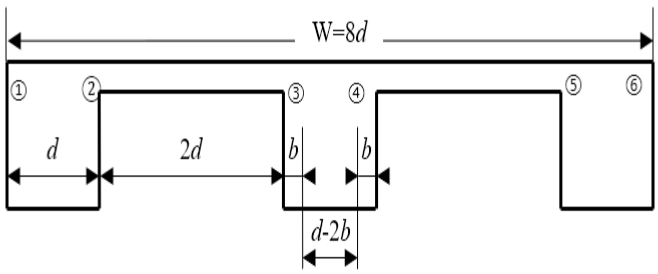


Figure 11. Geometric Size of pallet

From this equation we determine b as d/4.
On the other hand, there are six local peaks that means the edge of the pallet support logs in the integral sum difference plot as shown in Fig 10.
It is very important to select right points that reflect the geometric feature of the pallet from these six peak points to improve the detection accuracy.
Fig 10 shows the incorrect selected peaks.
The part respond to gray changes around the pallet support logs should form a peak.

© 2025 IJSRET

In our Experiment, we calculated the integral sum difference with the positive training samples.
We select the 10 maxima peaks in many peaks reflecting the transverse gray-scale change and there have to be peaks reflecting the geometric feature of the pallet in these 10 points. Therefore, we first select 10 maxima to detect the geometric feature of the pallet.
You have to be care of not to select two maximum points in the near neighboring region.

The peaks have to satisfy the following conditions for the pallet's detection:
1) The distance between the peaks reflecting both sides of the pallet support log should be almost similar.
2) The distance between the peaks reflecting the interval of the pallet support logs should be almost similar and larger than the width of the pallet support log.
This condition is described as following.

$$\frac{T_{i+3}-T_{i+2}}{T_{i+1}-T_i} \in U(1;\ \varepsilon), \quad \forall i = 1, 2, 3 \tag{9}$$

Where Ti represents the pixel value of the maximum peak in the absolute sum difference graph.
Where the variable ε reflects the flexibility of detection respect to the change of viewing angle for the pallet and is less than 0.1. When the forklift faces the pallet directly, ε is almost zero theoretically and the more view angle increases, the larger ε is.
We select 6 points in 10 ones randomly and arrange these points according to the pixel value. Then we operate according to Eq. (9) and finally complete the pallet detection

IV. EXPERIMENTS AND RESULT ANALYSIS

We simulated the pallet detection using a computer with an Intel Core i5-5300U 4 Core CPU with 2.3 GHz operating frequency and 8G RAM.
We designed the pallet detection software using python 3.6.3 and opencv3.4.2 on a Windows platform.
We used CVA50 as scanner, which has a frame rate of 25fps with a valid pixel size of 752 × 582.
We used 1830 positive samples and 3420 negative samples as a training data collected manually.
The positive samples are shown in Fig 12.
The positive samples were collected under the bright and dark environments even various illumination conditions, the illumination was controlled.

There are also samples under the varying pallet viewing angles in the range between [-15,15].
All samples were standardized to 100 × 28.
If we use rectangular feature shown in Fig 4, the number of Haar-type features is 4, 889, 808.

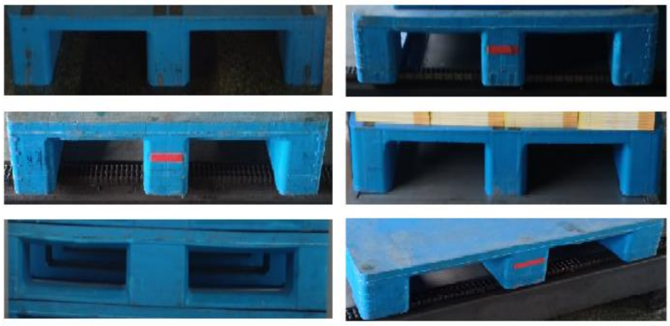


Figure 12. Positive Sample of pallets

Even though we have constructed a training dataset with a size of 32 × 14 to consider the relatively small size of the image that constitutes the sample set for generating Haar features, the number of Haar-like features are 112 and 840.
But the number of LBP is 2800 and number of our proposed features are 13832.
As you can see, the number of Haar-like features and proposed features are very different, the real-time performance of the proposed method will be satisfied enough for pallet detection.
We have defined N = 8 in Eq. (4).

Here, the number of weak classifiers that constitute the Adaboost classifier is determined as N = 169.
We did experiments in both bright and dark environments to verify the robustness of the proposed method to illumination variations.
We also did experiments in cases of loaded and unloaded pallets, and multiple pallets and under varying viewing angles.
Next, we did experiments with various plastic and wood pallets.

Table 1. Property comparison of proposed method and old method

Method	Detection rate (%)	Processing time (ms)	Luminance
Haar+ Adaboost	82.5%	42.13	Bright
MBLBP+ Adaboost	86.8%	18.14	
This work	99.3%	23.78	
Haar+ Adaboost	78.5%	42.13	Dark
MBLBP+ Adaboost	83.2%	19.28	
This work	98.2%	24.34	

Table 1 shows the detection rate while we only use Adaboost classifier and Haar-like features and while we use D-HMBLBP feature.

The experimental results in Table 1 illustrate that the proposed algorithm is reliable in complicated industrial environments, the detection error is less than 1.8%, and we can detect a pallet in real-time, especially shows high-detection rate in the dark environment.

V. CONCLUSIONS

The approach proposed here would be able to improve the pallet detection rate and satisfy the requirement of real-time performance under various conditions.

First, we used a feature descriptor, that reduces a dimension of features and also represents the pallet's features more accurately by combining Haar features and LBP features to improve the real-time performance, and distribution features obtained from distribution information of the training data and finally designed a more effective Adaboost classifier.

Second, we proposed pallet geometric analysis approach based on the geometric features of the pallet, which classify objects and non-objects with high accuracy using integral sum difference.

This can be widely used for real-time recognition of objects whose shapes are similar.

Acknowledgements

The Authors thank you Kim Mun Hyok for his valuable information and advice on the study. We hope for an objective assessment of the manuscript by editors and reviewers.

REFERENCES

1. Andrea, Motroni, Alice Buffi, Paolo Nepa. (2021). Forklift Tracking: Industry 4.0 Implementation in Large-Scale Warehouses through UWB Sensor Fusion, Advanced Sensors and Sensing Technologies for Indoor Localization. Appl. Science, 11(22), 10607.
2. Aref MM, Ghabcheloo R, Mattila J. (2014). A Macro-Micro Controller for Pallet Picking By an Articulated-Frame-Steering Hydraulic Mobile Machine. IEEE Int Conf Rob Autom, 6816-6822.
3. Guang-zhao Cui, Lin-sha Lu. (2010). A Robust Autonomous Mobile Forklift Pallet Recognition. 2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics, 286-290.
4. Junhao Xiao, Huimin Lu, Lilian Zhang, Jianhua Zhang. (2017). Pallet recognition and localization using an RGB-D camera. International Journal of Advanced Robotic Systems, DOI: 10.1177/1729881417737799, 1-10
5. Benjamin Molter, Johannes Fottner. (2018). Real-time Pallet Localization with 3D Camera Technology for Forklifts in Logistic Environments. 2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), 297-302
6. Ihab S. Mohamed, Alessio Capitanelli, Fulvio Mastrogiovanni, Stefano Rovetta, Renato Zaccaria. (2019). A 2D laser rangefinder scans dataset of standard EUR pallets. Data in brief, 24, 103837.
7. Chen G, Peng R, Wang Z, Zhao W. (2012). Pallet recognition and localization method for vision guided forklift. International Conference on Wireless Communications, Networking and Mobile Computing, 1–4.
8. M. Seelinger, J. D. Yoder. (2006). Automatic visual guidance of a forklift engaging a pallet. Robotics and Autonomous Systems, 54, 1026-1038.
9. Fengyuan Jia, Zhaosheng Tao, Fusong Wang. (2021). Wooden pallet image segmentation based on Otsu and marker watershed. Journal of Physics: Conference Series, doi:10.1088/1742-6596/1976/1/012005
10. A. Panning, A. K. Al-Hamadi, R. Niese, B. Michaelis. (2008). Facial Expression Recognition Based on Haar-Like Feature Detection. Pattern Recognition and Image Analysis, 18(3), 447–452.
11. Songyan Ma, Lu Bai. (2016). A Face Detection Algorithm Based on Adaboost and New Haar-Like Feature. Inner Mongolia Autonomous. 200362003@163.com
12. Timo Ojala, Matti Pietikäinen. (2002). Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 24(7), 971-986.
13. Robert Varga and Sergiu Nedevschi. (2016). Robust Pallet Detection for Automated Logistics Operations. In Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 4, 470-477.
14. J. Pagès, X. Armangué, J. Salvi, J. Freixenet, J. Martí. (2011). A Computer Vision System for Autonomous Forklift Vehicles in Industrial Environments. Institute of Informatics and Applications.
15. Ryosuke Iinuma, Yusuke Kojima, Hiroyuki Onoyama, Takanori Fukao, Shingo Hattori, Yasunori Nonogaki. (2020). Pallet Handling System with an Autonomous Forklift for Outdoor Fields. Journal of Robotics and Mechatronics, 32(5),1071-1078.
16. Fernando Casado, Yago Luis lapido, Diego P. Losada, Alejandro Santana-Alonso. (2017). Pose estimation and object tracking using 2D images. Procedia Manufacturing, 63–71.
17. Jia-Liang Syu, Hsin-Ting Li, Jen-Shiun Chiang, Chih-Hsien Hsia, Po-Han Wu, Chi-Fang Hsieh, Shih-An Li. (2016). A computer vision assisted system for autonomous forklift vehicles in real factory environment, Springer Science + Business Media New York 2016.
18. Qi Song. Pallet detection and localization with RGB image and depth data using deep learning techniques. (2023)
19. [Yiping Shao, Zhengshuai Fan, Baochang Zhu, Jiansha Lu, Yiding Lang. (2023). A Point Cloud Data-Driven Pallet

© 2025 IJSRET

Pose Estimation Method Using an Active Binocular Vision Sensor. Sensors (Basel). 23(3): 1217.
20. Michela Zaccaria, Riccardo Monica, Jacopo Aleotti. (2020). A Comparison of Deep Learning Models for Pallet Detection in Industrial Warehouses. IEEE Xplore. Restrictions apply. 417-422
21. L. Zhang, R. Chu, S. Xiang, S. Liao, S. Li. (2007). Face detection based on multi-block LBP representation. 2nd International Conference on Biometrics, 11–18.
22. S.K. Pavani, D. Delgado, A.F. Frangi. (2009). Haar-like features with optimally weighted rectangles for rapid object detection. Pattern Recognition 43, 160–172.
23. Songyan Ma, Lu Bai. (2016). A Face Detection Algorithm Based on Adaboost and New Haar-Like Feature, 651-654.
24. Viola P, Jones M. (2001), Rapid object detection using a boosted cascade of simple features. IEEE Conference on Computer Vision and Pattern Recognition, 511-518.
25. Qiu Qin-jun, Liu Yong, Cai Da-wei. (2014). Vehicle detection based on LBP features of the Haar-like Characteristics. Proceeding of the 11th World Congress on Intelligent Control and Automation, 1050-1055.
26. Perreault, S. and Hébert, P. (2007). Median filtering in constant time. IEEE Trans. on Image Processing, 16(9), 2389–2394.

Author's details
1 Chung Hyok Pak, Faculty of Automation Engineering, Kim Chaek University of Technology, Pyongyang, Democratic People's Republic of Korea, pch87929@star-co.net.kp
2 Un Sim Ri, Faculty of Automation Engineering, Kim Chaek University of Technology, Pyongyang, Democratic People's Republic of Korea, rus89115@star-co.net.kp
3 Se Hyon Kim, Faculty of Automation Engineering, Kim Chaek University of Technology, Pyongyang, Democratic People's Republic of Korea, ksh921225@star-co.net.kp

© 2025 IJSRET