PERFORMANCE COMPARISON OF 2D OBJECT RECOGNITION TECHNIQUES

Markus Ulrich[a,b,*] and Carsten Steger[b]

[a]Chair for Photogrammetry and Remote Sensing, Technische Universität München,
Arcisstraße 21, 80290 München, Germany - markus.ulrich@bv.tu-muenchen.de
[b]MVTec Software GmbH, Neherstraße 1, 81675 München, Germany - {ulrich,steger}@mvtec.com

Commission III, Working Group III/5

KEY WORDS: Computer Vision, Object Recognition, Performance Comparison, Industrial Application

ABSTRACT

We propose an empirical performance evaluation of five different 2D object recognition techniques. For this purpose, two novel recognition methods that we have developed with the aim to fulfill increasing industrial demands are compared to the normalized cross correlation and the Hausdorff distance as two standard similarity measures in industrial applications, as well as to PatMax® — an object recognition tool developed by Cognex. Additionally, a new method for refining the object's pose based on a least-squares adjustment is included in our analysis. After a description of the respective methods, several criteria that allow an objective evaluation of object recognition approaches are introduced. 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.

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 or quality control. 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. The latter are usually taken as an approximation to the true perspective transformations an object may undergo.

A large number of object recognition strategies exist. All approaches to object recognition examined in this paper — possibly with the exception of PatMax® — use pixels as their geometric features, i.e., not higher level features like lines or elliptic arcs. Since PatMax® is a commercial software tool, a detailed technical description is not available and therefore no statements about the used features can be made within the scope of this paper. Nevertheless, we included PatMax® in our evaluation because it is one of the most powerful commercial object recognition tools. Thus, we are able to rate the performance of our two novel approaches not only by comparing them to standard recognition techniques but also to a high-end software product.

Several methods have been proposed to recognize objects in images by matching 2D models to images. A survey of matching approaches is given in (Brown, 1992). In most 2D matching object recognition implementations the search is usually done in a coarse-to-fine manner, e.g., by using image pyramids (Tanimoto, 1981). 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 (see (Brown, 1992) or (Lai and Fang, 1999), for example). 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. Two example representatives of this class are the hierarchical chamfer matching (Borgefors, 1988) and the Hausdorff distance (see (Rucklidge, 1997) or (Olson and Huttenlocher, 1997)). Finally, another class of edge based object recognition algorithms is based on the generalized Hough transform (GHT) (Ballard, 1981). 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 this paper our two new approaches are analyzed and their performance is compared to that of PatMax® and two of the above mentioned approaches. Additionally, our new method for refining the object's pose, i.e., improving the accuracy of the transformation parameters, based on a least-squares adjustment is included in our evaluation. The analysis of the performance characteristics of object recognition methods is an important issue. First, 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. Second, it makes an algorithm comparable to other algorithms, thus helping users in selecting the appropriate method for the task they have to solve. Therefore, in this paper an attempt is made to characterize the performance of five different object recognition approaches, which are briefly introduced in the following section. A more detailed description of the approaches can be found in the corresponding references or in (Ulrich and Steger, 2001) and (Ulrich and Steger, 2002), where also the evaluation is described more extensively.

2 EVALUATED OBJECT RECOGNITION METHODS

First of all, we introduce some definitions that facilitate the comparison between the seven 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. A region of interest (*ROI*) R specifies the object's location in the image. The image part defined by R is called *reference image* I^{r}. 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 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.

The first method to be analyzed is the *Normalized Cross Correlation* (Brown, 1992) because it is a rather widely spread method in industry and therefore well known in the application area of object recognition. The *Hausdorff Distance* (Rucklidge, 1997) is the second candidate, which is also the core of many recognition implementations, because of its higher robustness against occlusions and clutter in contrast to the normalized cross correlation. Additionally, *PatMax*® and two novel approaches, which are referred to as *Shape-Based Matching* (Steger, 2001) and *Modified Hough Transform* (Ulrich, 2001, Ulrich et al., 2001a, Ulrich et al., 2001b) below, are included in our analysis. The least-squares adjustment of the object's pose assumes approximate values for the transformation parameters and therefore, is no self-contained object recognition method. Thus, it can be used to improve the accuracy of the returned parameters from any recognition technique that uses the edge position and edge orientation as features by a subsequent execution of the least-squares adjustment. In our current study we use the shape-based matching as basis for the least-squares adjustment. The development of our new 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 six candidate algorithms under different external circumstances.

2.1 Normalized Cross Correlation

For the purpose of evaluating the performance of the normalized cross correlation (see (Brown, 1992), for example) we use — as one typical representative — the current implementation of the *Matrox Imaging Library* (MIL), which is a software development toolkit of *Matrox Electronic Systems Ltd.* (Matrox, 2001). In this implementation image pyramids are used for speed up. The quality of the match is returned by mapping the normalized cross correlation to a score value between 0.0 and 1.0. Subpixel accuracy is obtained by a subsequent refinement of the position and orientation parameters by interpolation.

2.2 Hausdorff Distance

The Hausdorff distance measures the extent to which each pixel of the binarized reference image lies near some pixel of the binarized search image and vice versa. We use the implementation of (Rucklidge, 1997), which uses the symmetric partial undirected Hausdorff distance to reduce the sensitivity to outliers applying a forward and a reverse fraction of points that must fulfill a certain distance criterion. 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.

2.3 PatMax®

As described in its documentation (Cognex, 2000) PatMax® uses geometric information. The model representation, which can be visualized by PatMax®, apparently consists of subpixel precise edge points and respective edge directions. From this we can conclude that PatMax® uses similar features as the shape-based matching. To speed up the search, a coarse-to-fine approach is implemented. To indicate the quality of the match, PatMax® computes a score value between 0.0 and 1.0.

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 (Steger, 2001).

The model consists of a set of points and their corresponding direction vectors. In the matching process, a transformed model is 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 of the image. The normalized similarity measure 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. 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 horizontal translation, vertical translation, rotation, and scale) 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) (Ballard, 1981) 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 (Ulrich, 2001), (Ulrich et al., 2001a), and (Ulrich et al., 2001b) we introduce our novel approach that eliminates the major drawbacks of the GHT using a hierarchical search strategy in combination with an effective limitation of the search space. The resulting pose parameters of position and orientation are refined using the method describe 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.

2.6 Shape-Based Matching Using Least-Squares Adjustment

To improve the accuracy of the transformation parameters, we developed a method that minimizes the distance between tangents of the model shape and potential edge pixels in the search

image using a least-squares adjustment (see also (Wallack and Manocha, 1998)). Approximate transformation parameters are assumed to be known, which can be obtained by any preceding object recognition method that uses the edge position and orientation as features, e.g., the shape-based matching or the modified Hough transform. The minimization is realized using a single step algorithm (Press et al., 1992). This approach is described more extensively in (Ulrich and Steger, 2002). We implemented the least-squares adjustment as an extension of the shape-based matching, which returns the requested approximate values accurately enough.

3 EVALUATION

3.1 Evaluation Criteria

We use three main criteria to evaluate the performance of the six object recognition methods and to build a common basis that facilitates 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 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 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, which is defined as the number of images in which the object was correctly recognized divided by the total number of images.

The second criterion is 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 of modern microprocessors, 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 five remaining candidates. The least-squares adjustment is implemented as a subsequent refinement step in combination with the shape-based matching. Therefore, only the accuracy and the recognition time of the least-squares adjustment are analyzed, since the robustness is not affected and hence is the same as the robustness of the underlying recognition approach.

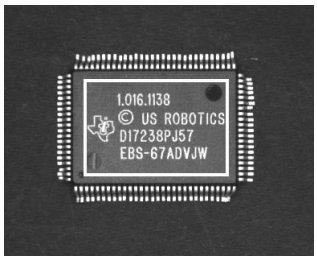

Figure 1: An IC is used as the object to be recognized.

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 1, 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. For the recognition methods that segment edges during model creation (Hausdorff distance, shape-based matching, modified Hough transform, least-squares adjustment) 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 using image pyramids, four pyramid levels were used to speed up the search, which we found to be the optimum number for our specific object. When using PatMax®, there is no parameter that allows to explicitly specify the number of image pyramids to use. Instead, the parameter *coarse grain limit* can be used to control the depth of the hierarchical search, which has a similar meaning as, but can not be equated with, the number of pyramid levels. Since this parameter can be set automatically, we assumed the automatically determined value as the optimum one and did not use a manual setting.

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 to 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.

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 2 shows two of the 500 images that we used to test the robustness against occlusion. For the approaches that segment edges in the search image (modified Hough transform and Hausdorff distance) the minimum edge amplitude in the search image was set to 30, i.e., to the same value as in the reference image. 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. In addition to the recognition rate, the correlation between the actual occlusion and the returned score values are examined, because the correlation between the visibility of the object and the returned score value is also an indicator for robustness. If, for example, only half of the object is visible in the image then, intuitively, also the score should be 50%, i.e., we expect a very high correlation in the ideal case. 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

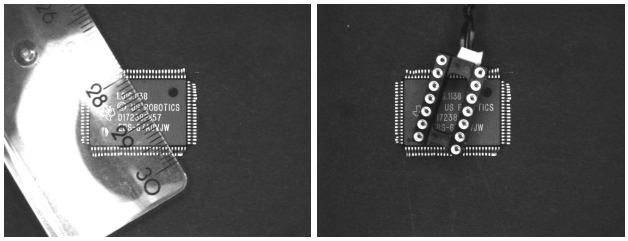

Figure 2: Two of the 500 images that were used to test the robustness against occlusions.

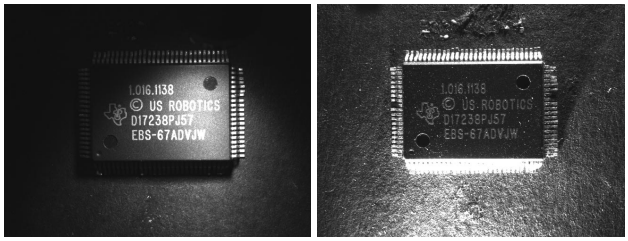

Figure 3: Two of the 200 images that were used to test the robustness against arbitrary illumination changes.

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 entire image. Different values for the parameter of the minimum score were applied, which can be chosen for all approaches. The minimum score specifies the score a match must at least have to be interpreted as a found object instance. 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.

To test the robustness against arbitrary illumination changes, a second sequence of images of the IC was taken, which includes various illumination situations. Two example images are displayed in Figure 3. Due to a smaller distance between the IC and the camera, the ROI is now 255 $\times$ 140 pixels containing 3381 model points on the lowest pyramid level. The parameter settings for the six methods is equivalent to the settings for testing the robustness against occlusions.

3.2.2 Accuracy In this section the experimental set-up that we used to determine 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 PatMax® and the shape-based matching are the only candidates that are able to recognize scaled objects, only the position and orientation accuracy of the five approaches were tested.

To test the accuracy, the IC was mounted onto a table that can be shifted with an accuracy of 1 μm 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 μm 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.

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 computation time is considered.

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 analysis of the plots are structured as in the previous section, i.e., first the results of the robustness, then the accuracy, and finally the computation time are presented.

3.3.1 Robustness

Occlusion. First, the sequence of the occluded IC was tested. A complete comparison of all approaches concerning the robustness against occlusion is shown in Figure 4. In the left plot the recognition rate, which is an indicator for the robustness, is plotted depending on the minimum score (see Section 3.2.1). Here, the superiority of our two novel approaches to the standard approaches becomes clear. Note that the robustness of the modified Hough transform hardly differs from the robustness achieved by the shape-based matching. Looking at the other approaches, only PatMax® reaches a comparable result, which is, however, slightly inferior in all cases. Furthermore, when using a restricted parameter space, which is limited to only translations as described in Section 3.2, the recognition rate of PatMax® was up to 14% lower as when using a narrow angle tolerance interval of [-5°;+5°].

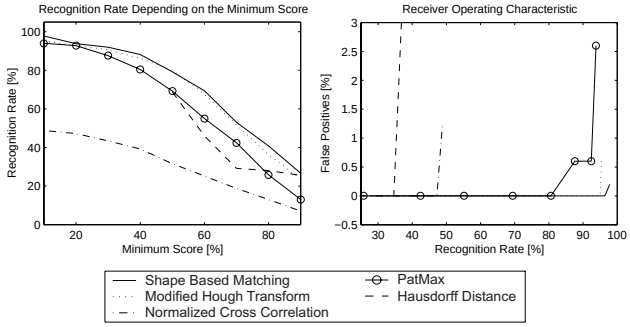

Figure 4: The recognition rate of different approaches indicates the robustness against occlusions. The left figure shows the recognition rate of the five candidates depending on the minimum score. In the right figure the receiver operating characteristic is shown.

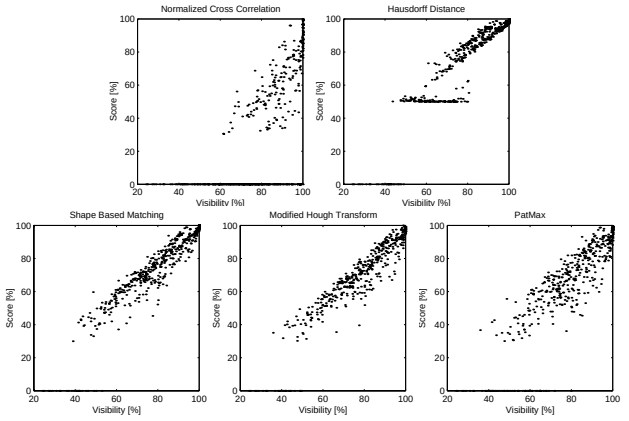

Figure 5: Extracted scores plotted against the visibility of the object.

To avoid that this peculiarity results in an unfavorable comparison for PatMax® we decided to take the angle tolerance interval into account when using PatMax®. The recognition rate of the normalized cross correlation does not reach 50% at all, even if the minimum score is chosen small. In the right plot of Figure 4 the receiver operating characteristic curve is shown, i.e., the false positive rate is plotted depending on the recognition rate. Even for a small recognition rate the number of false positives dramatically increases up to 32% (not visible in the plot due to axis scaling) when using the Hausdorff distance. The normalized cross correlation also tends to return false positives if the recognition rate approaches the maximum value of about 50%. For high recognition values even PatMax®returns wrong matches. Also here, the best results are obtained using the modified Hough transform and the shape-based matching.

Figure 5 displays a plot of the returned score value against the estimated visibility of the object, i.e., the correlation between the visibility of the object and the returned score value is visualized. 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 all approaches except for the Hausdorff distance the minimum score was set to 30%, i.e., in those images in which the object has a visibility of more than 30%, it should be found by the recognition method. For the Hausdorff distance a minimum forward fraction of 50% was used. In the plot of the Hausdorff distance the wrong matches either have a forward fraction of 0% or close to 50%, because of some false positives. Here, a noticeable positive correlation can be observed, but several objects with a visibility of far greater than 50% could not be recognized. This explains the

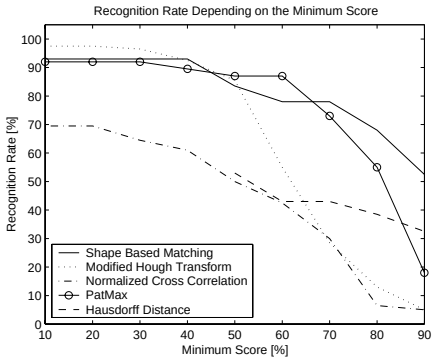

Figure 6: The recognition rate of the different approaches indicates the robustness against arbitrary illumination changes. This figure shows the recognition rate of the five candidates depending on the minimum score.

lower recognition rate in comparison to our approaches, which was mentioned above. The normalized cross correlation also shows positive correlation but 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 evident 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. Also PatMax® results in a nearly ideal point distribution. Nevertheless, in some occlusion cases the object was not found even though the visibility was significantly higher than 30%.

Illumination. Figure 6 shows a comparison of the robustness of all approaches. The recognition rate of the normalized cross correlation is now substantially better than in the case of occlusions. This can be attributed to its normalization, which compensates at least global illumination changes. The Hausdorff distance shows also good results especially in the case of large values for the minimum score, 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. Here, also PatMax® shows very good results: the recognition rate is nearly constant when increasing the minimum score from 10% to 60% but also drops down during further increase.

3.3.2 Accuracy Since the Hausdorff distance does not return the object position in subpixel accuracy, only the accuracy of the five remaining candidates are evaluated in this section. To assess 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 7 and 8, are an extremely good indication of the achievable accuracy.

As can be seen from the Figure 7 the position accuracy of the normalized cross correlation, PatMax®, the modified Hough transform, the shape-based matching and the least-squares adjustment and are very similar. The corresponding errors are in most cases smaller than 1/20 pixel. The two conspicuous peaks in the error plot of Figure 7 occur for all three approaches with similar magnitude. Therefore, and because of the nearly identical lines, it is probable that the chip was not shifted exactly and thus, the error must be attributed to a deficient acquisition. Since the errors in y during a vertical translation approximately have the same magnitude as the errors in x we refrain from presenting these plots.

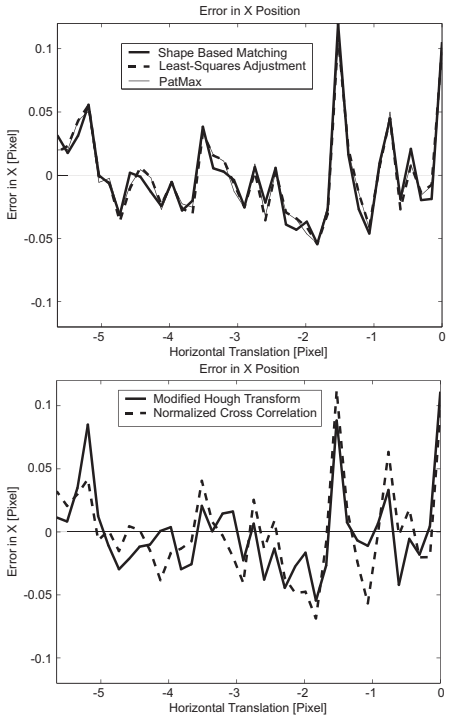

Figure 7: Position accuracy as the difference between the actual x coordinate of the IC and the x coordinate returned by the recognition approach while shifting the chip successively by $1/7$ pixel to the left.

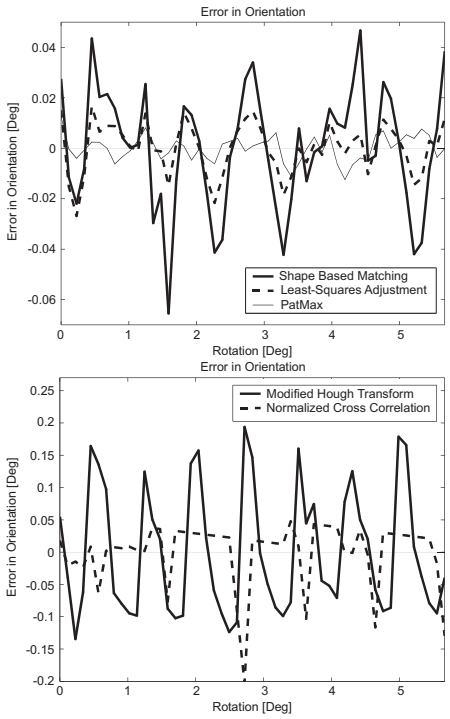

Figure 8: Orientation accuracy as the difference between the actual object orientation of the IC and the returned angle by the recognition approach while rotating the chip successively by approx. $1/9^\circ$ counterclockwise.

Figure 8 shows the corresponding errors in orientation. Here, the least-squares adjustment and PatMax® are superior to all other candidates reaching maximum errors of 1/50° and 1/100° in this example. In comparison, when looking at the result of the shape based matching, the improvement of the least-squares adjustment is evident: the maximum error is reduced to 1/50° compared to the result of the shape-based matching without least-squares adjustment, which was about 1/16°. The error magnitude of the other three approaches is higher 1/6° (10') in this example.

3.3.3 Computation Time The last criterion that was applied is the computation time of the recognition approaches. In Table 1 the mean recognition times of all approaches using the sequence with the shifted IC as well as using the sequence with the rotated IC are listed. Additionally, the time increase Δ from the restricted to the unrestricted angle interval is printed in percent.

	Shifted IC			Rotated IC		
	R [ms]	UR [ms]	Δ [%]	R [ms]	UR [ms]	Δ [%]
SBM	57	126	121	50	100	100
LSA	65	133	105	60	110	83
MHT	72	112	56	62	80	29
PM	55	88	60	80	193	141
NCC	132	281	113	294	373	27

Table 1: Mean computation times of the shape-based matching (SBM), the least-squares adjustment (LSA), the modified Hough transform (MHT), PatMax® (PM), and the normalized cross correlation (NCC) on a 400 MHz Pentium II. Additionally, the time increase Δ from the restricted (R) to the unrestricted (UR) angle interval is printed in percent.

First the angle interval was restricted to [-30°;+30°]. In this respect the shape-based matching approach, the least-squares adjustment, the modified Hough transform and PatMax® are substantially faster than existing traditional approaches using the normalized cross correlation. The results when using an unrestricted angle interval are shown in an extra column of table 1. Here, the modified Hough transform is slightly faster than the shape-based matching approach, which indicates an advantage of the modified Hough transform over the shape-based matching if the transformation space increases. This assumption is also supported when looking at the percental time increase of the modified Hough transform: The computation time merely increases by 56%, which is the smallest value in this test. Also PatMax® shows only a small increase, whereas the computation time of the normalized cross correlation increases more dramatically. In this example the computation times of our new approaches are only 0.2 to 0.5 times as high as those of the traditional methods but about 1.0 to 1.5 times as high as that of PatMax®.

For the most methods, a similar behavior is obtained when searching for the rotated IC. What we recognized during evaluation 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 (Matrox, 2001) 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. Thus, the computation time of the normalized cross correlation 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. Also the corresponding Δ-values would be higher.

Also in the case of the rotated IC, the modified Hough transform seems to be the method that is most suitable when dealing with large parameter spaces because of its small time increase — in this case — of only 29%. To get the effective computation time of the least-squares adjustment we have to subtract the computation time of the shape-based matching. The difference is in the range of 7 to 10 milliseconds and does not depend on the size of the parameter space. Therefore, the larger the parameter space the less the influence of this constant part, which is the reason for the smaller Δ-values in Table 1 of the least-squares adjustment in

contrast to the shape-based matching. By all means, it should be pointed out that the computation time of PatMax® in this example is much slower than in the example above. Thus, in this case our new approaches are not only dramatically faster than the traditional methods but also considerably faster than PatMax®. The reason for the totally different computation times of PatMax® in the two example sequences is the automatic computation of the coarse grain limit (see Section 3.2). During the first sequence using the shifted IC the grain limit was automatically set to 3.72 and during the second sequence using the rotated IC the grain limit was automatically set to only 2.92. There is no obvious reason for this difference, because the object was the same in both cases. Experiments have shown that the automatic computation of the grain limit may result in a totally different value if the region of interest is shifted by just 1 pixel without changing the number of edge points within the region.

4 CONCLUSIONS

We presented an extensive performance evaluation of six object recognition methods. For this purpose, the normalized cross correlation and the Hausdorff distance as two standard similarity measures in industrial applications were compared to PatMax® — an object recognition tool developed by (Cognex, 2000) — and two novel recognition methods that we have developed with the aim to fulfill increasing industrial demands. Additionally, a new method for refining the object's pose based on a least-squares adjustment was included in our analysis. We showed that our new approaches have considerable advantages and are substantially superior to the existing traditional methods especially in the field of robustness and recognition time. Even in comparison to PatMax® particularly the shape based matching in combination with the least-squares adjustment shows very good results. In contrast, we exposed some inconsistencies when using PatMax®: restricting the parameter space to translations causes the recognition rate to drop down dramatically. The automatic setting of the grain limits is problematical since significantly different results are obtained using the same object and, although a high value for the grain limits leads to a fast computation it also results in a high risk of returning false positives. In contrast, a low value means higher robustness but slow computation.

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.

The breakdown points of the normalized cross correlation are its low robustness against occlusions/clutter and non-linear illumination changes. The relatively slow computation is another factor that limits its applicability. As a breakdown point of the Hausdorff distance its trend to return false positives should be mentioned, which often occur in the presence of clutter.

Aside from these conclusions, it should be pointed out that some of the results might change if we chose, for example, other implementations of the approaches, other parameter constellations or other image sequences. Therefore, our comparison is more of a qualitative nature rather than of a quantitative one. Nevertheless, our results are very objective and help potential users to find the optimum approach for their specific application. To facilitate an extended comparison including other recognition methods, interested parties are requested to send an e-mail to the authors in order to get the sequences that we used for the evaluation.

REFERENCES

Ballard, D. H., 1981. Generalizing the Hough transform to detect arbitrary shapes. Pattern Recognition 13(2), pp. 111–122.

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

Brown, L. G., 1992. A survey of image registration techniques. ACM Computing Surveys 24(4), pp. 325–376.

Cognex, 2000. Cognex MVS-8000 Series - CVL Vision Tool Guide. Cognex Corporation. CVL 5.5.1.

Lai, S.-H. and Fang, M., 1999. Robust and efficient image alignment with spatially varying illumination models. In: Computer Vision and Pattern Recognition, Vol. II, pp. 167–172.

Matrox, 2001. Matrox Imaging Library - User Guide. Matrox Electonic Systems Ltd. Version 6.1.

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

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P., 1992. Numerical Recipes in C: The Art of Scientific Computing. 2nd edn, Cambridge University Press, Cambridge.

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

Steger, C., 2001. Similarity measures for occlusion, clutter, and illumination invariant object recognition. In: B. Radig and S. Florczyk (eds), Mustererkennung 2001, Springer, München, pp. 148–154.

Tanimoto, S. L., 1981. Template matching in pyramids. Computer Graphics and Image Processing 16, pp. 356–369.

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

Ulrich, M. and Steger, C., 2001. Empirical performance evaluation of object recognition methods. In: H. I. Christensen and P. J. Phillips (eds), Empirical Evaluation Methods in Computer Vision, IEEE Computer Society Press, Los Alamitos, CA, pp. 62–76.

Ulrich, M. and Steger, C., 2002. Performance evaluation of 2d object recognition techiques. Technical Report PF-2002-01, Lehrstuhl für Photogrammetrie und Fernerkundung, Technische Universität München.

Ulrich, M., Steger, C., Baumgartner, A. and Ebner, H., 2001a. Real-time object recognition in digital images for industrial applications. In: 5th Conference on Optical 3-D Measurement Techniques, Vienna, pp. 308–318.

Ulrich, M., Steger, C., Baumgartner, A. and Ebner, H., 2001b. Real-time object recognition using a modified generalized Hough transform. In: 21. Wissenschaft-Technische Jahrestagung der DGPF, Vol. 10, Konstanz, pp. 571–578.

Wallack, A. and Manocha, D., 1998. Robust algorithms for object localization. International Journal of Computer Vision 27(3), pp. 243–262.