Empirical Performance Evaluation of Object Recognition Methods

Markus Ulrich[1,2] and Carsten Steger[1]

[1]MVTec Software GmbH
Neherstraße 1, 81675 München, Germany
Ph.: +49-89-457695-0, Fax: +49-89-457695-55
URL: www.mvtec.com, e-mail: {ulrich,steger}@mvtec.com

[2]Chair for Photogrammetry and Remote Sensing, Technische Universität München
Arcisstraße 21, 80290 München, Germany
Ph.: +49-89-289-22671, Fax: +49-89-280-9573
URL: www.photo.verm.tu-muenchen.de, e-mail: ulrich@bv.tu-muenchen.de

Abstract

We propose an empirical performance evaluation of five different object recognition methods. For this purpose, the normalized cross correlation and the sum of absolute differences as two standard similarity measures in industrial applications are compared to the Hausdorff distance and two novel recognition methods that we developed with the aim to fulfil increasing industrial demands. After a description of the respective methods, several criteria are introduced that allow an objective evaluation of object recognition approaches. Experiments on real images are used to apply the proposed criteria. The experimental set-up used for the evaluation measurements is explained in detail. The results are illustrated and analyzed extensively. It is shown that our novel recognition approaches perform substantially better than the existing approaches.

1. Introduction

Object recognition is used in many computer vision applications. It is particularly useful for industrial inspection tasks, where often an image of an object must be aligned with a model of the object. The transformation (pose) obtained by the object recognition process can be used for various tasks, e.g., pick and place operations, quality control, or inspection tasks. In most cases, the model of the object is generated from an image of the object. This 2D approach is taken because it usually is too costly or time consuming to create a more complicated model, e.g., a 3D CAD model. Therefore, in industrial inspection tasks one is usually interested in matching a 2D model of an object to the image. The object may be transformed by a certain class of transformations, e.g., rigid transformations, similarity transformations, or general 2D affine transformations (which are usually taken as an approximation to the true perspective transformations an object may undergo).

A large number of object recognition strategies exist. The approaches to object recognition examined in this paper use pixels as their geometric features, i.e., not higher level features like lines or elliptic arcs. Therefore, in the following only similar pixel-based strategies will be reviewed.

Several methods have been proposed to recognize objects in images by matching 2D models to images. A survey of matching approaches is given in [3]. In most 2D matching approaches the model is systematically compared to the image using all allowable degrees of freedom of the chosen class of transformations. The comparison is based on a suitable similarity measure (also called match metric). The maxima or minima of the similarity measure are used to decide whether an object is present in the image and to determine its pose. To speed up the recognition process, the search is usually done in a coarse-to-fine manner, e.g., by using image pyramids [12].

The simplest class of object recognition methods is based on the gray values of the model and image itself and uses normalized cross correlation or the sum of squared or absolute differences as a similarity measure [3]. Normalized cross correlation is invariant to linear brightness changes but is very sensitive to clutter and occlusion as well

as nonlinear contrast changes. The sum of gray value differences is not robust to any of these changes, but can be made robust to linear brightness changes by explicitly incorporating them into the similarity measure, and to a moderate amount of occlusion and clutter by computing the similarity measure in a statistically robust manner [6].

A more complex class of object recognition methods does not use the gray values of the model or object itself, but uses the object's edges for matching ([2], [9]). In all existing approaches, the edges are segmented, i.e., a binary image is computed for both the model and the search image. Usually, the edge pixels are defined as the pixels in the image where the magnitude of the gradient is maximum in the direction of the gradient. Various similarity measures can then be used to compare the model to the image. The similarity measure in [2] computes the average distance of the model edges and the image edges. The disadvantage of this similarity measure is that it is not robust to occlusions because the distance to the nearest edge increases significantly if some of the edges of the model are missing in the image.

The Hausdorff distance similarity measure used in [9] tries to remedy this shortcoming by calculating the maximum of the k-th largest distance of the model edges to the image edges and the l-th largest distance of the image edges and the model edges. If the model contains n points and the image contains m edge points, the similarity measure is robust to $100k/n\%$ occlusion and $100l/m\%$ clutter. Unfortunately, an estimate for m is needed to determine l, which is usually not available.

All of these similarity measures have the disadvantage that they do not take into account the direction of the edges. In [8] it is shown that disregarding the edge direction information leads to false positive instances of the model in the image. The similarity measure proposed in [8] tries to improve this by modifying the Hausdorff distance to also measure the angle difference between the model and image edges. Unfortunately, the implementation is based on multiple distance transformations, which makes the algorithm too computationally expensive for industrial inspection.

Finally, another class of edge based object recognition algorithms is based on the generalized Hough transform [1]. Approaches of this kind have the advantage that they are robust to occlusion as well as clutter. Unfortunately, the GHT in the conventional form requires large amounts of memory and long computation time to recognize the object.

In all of the above approaches, the edge image is binarized. This makes the object recognition algorithm invariant only against a narrow range of illumination changes. If the image contrast is lowered, progressively fewer edge points will be segmented, which has the same effects as progressively larger occlusion.

In this paper three of the above mentioned approaches are analyzed and their performance is compared to those of our new approaches. The analysis of the performance characteristics of object recognition methods is very important. First, it makes an algorithm comparable to other algorithms, thus helping users in selecting the appropriate method for the task they have to solve. Second, it helps to identify breakdown points of the algorithm, i.e., areas where the algorithm cannot be used because some of the assumptions it makes are violated. Therefore, in this paper an attempt is made to characterize the performance of five selected different object recognition approaches: The first two methods to be analyzed are the *Normalized Cross Correlation* [3] and the *Sum of Absolute Differences*, because they are rather wide spread methods in industry and therefore well known in the application area of object recognition. The *Hausdorff Distance* [9] is the third candidate, which is also the core of many recognition implementations, because of its higher robustness against occlusions and clutter in contrast to the sum of absolute differences and even to the normalized cross correlation. Additionally, two novel approaches, which are referred to as *Shape-Based Matching* [11] and *Modified Hough Transform* [13, 14, 15] below, are included in our analysis. The development of the latter two approaches was motivated by the increasing industrial demands like real-time computation and high recognition accuracy. Therefore, the study is mainly concerned with the robustness, the subpixel accuracy, and the required computation time of the five candidate algorithms under different external circumstances.

The paper is organized as follows. In section 2 the mentioned recognition methods to be analyzed are introduced. Since the normalized cross correlation, the sum of absolute differences and the Hausdorff distance are standard computer vision techniques, the main focus is set to our two novel approaches. The performance evaluation is presented in section 3, which includes the description of the evaluation criteria and the employed experimental set-up as well as the analysis of the obtained results. The conclusions in section 4 complete this study.

2. Evaluated Object Recognition Methods

First of all we want to introduce some definitions that facilitate the comparison between the five techniques. All recognition methods have in common that they require some form of representation of the object to be found, which will be called *model* below. The model M is generated from an image of the object to be recognized, called *reference image* I^{r}. An arbitrary region of interest (*ROI*) R specifies the object's location in the image. The image, in which the object should be recognized, will be referred to as the *search image* I^{s}. Almost all object recognition approaches can be split into two successive phases: the *offline phase* including

the generation of the model and the *online phase*, in which the constructed model is used to find the object in the search image.

The *transformation class* $\mathcal{T}$, e.g., translations or Euclidean, similarity, affine, or arbitrary projective transformations, specifies the degrees of freedom of the object, i.e., which transformations the object may undergo in the search image. For all similarity measures the object recognition step is performed by transforming the model to a user-limited range of discrete transformations $T_i \in \mathcal{T}$ within the transformation class. For each transformed model $M_i^{\mathrm{t}} = T_i M$ the similarity measure is calculated between M^{t} and the corresponding representation of the search image. The representation can, for example, be described by the raw gray values in both images (e.g., when using the normalized cross correlation or the sum of absolute differences) or by the corresponding binarized edges (e.g., when using the Hausdorff distance). The maximum or minimum of the match metric then indicates the pose of the recognized object.

2.1. Normalized Cross Correlation

The normalized cross correlation C at position (x, y) in I^s is computed as follows:

$$\begin{aligned} C(x,y) &= \frac{1}{n} \sum_{(u,v)\in R} \left(\frac{I^{\mathrm{r}}(u,v) - \mu_r}{\sigma_r} \right. \\ &\quad \left. \cdot \frac{I^{\mathrm{s}}(x+u, y+v) - \mu_s(x,y)}{\sigma_s(x,y)} \right) . \end{aligned} \quad (1)$$

Here, μ and σ are the mean and the standard deviation of the gray values in the reference and the search image, respectively, and n specifies the number of points in R. The normalization causes the cross correlation to be unaffected by linear brightness changes in the search image (see [3]).

For the purpose of evaluating the performance of the normalized cross correlation we use — as one typical representative — the current implementation of the *Matrox Imaging Library* (MIL), which is a software development toolkit of *Matrox Electonic Systems Ltd* [7]. Some specific implementation characteristics should be explained to ensure the correct appraise of the evaluation results: The pattern matching algorithm is able to find a predefined object under rigid motion, i.e., translation and rotation. A hierarchical search strategy using image pyramids is used to speed up the recognition. The quality of the match is returned by calculating a match score value as $\mathrm{Score} = max(C, 0)^2 \cdot 100\%$. The subpixel accuracy of the object position is achieved by a surface fit to the match scores around the peak. Thus, the exact peak position can be calculated from the equation of the surface. The refinement of the object orientation is not comprehensively explained in the documentation, but we suppose the refinement of the obtained discrete object orientation is realized by a finer resampling of the orientation in the angle neighborhood of the maximum score and recalculating the cross correlation for each refined angle.

2.2. Sum of Absolute Differences

The similarity measure based on the sum of absolute differences D [3] at $I^s(x, y)$ is calculated by:

$$\begin{aligned} \mathrm{Error}(x,y) &= D(x,y) \\ &= \frac{1}{n} \sum_{(u,v)\in R} \left| I^{\mathrm{s}}(x-u, y-v) - I^{\mathrm{r}}(u,v) \right| . \end{aligned} \quad (2)$$

Thus, D indicates the average difference of the grayvalues. Since the sum of absolute differences is a measure of dissimilarity, we also denote D as Error.

The special implementation we investigate considers rigid motion and makes use of image pyramids to speed up the recognition process. The position and orientation of the best match, i.e., the match with smallest Error, is returned. Additionally, the user is able to specify the maximum average error of the match. The lower this value is, the faster the operator runs, since fewer matches must be traced down the pyramid. Subpixel accuracy of position and the refinement of the discrete orientation are calculated by extrapolating the minimum of D.

2.3. Hausdorff Distance

The Hausdorff distance [9] measures the extent to which each pixel of the binarized reference image lies near some pixel of the binarized search image and vice versa. The binarization can be done by a number of different image processing algorithms, e.g., edge detectors like Sobel or Canny [4] operators, line detectors [10], or corner detectors [5]. To be able to compare the results of the Hausdorff distance to the shape-based similarity measure and the modified Hough transform the Sobel filter is used, which is also used in our two approaches. To reduce the sensitivity to outliers, the symmetric partial undirected Hausdorff distance is used. Let P^r and P^s be the two point sets obtained in the reference image and in the search image respectively. The symmetric partial undirected Hausdorff distance is then computed by

$$H^{f_F f_R}(P^r, P^s) = \max(h^{f_F}(P^r, P^s), h^{f_R}(P^s, P^r)) \quad (3)$$

$$h^{f_F}(P^r, P^s) = \underset{p_i^r \in P^r}{f\mathrm{th}} \min_{p_i^s \in P^s} \|p_i^r - p_i^s\| \quad (4)$$

$$h^{f_R}(P^s, P^r) = \underset{p_i^s \in P^s}{f\mathrm{th}} \min_{p_i^r \in P^r} \|p_i^s - p_i^r\| \ . \quad (5)$$

where f_F and f_R are called the *forward fraction* and the *reverse fraction* and $f\mathrm{th}$ denotes the f-th largest value.

Due to the lack of time we did not implement the Hausdorff distance by ourselves but use the original implementation by Rucklidge [9]. The program expects the forward and the reverse fraction as well as the thresholds for the forward and the reverse distance as input data. Since the method of [9] returns all matches that fulfill its score and distance criteria, the best match was selected based on the minimum forward distance. If more than one match had the same minimum forward distance, the match with the maximum forward fraction was selected as the best match. Only translations of the object can be recognized and no subpixel refinement is included. Although the parameter space is treated in a hierarchical way there is no use of image pyramids, which makes the algorithm very slow.

2.4. Shape-Based Matching

In this section the principle of our novel similarity measure is briefly explained. A detailed description can be found in [11].

Here, the model consists of a set of points $p_i = (x_i, y_i)^T$ with a corresponding direction vector $d_i = (t_i, u_i)^T$, $i = 1, \ldots, n$. The direction vectors can be generated by a number of different image processing operations, e.g., edge, line, or corner extraction.

The search image can be transformed into a representation in which a direction vector $d_{x,y} = (v_{x,y}, w_{x,y})^T$ is obtained for each image point (x, y). In the matching process, a transformed model must be compared to the image at a particular location by a similarity measure. We suggest to sum the normalized dot product of the direction vectors of the transformed model and the search image over all points of the model to compute a matching score at a particular point $q = (x, y)^T$ of the image. If the model is generated by edge or line filtering, and the image is preprocessed in the same manner, this similarity measure fulfills the requirements of robustness to occlusion and clutter. If a user specifies a threshold on the similarity measure to determine whether the model is present in the image, a similarity measure with a well defined range of values is desirable. The following similarity measure achieves this goal:

$$\begin{aligned} s &= \frac{1}{n}\sum_{i=1}^{n}\frac{\langle d'_i, e_{q+p'}\rangle}{\|d'_i\| \cdot \|e_{q+p'}\|} \qquad (6)\\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{t'_i v_{x+x'_i,y+y'_i} + u'_i w_{x+x'_i,y+y'_i}}{\sqrt{t'^2_i + u'^2_i}\cdot\sqrt{v^2_{x+x'_i,y+y'_i} + w^2_{x+x'_i,y+y'_i}}} \ . \end{aligned}$$

Because of the normalization of the direction vectors, this similarity measure is invariant to arbitrary illumination changes. What makes this measure robust against occlusion and clutter is the fact that if a feature is missing, either in the model or in the image, noise will lead to random direction vectors, which, on average, will contribute nothing to the sum.

The normalized similarity measure (6) has the property that it returns a number smaller than 1 as the score of a potential match. A score of 1 indicates a perfect match between the model and the image. Furthermore, the score roughly corresponds to the portion of the model that is visible in the image.

A desirable feature of this similarity measure is that it does not need to be evaluated completely when object recognition is based on a threshold $s_{\min}$ for the similarity measure that a potential match must achieve. Let s_j denote the partial sum of the dot products up to the j-th element of the model:

$$s_j = \frac{1}{n}\sum_{i=1}^{j}\frac{\langle d'_i, e_{q+p'}\rangle}{\|d'_i\| \cdot \|e_{q+p'}\|} \ . \qquad (7)$$

Obviously, all the remaining terms of the sum are all ≤ 1. Therefore, the partial score can never achieve the required score $s_{\min}$ if $s_j < s_{\min} - 1 + j/n$, and hence the evaluation of the sum can be discontinued after the j-th element whenever this condition is fulfilled. This criterion speeds up the recognition process considerably.

Nevertheless, further speed-ups are highly desirable. Another criterion is to require that all partial sums have a score better than $s_{\min}$, i.e., $s_j \geq s_{\min}$. When this criterion is used, the search will be very fast, but it can no longer be ensured that the object recognition finds the correct instances of the model because if missing parts of the model are checked first, the partial score will be below the required score. To speed up the recognition process with a very low probability of not finding the object although it is visible in the image, the following heuristic can be used: the first part of the model points is examined with a relatively safe stopping criterion, while the remaining part of the model points are examined with the hard threshold $s_{\min}$. The user can specify what fraction of the model points is examined with the hard threshold with a parameter g, which will be called *greediness* below. If $g = 1$, all points are examined with the hard threshold, while for $g = 0$, all points are examined with the safe stopping criterion. With this, the evaluation of the partial sums is stopped whenever $s_j < \min(s_{\min} - 1 + fj/n, s_{\min})$, where $f = (1 - gs_{\min})/(1 - s_{\min})$.

Our current implementation is able to recognize objects under similarity transformations. To speed up the recognition process, the model is generated in multiple resolution levels, which are constructed by building an image pyramid from the original image. Because of runtime considerations the Sobel filter is used for feature extraction.

In the online phase an image pyramid is constructed for the search image. For each level of the pyramid, the same filtering operation that was used to generate the model is applied to the search image. This returns a direction vector for each image point. Note that the image is not segmented, i.e., thresholding or other operations are not performed. This results in true robustness to illumination changes.

To identify potential matches, an exhaustive search is performed for the top level of the pyramid. With the termination criteria using the threshold $s_{\min}$, this seemingly brute-force strategy actually becomes extremely efficient. After the potential matches have been identified, they are tracked through the resolution hierarchy until they are found at the lowest level of the image pyramid.

Once the object has been recognized on the lowest level of the image pyramid, its position, rotation, and scale are extracted to a resolution better than the discretization of the search space by fitting a second order polynomial (in the four pose variables) to the similarity measure values in a $3 \times 3 \times 3 \times 3$ neighborhood around the maximum score.

2.5. Modified Hough Transform

One weakness of the Generalized Hough Transform (GHT) [1] algorithm is the — in general — huge parameter space. This requires large amounts of memory to store the accumulator array as well as high computational costs in the online phase caused by the initialization of the array, the incrementation, and the search for maxima after the incrementation step. In this section we introduce our novel approach which is able to recognize objects under translation and rotation: A hierarchical search strategy in combination with an effective limitation of the search space is introduced. Further details can be found in [13], [14], and [15] .

To reduce the size of the accumulator array and to speed up the online phase both the model and the search image are treated in a hierarchical manner using image pyramids. To build the *R*-table the edges and the corresponding gradient directions are computed using the Sobel filter.

In the online phase the recognition process starts on the top pyramid level without any à priori information about the searched transformation parameters x, y and θ using the conventional GHT. The cells in the accumulator array that are local maxima and exceed a certain threshold are stored and used to initialize approximate values on the lower levels.

At lower levels approximate values $\tilde{x}$, $\tilde{y}$, and $\tilde{\theta}$ are known from a higher level. To restrict the edge extraction region and to minimize the incrementation effort, only those pixels in the search image that lie beneath the *blurred region* are taken into account. The blurred region is defined by dilating and rotating the shape, i.e., the edge regions, corresponding to the error of the approximate values and translating it to the approximate parameter values. The blurred regions are calculated for every quantized orientation in the offline phase and stored together with the model. By this, the size of the accumulator array can be narrowed to a size corresponding to the uncertainties of the à priori parameters, which decreases the memory amount drastically.

After the edge extraction another improvement is utilized. The principle of the conventional GHT is shown in Figure 1. The à priori information is displayed as a dark gray box representing the maximum error of the approximate translation values from the next higher level. The edge pixels $\boldsymbol{p}_1^{\mathbf{r}}$, $\boldsymbol{p}_2^{\mathbf{r}}$, and $\boldsymbol{p}_3^{\mathbf{r}}$ have identical gradient directions. Thus, if any of these edge pixels is processed in the online phase of the conventional GHT each of the three vectors $\boldsymbol{r_1}$, $\boldsymbol{r_2}$, and $\boldsymbol{r_3}$ is added and the corresponding three cells are incremented, which is inefficient. Our solution greatly improves the efficiency: Already in the offline phase the information about the position of the edge pixels relative to the centroid is calculated and stored in the model. This is done by overlaying a grid structure over the rotated reference image and splitting the image into tiles (Fig. 2). In the online phase the current tile is calculated using the approximate translation parameters and the position of the current edge pixel. Only the vectors in the respective tile with the appropriate gradient direction are used to calculate the incrementation cells.

Furthermore, the inherent quantization problems of the GHT are solved by smoothing the accumulator array and establishing overlapping gradient intervals in the *R*-table.

To refine the parameters of position and orientation we use the three dimensional equivalent approach as for the shape-based matching described in section 2.4. To evaluate the quality of a match, a score value is computed as the peak height in the accumulator array divided by the number of model points.

3. Evaluation

3.1. Evaluation Criteria

We use three main criteria to evaluate the performance of the five object recognition methods and to build a common basis which allows an objective comparison.

The first criterion to be considered is the *robustness* of the approach. This includes the robustness against occlusions, which often occur in industrial applications, e.g., caused by overlapping objects on the assembly line or severe defects of the objects to be inspected. Non-linear as well as local illumination changes are also crucial situations, which cannot be avoided in many applications over the entire field of view. Therefore, the robustness against

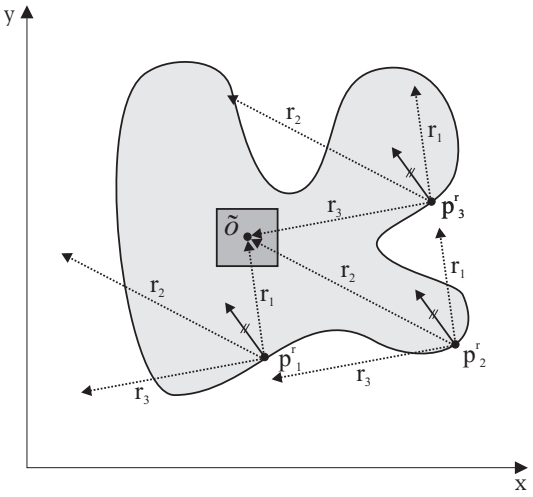


Figure 1. The conventional GHT without using the à priori information of the translation parameters. Many unnecessary increments are executed.

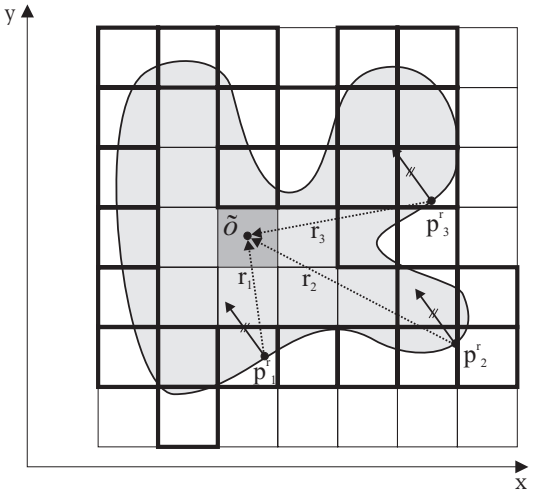


Figure 2. Taking advantage of the à priori information using a tile structure. For each occupied tile (bold border) a separate *R*-table is generated.

arbitrary illumination changes is also examined. A multitude of images were taken to simulate different overlapping and illumination situations (see section 3.2). We measure the robustness using the recognition rate that is defined by the number of images in which the object was correctly recognized divided by the total number of images.

The second criterion affects the *accuracy* of the methods. Most applications need the exact transformation parameters of the object as input for further investigations like precise metric measurements. In the area of quality control, in addition, the object in the search image must be precisely aligned with the transformed reference image to ensure a reliable recognition of defects or other variations that influence certain quality criteria, e.g., by subtracting the gray values of both images. We determine the subpixel accuracy by comparing the exact (known) position and orientation of the object with returned parameters of the different candidates.

The *computation time* represents the third evaluation criterion. Despite the increasing computation power efficient and fast algorithms are more important than ever. This is particularly true in the field of object recognition, where a multitude of applications enforce real time computation. Indeed, it is very hard to compare different recognition methods using this criterion because the computation time strongly depends on the individual implementation of the recognition methods. Nevertheless, we tried to find parameter constellations (see section 3.2) for each of the investigated approaches that at least allow a qualitative comparison.

Since the Hausdorff distance does not return the object position in subpixel accuracy and in addition does not use image pyramids resulting in unreasonably long recognition times, the criteria of accuracy and computation time are only applied to the four remaining candidates.

3.2. Experimental Set-Up

In this section the experimental set-up for the evaluation is explained in detail. We chose an IC, which is shown in Figure 3, as the object to be found in the subsequent experiments. Only the part within the bounding box of the print on the IC formes the ROI, from which the models of the different recognition approaches are created (see Figure 4). For the recognition methods that segment edges during model creation (Hausdorff distance, shape-based matching, modified Hough transform) the threshold for the minimum edge amplitude in the reference image was set to 30 during all our experiments. The images we used for the evaluation are 8 bit gray scale of size 652×494 pixels. For all recognition methods we used four pyramid levels except for the implementation of the Hausdorff distance, which does not support a pyramid approach.

3.2.1. Robustness

To apply the first criterion of robustness and determine the recognition rate two image sequences were taken, one for testing the robustness against occlusions the other for testing the sensibility against illumination changes. We defined the recognition rate as the number of images, in which the object was recognized at the correct position divided by the total number of images.

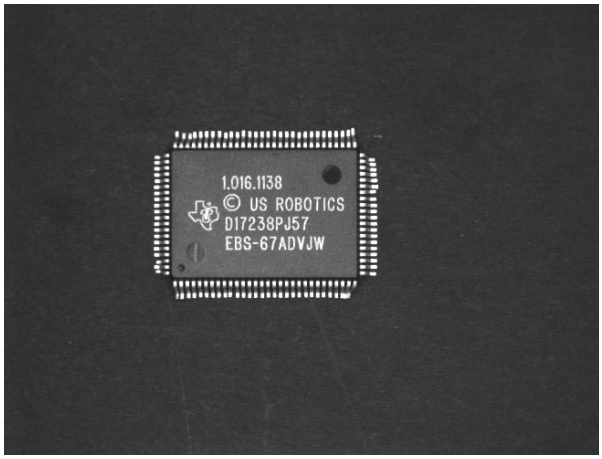


Figure 3. An IC is used as the object to be recognized.

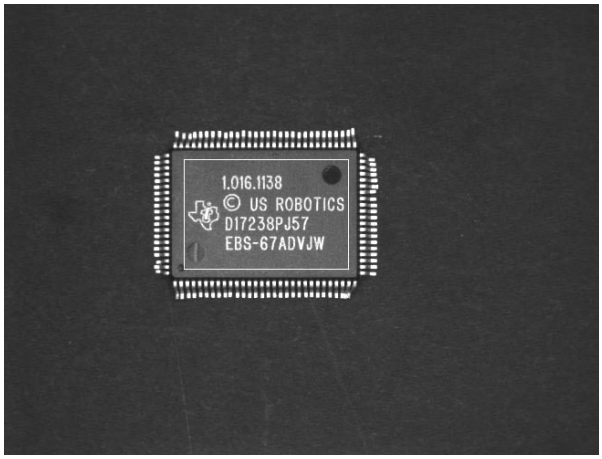


Figure 4. The model is created from the print of the IC using a rectangular ROI.

The first sequence contains 500 images of the IC, which was occluded to various degrees with various objects, so that in addition to occlusion, clutter of various degrees was created in the image. Figure 5 shows six of the 500 images that we used to test the robustness against occlusion.

The size of the bounding box is 180×120 pixels at the lowest pyramid level, i.e., at original image resolution, containing 2127 edge points extracted by the Sobel filter. Beside the recognition rate, in addition, the correlation between the actual occlusion and the returned score values are examined. For this purpose an effort was made to keep the IC in exactly the same position in the image in order to be able to measure the degree of occlusion. Unfortunately, the IC moved very slightly (by less than one pixel) during the acquisition of the images. The true amount of occlusion was determined by extracting edges from the images and intersecting the edge region with the edges within the ROI in the reference image. Since the objects that occlude the IC generate clutter edges, this actually underestimates the occlusion.

The transformation class was restricted to translations, to reduce the time required to execute the experiment. However, the allowable range of the translation parameters was not restricted, i.e., the object is searched in the whole image. For the normalized cross correlation, the Hausdorff distance, the shape-based matching approach, and the modified Hough transform, different values for the parameter of the minimum score were applied. The forward fraction of the Hausdorff distance was interpreted as score value. Initial tests with the forward and reverse fractions set to 30% resulted in run times of more than three hours per image. Therefore, the reverse fraction was set to 50% and the forward fraction was successively increased from 50% to 90% using an increment of 10%. The parameter for the maximum forward and reverse distance were set to 1. For the other three approaches the minimum score was varied from 10 to 90 percent. In the case of the sum of the absolute differences the maximum error instead of the minimum score was varied. We limited this range to a maximum error of 30. Tolerating higher values was also too computationally expensive. As explained in section 2.4 the recognition rate of the shape-based matching approach depends on the parameter greediness. Therefore, we additionally varied this parameter in the range of 0 to 1 using increments of 0.2.

To test the robustness against arbitrary illumination changes a second sequence of images of the IC was taken, which includes various illumination situations. Three example situations are displayed in Figure 6. Due to a smaller distance between the IC and the camera, the ROI is now 255×140 pixels containing 3381 model points on the lowest pyramid level. The parameter settings for the five methods is equivalent to the settings for testing the robustness against occlusions. Since the modified Hough transform segments the search image, additionally the threshold for the minimum edge amplitude in the online phase is varied from 5 to 30 using an increment of 5.

3.2.2. Accuracy

In this section the experimental set-up, which we used to specify the accuracy of the algorithms, is explained. This criterion is not applied to the Hausdorff distance, because subpixel accuracy is not achieved by the used implementation. Generally, it seems to be very difficult to compute a refinement of the returned parameters directly based on the forward or reverse fraction. Since the shape-based matching is the only candidate that is able to recognize scaled objects, only the position and orientation accuracy of the four approaches are tested.

To test the accuracy, the IC was mounted onto a table

Figure 5. Six of the 500 images that were used to test the robustness against occlusions.

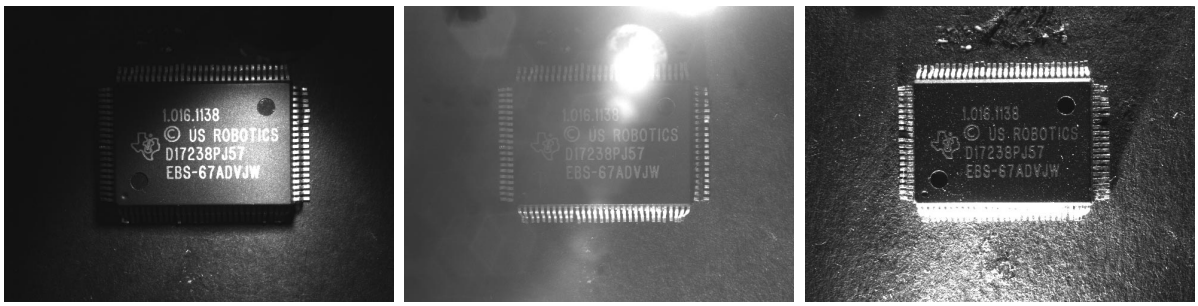


Figure 6. Three of the 200 images that were used to test the robustness against arbitrary illumination changes.

that can be shifted with an accuracy of $1\mu m$ and can be rotated with an accuracy of $0.7'$ ($0.011667°$). Three image sequences were acquired: In the first sequence, the IC was shifted in $10\mu m$ increments to the left in the horizontal direction, which resulted in shifts of about 1/7 pixel in the image. A total of 40 shifts were performed, while 10 images were taken for each position of the object. The IC was not occluded in this experiment and the illumination was not changed. In the second sequence, the IC was shifted in the vertical direction with upward movement in the same way. However, a total of 50 shifts were performed. The intention of the third sequence was to test the accuracy of the returned object orientation. For this purpose, the IC was rotated 50 times for a total of $5.83°$. Again, 10 images were taken in every orientation.

During all accuracy tests Euclidean motion was used as transformation class. The search angle for all approaches was restricted to the range of $[-30°;+30°]$, whereas the range of translation parameters again was not restricted. The increment of the quantized orientation step was set to $1°$, which results in the models containing 61 rotated instances of the template image at the lowest pyramid level. Since no occlusions were present the threshold for the minimum score could be uniformly set to 80% for all approaches. For the maximum error of the sum of absolute differences we used a value of 25. Lower values resulted in missed objects, if the IC was rotated in the middle between two quantized orientations, e.g. $0.5°$, $1.5°$, etc., whereas higher values resulted in expensive computations. In shape-based matching the greediness parameter was set to 0.5, which represents a good compromise between recognition rate and computation time.

3.2.3. Computational Time

In order to apply the third criterion, exactly the same configuration was employed as it was used for the accuracy test described in section 3.2.2. The computation time of the recognition processes was measured on a 400 MHz Pentium II for each image of the three sequences and for each recognition method (excluding again the Hausdorff distance for the reason mentioned above). In order to assess the correlation between restriction of parameter space and computation time, additionally, a second run was performed without restricting the angle interval.

In this context it should be noted that the modified Hough transform is the only candidate that is able to recognize the object, even if it partially lies outside the search image. The translation range of the other approaches is restricted automatically to the positions at which the object lies completely in the search image. Particularly in the case of large objects this results in an unfair comparison between the Hough transform and the other candidates, when considering computation time.

3.3. Results

In this section we present the results of the experiments described in section 3.2. Several plots illustrate the performance of the examined recognition methods. The description and the analyses of the plots are structured as in the previous section, i.e., at first the results of the robustness, then the accuracy, and finally the computation time are presented.

3.3.1. Robustness

At first, the sequence of the occluded IC was tested. Figure 7 shows the recognition rate of the shape-based matching approach depending on the minimum score and the greediness parameter. As expected, the number of correctly recognized objects decreases with increasing minimum score, i.e., the higher the degree of occlusion the smaller the parameter of the minimum score must be chosen to find the object. What also can be seen from this figure is that apparently the greediness parameter must be adjusted very carefully when dealing with occluded objects: For the given minimum score of 50% the recognition rate varies in the range between 48% and 82% corresponding to the two extreme greediness values of 1 and 0. On the other hand decreasing the greediness parameter from 1, which would never be chosen in practice, to 0.8 already improves the recognition rate to 64%. When we look at the curve corresponding to the greediness value 0, we can see that if the minimum score was chosen small enough the object was found in almost all images . The correlation between the score and the visibility of the object is studied beneath.

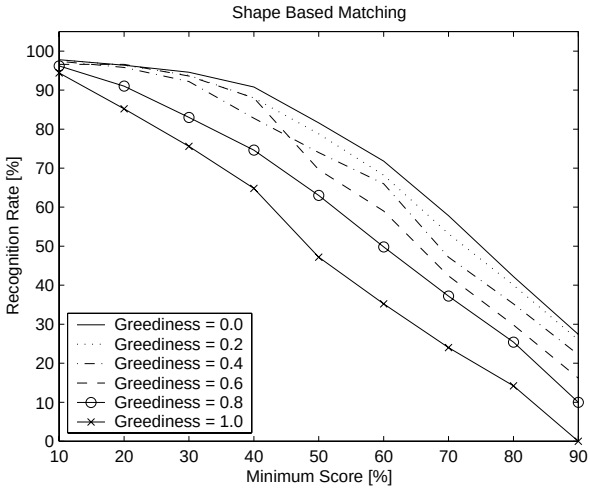
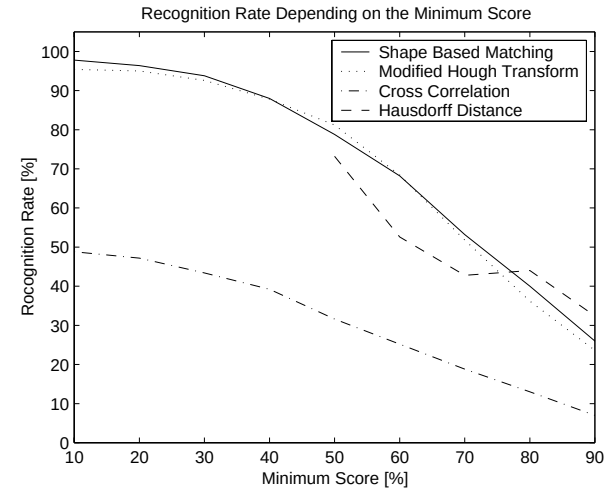


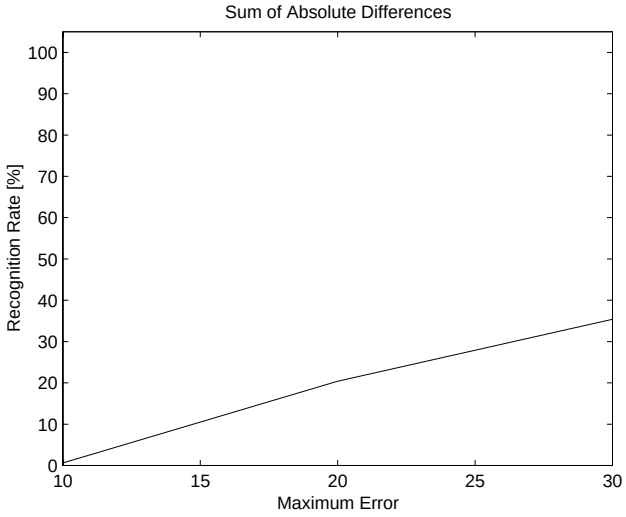
Figure 7. Recognition rate in the case of occlusions of shape-based matching using different values for the parameter greediness. The "greedier" the search the more matches are missed.

A complete comparison of all approaches concerning the robustness against occlusion is shown in Figure 8. Here, the superiority of our two novel approaches becomes clear. Note that the robustness of the modified Hough transform hardly differs from the robustness achieved by the shape-based matching when using a greediness of 0. Looking at the other approaches, only the Hausdorff distance reaches a comparable result, which is, however, inferior in most cases. For example, when using a threshold for the minimum forward fraction of 50% the recognition rate is about 72%, i.e., the shape-based matching performed 10% better and the modified Hough transform — with a corresponding recognition rate of 83% — 11% better than the method using the Hausdorff distance. The recognition rate of the normalized cross correlation does not reach 50% at all, even if the minimum score is chosen small. The approach using the sum of absolute differences shows a similar behavior. Although the expectation is fulfilled that the robustness increases when the maximum error is set to a higher value, even relatively high values for the mean maximum error (e.g., 30) only lead to a small recognition rate of about 35%. A further increase of the maximum error is not reasonable because of two reasons: first, the computation time would make the algorithm unsuitable for practical use and second, this would lead to many false positives, i.e., an occluded instance could not be distinguished from clutter in the search image.

Figure 9 displays a plot of the extracted score against the estimated visibility of the object. The instances in which the model was not found are denoted by a score of 0, i.e., they lie on the x axis of the plot. For the sum of absolute differ-



(a)



(b)

Figure 8. The recognition rate of different approaches indicates the robustness against occlusions. This figure shows the recognition rate (a) of four candidates depending on the minimum score and (b) of the approach using the sum of the absolute differences depending on the maximum error.

ences again the error is plotted instead of the score. What can be seen is that the error is negatively correlated with the visibility. Nevertheless the points are widely spread and not close to the ideal virtual line with negative gradient. In addition, despite of a very high degree of visibility many objects were not recognized. One possible reason for this behavior could be that in some images the clutter object does not occlude the IC yet, but throws its shadow on the IC, which strongly influences this metric.

In the plot of the Hausdorff distance the wrong matches either have a forward fraction of 0% or close to 50%. Here, a noticeable positive correlation can be observed, but several objects with a visibility of far beyond 50% could not be recognized. This explains the lower recognition rate in comparison to our approaches, which was mentioned above.

The normalized cross correlation also shows positive correlation but similar to the sum of absolute differences the points in the plot are widely spread and many objects with high visibility were not recognized.

In contrast, the plots of our new approaches show a point distribution that is much closer to the ideal: The positive correlation is conspicuous and the points lie close to a fitted line, the gradient of which is close to 1. In addition, objects with high visibility are recognized with a high probability.

Another very interesting plot is shown in Figure 10, in which the position error of the five approaches under the stringent condition of occlusion is described. The reason why this analysis is shown in this section instead of section 3.3.2 is that robustness against occlusion not only means to recognize the object at all, but to find it at the correct position. It can be seen that the IC was accidentally shifted twice. The position errors are all very close to the three cluster centers except for the Hausdorff distance, for which in some instances the best match was more than 30 pixels from the true location. Concerning the other approaches some of the larger errors in the y coordinate result from refraction effects caused by the transparent ruler that was used in some images to occlude the IC.

In the following the robustness against arbitrary illumination changes is analyzed. In Figure 11 the recognition rate of the shape-based matching approach is plotted using different values for the greediness parameter, as in the case of occlusion. Additionally, the robustness of the modified Hough transform depending on the minimum edge amplitude in the search image is also analyzed.

Here, the effect of different greediness values is smaller than in the case of occlusion. Disregarding the result obtained with greediness=1, the discrepancy is smaller than 10%. In contrast, the recognition rate of the modified Hough transform strongly depends on the chosen threshold for edge extraction in the search image. The higher the minimum edge amplitude the more edge pixels were missed, because dimming the light as well as stronger ambient illumination reduces the contrast. Thus, this effect is comparable to the effect of higher occlusion. Therefore, a high recognition rate can be obtained by setting the minimum score to a lower value or by choosing a lower threshold for the edge amplitudes. For example, a minimum score of 50% and an edge threshold of 10 leads to a recognition rate of 84%. Nevertheless, the true invariance of the shape-based matching approach could not be reached by the modified Hough transform.

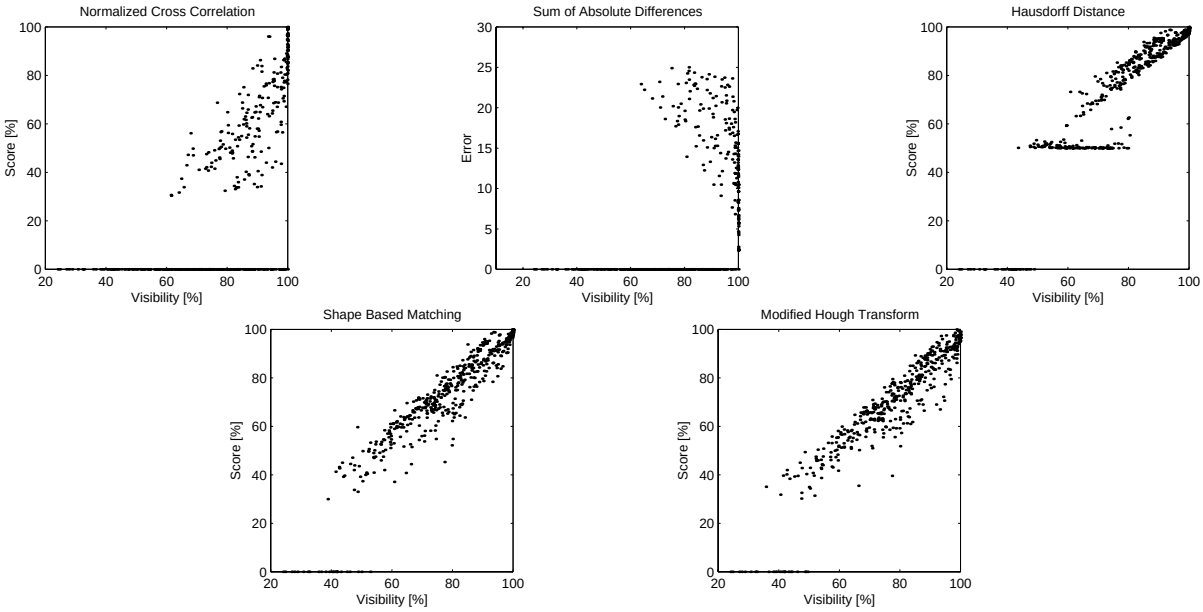


Figure 9. Extracted scores plotted against the visibility of the object. For the approach using the sum of absolute differences the error is plotted instead of the score.

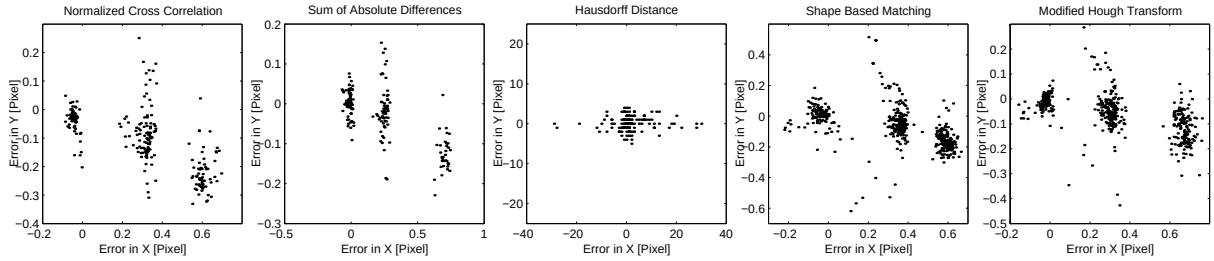
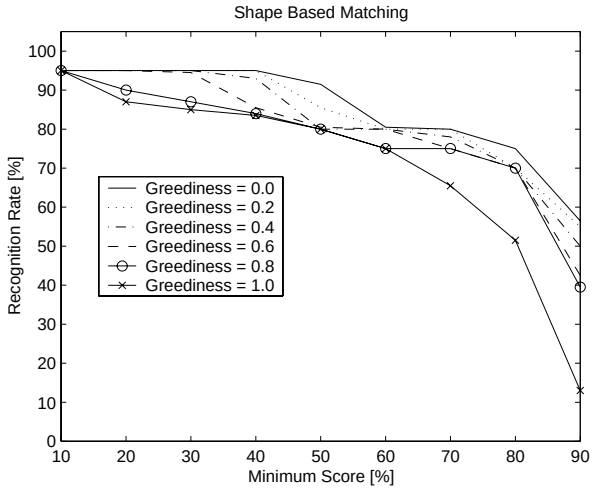


Figure 10. Position errors of the approaches using the 500 images of the occluded IC.

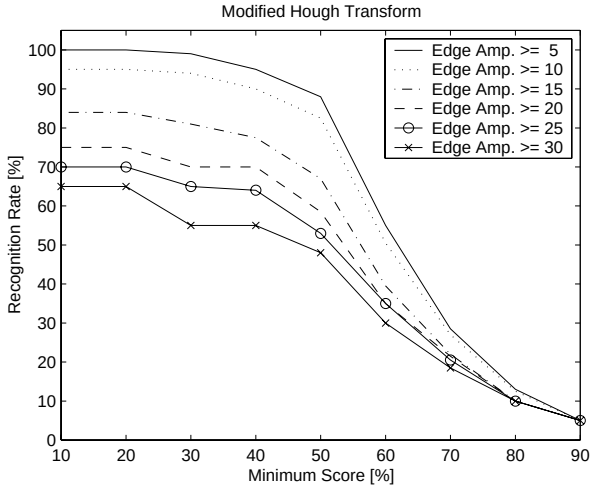
Figure 12 shows a comparison of the robustness of all approaches. Again, the sum of absolute differences shows low recognition rates: in comparison to the robustness against occlusion the overall recognition rate even decreases. Now, the best recognition rate that could be obtained using a maximum error of 30 was only 11%. By comparing this value to the result obtained for a maximum error of 20, which is also 11%, it is obvious that even by further increasing the maximum error no improvement can be reached. In comparison, the recognition rate of the normalized cross correlation is substantially better. This can be attributed to its normalization, which compensates at least global illumination changes. The results obtained by the Hausdorff distance are superior to that of the both approaches last named but could not reach the performance of the shape-based matching approach by far. If the minimum score is set low enough, the recognition rate of the modified Hough transform surpasses that of the shape-based matching, however, for higher values its recognition rate rapidly falls.

3.3.2. Accuracy

Since the Hausdorff distance does not return the object position in subpixel accuracy, only the accuracy of the four remaining candidates are evaluated in this section. To asses the accuracy of the extracted model position and orientation a straight line was fitted to the mean extracted coordinates of position and orientation. This is legitimated by the linear variation of the position and orientation of the IC in the world as described in section 3.2. The residual errors of the line fit, shown in the Figures 13 – 15, are an extremely good indication of the achievable accuracy. As can be seen

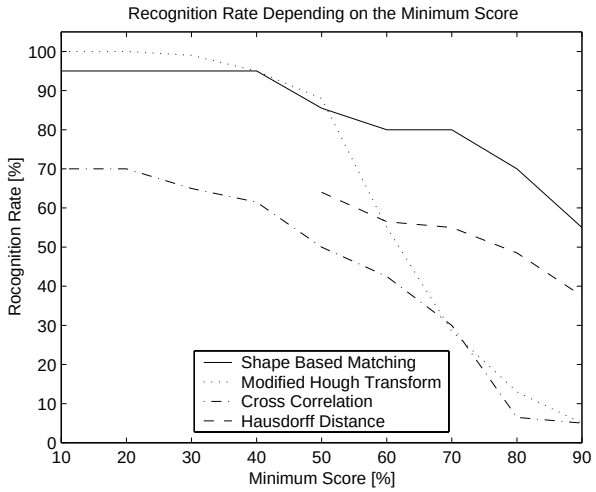


(a)

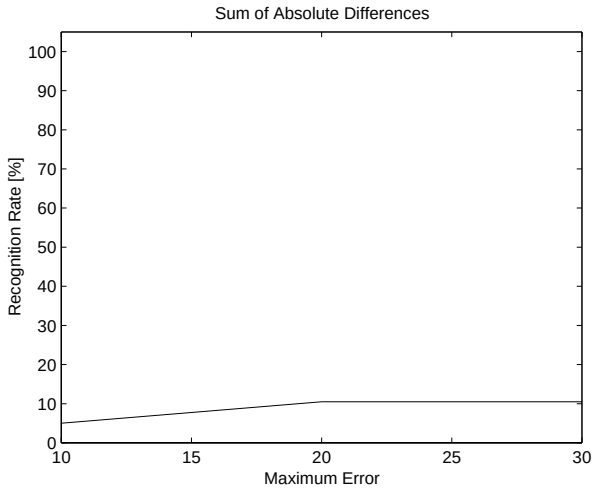


(b)

Figure 11. Recognition rate in the case of illumination changes of (a) the shape-based matching using different values for the parameter greediness and (b) the modified Hough transform using different values for the minimum edge amplitude in the search image.



(a)



(b)

Figure 12. The recognition rate of different approaches indicates the robustness against arbitrary illumination changes. This figure shows the recognition rate (a) of four candidates depending on the minimum score and (b) of the approach using the sum of the absolute differences depending on the maximum error.

from the Figures 13 and 14 the errors in x approximately have the same magnitude as y. The position accuracy of the normalized cross correlation, the modified Hough transform and the shape-based matching approach are very similar. They corresponding errors are in most cases smaller than 1/20 pixel. The two conspicuous peaks in the error plot of Figure 13 occur for all three approaches with similar magnitude. Therefore, it is most probably, that the chip was not shifted exactly and thus, the error must be attributed to a deficient acquisition. However, the high errors in x and y of about 1/10 pixel of the approach using the sum of absolute differences can not be attributed to this. These errors oscillates with a period of 1 pixel: The minima are reached for shifts of $\frac{1}{2} \cdot n$ pixels, the maxima occur at $\frac{1}{2} \cdot n + \frac{1}{4}$ pixels, where n is an integer.

Figure 15 shows the corresponding errors in orientation. Here, the shape based matching approach is superior to all other candidates reaching a maximum error of $1/12^\circ$ (5') in this example. The angle accuracy of the other three candidates is about $1/6^\circ$ (10').

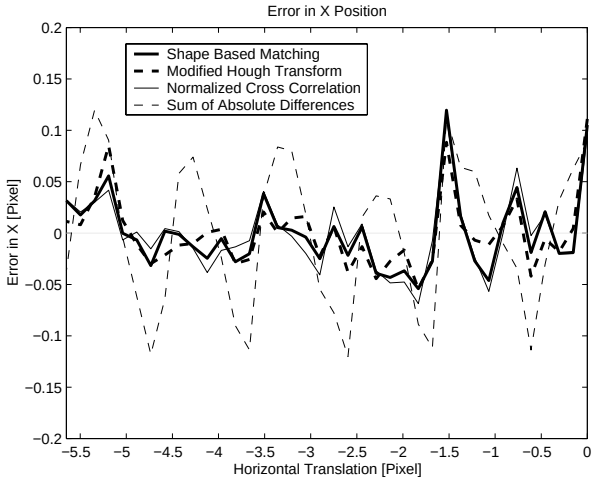


Figure 13. Position accuracy as the difference between the actual x coordinate of the IC and the x coordinate returned by the recognition approach while shifting the chip successively by $1/7$ pixel to the left.

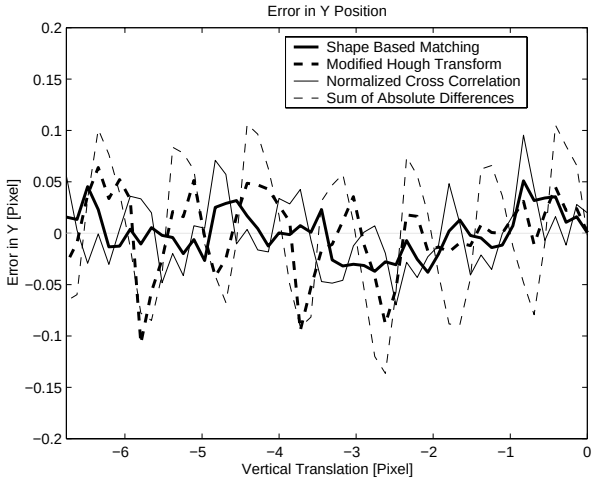


Figure 14. Position accuracy as the difference between the actual y coordinate of the IC and the y coordinate returned by the recognition approach while shifting the chip successively by $1/7$ pixel upwards.

3.3.3. Computation Time

The last criterion, which was applied, is the computation time of the recognition approaches. Figure 16 shows the mean recognition time of the four approaches for each shift of the IC. In Figure 16 (a) the angle interval was restricted to $[-30°;+30°]$. In this respect the shape-based matching approach (50 ms) and the modified Hough transform (60 ms) are substantially faster than existing approaches using the

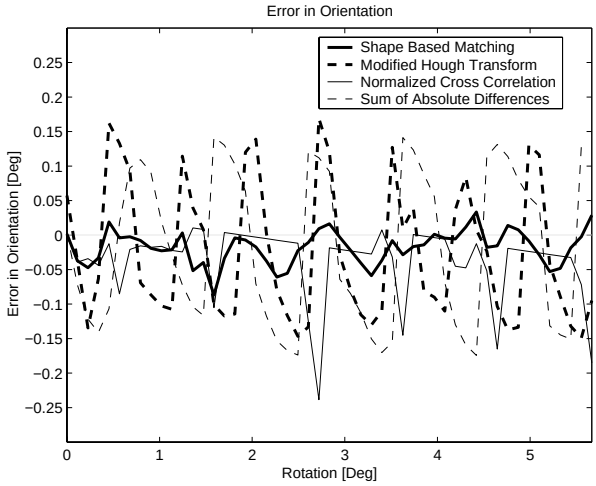
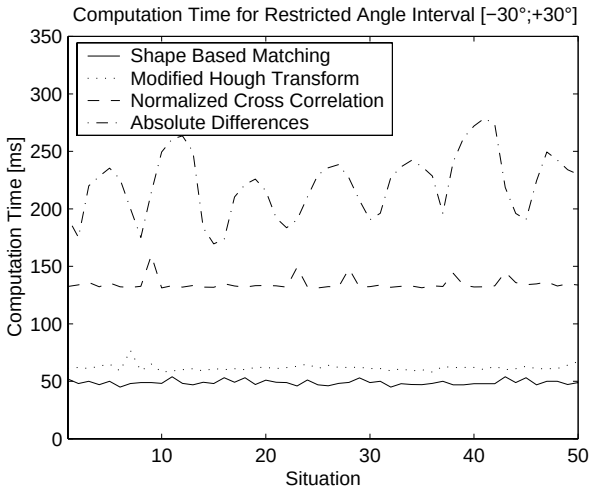


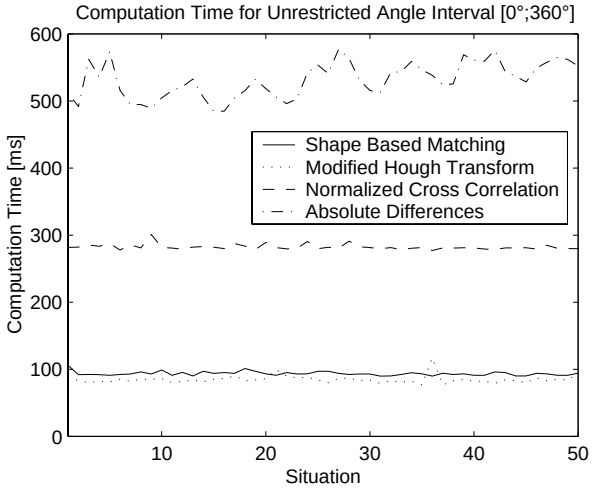
Figure 15. Orientation accuracy as the difference between the actual object orientation of the IC and the returned angle by the recognition approach while rotating the chip successively by approx. $1/9°$ counterclockwise.

normalized cross correlation (140 ms) or the sum of absolute differences (220 ms). The results when using an unrestricted angle interval are shown in Figure 16 (b). Now, the modified Hough transform (90 ms) is slightly faster than the shape-based matching approach (100 ms), which indicates an advantage of the modified Hough transform over the shape-based matching if the transformation space increases. The computation time of the normalized cross correlation (290 ms) and the sum of absolute differences (530 ms) approximately increase with the same ratio. That is, in this example our new approaches are 2.5 to 5.3 times faster than the existing methods.

A similar behavior is obtained when searching for the rotated IC. In Figure 17 the mean recognition time of the four approaches for each orientation of the IC is shown. What should be noted is that the more the IC is rotated relatively to the reference orientation the longer the computation time of the normalized cross correlation. Obviously, the implementation of [7] does not scan the whole orientation range at the highest pyramid level before the matches are traced through the pyramid but starts with a narrow angle range close to the reference orientation. The adjacent orientations are not scanned until no matches could be found in the angle range close to the reference orientation. Thus, the computation time of the normalized cross correlation shown in Figure 16 is not directly comparable to the other approaches, because the orientation range of $[-30°;+30°]$ or $[0°;+360°]$ is not really scanned, i.e., a comparable computation time would be still higher.

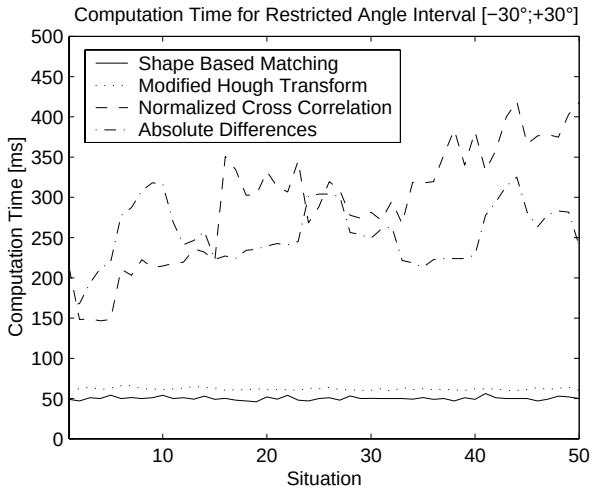


(a)

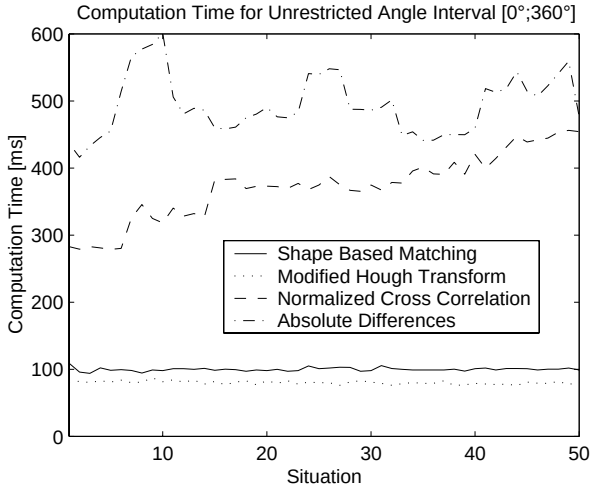


(b)

Figure 16. Computation time of the different approaches on a 400 MHz Pentium II using the shifted IC (a) with angle restriction to $[-30°;+30°]$ and (b) without angle restriction.



(a)



(b)

Figure 17. Computation time of the different approaches on a 400 MHz Pentium II using the rotated IC (a) with angle restriction to $[-30°;+30°]$ and (b) without angle restriction.

4. Conclusions

We presented an extensive performance evaluation of five object recognition methods. For this purpose, the normalized cross correlation and the sum of absolute differences as two standard similarity measures are compared to the Hausdorff distance and two novel recognition methods that we developed with the aim to fulfil increasing industrial demands. We showed that our new approaches have considerable advantages and are substantially superior to the existing methods. In most cases the shape-based matching approach and the modified Hough transform show equivalent behavior. The shape-based matching approach should be preferred when dealing with intense illumination changes and situations where it is important to know the exact orientation of the object. In contrast, the modified Hough transform is better suited when either the dimensionality or the extension of the parameter space increases and the computation time is a critical factor.

References

[1] D. H. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.

[2] G. Borgefors. Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(6):849–865, Nov. 1988.

[3] L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, Dec. 1992.

[4] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, June 1986.

[5] W. Förstner. A framework for low level feature extraction. In J.-O. Eklundh, editor, *Third European Conference on Computer Vision*, volume 801 of *Lecture Notes in Computer Science*, pages 383–394, Berlin, 1994. Springer-Verlag.

[6] S.-H. Lai and M. Fang. Robust and efficient image alignment with spatially varying illumination models. In *Computer Vision and Pattern Recognition*, volume II, pages 167–172, 1999.

[7] Matrox. *Matrox Imaging Library - User Guide*. Matrox Electonic Systems Ltd, Mar. 2001. Version 6.1.

[8] C. F. Olson and D. P. Huttenlocher. Automatic target recognition by matching oriented edge pixels. *IEEE Transactions on Image Processing*, 6(1):103–113, Jan. 1997.

[9] W. J. Rucklidge. Efficiently locating objects using the Hausdorff distance. *International Journal of Computer Vision*, 24(3):251–270, 1997.

[10] C. Steger. An unbiased detector of curvilinear structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(2):113–125, Feb. 1998.

[11] C. Steger. Similarity measures for occlusion, clutter, and illumination invariant object recognition. In B. Radig and S. Florczyk, editors, *Mustererkennung 2001*, pages 148–154, München, 2001. Springer.

[12] S. L. Tanimoto. Template matching in pyramids. *Computer Graphics and Image Processing*, 16:356–369, 1981.

[13] M. Ulrich. Real-time object recognition in digital images for industrial applications. Technical Report PF-2001-01, Lehrstuhl für Photogrammetrie und Fernerkundung, Technische Universität München, 2001.

[14] M. Ulrich, C. Steger, A. Baumgartner, and H. Ebner. Real-time object recognition in digital images for industrial applications. In *5th Conference on Optical 3-D Measurement Techniques*, pages 308–318, Wien, Oct. 2001.

[15] M. Ulrich, C. Steger, A. Baumgartner, and H. Ebner. Real-time object recognition using a modified generalized hough transform. In *21. Wissenschaft-Technische Jahrestagung der DGPF*, volume 10, pages 571–578, Konstanz, Sept. 2001.