Cascaded Reduction and Growing of Result Sets for Combining Object Detectors

Uwe Knauer[1,2] and Udo Seiffert[2]

[1] Humboldt University, Unter den Linden 6, 10099 Berlin, Germany
[2] Fraunhofer IFF, Sandtorstr. 22, 39106 Magdeburg, Germany
knauer@informatik.hu-berlin.de

Abstract. In this paper *cascaded reduction and growing of result sets* is introduced as a principle for combining the results of different object detectors. First, different candidate operating points are selected for each object detection algorithm. This procedure is based on the analysis of precision and recall of the individual methods. Selecting an appropriate operating point prior to fusion is important because it regulates the cardinal number of the result set. As diversity and correlation between object detectors also depend on the elements of the result sets, this and the application of set operations allow to create a final set of detected objects by including missing and excluding false detections. The approach allows both diverse and correlated detectors to contribute to the performance of the combined detector. The performance of the proposed algorithm is compared to other combining algorithms. It outperforms or competes with existing state of the art combiners for several datasets. Additionally, the results provide a significant improvement in the interpretability of the combining rules. As a unique feature of the proposed algorithm, the found operating points can be used to reconfigure the object detection algorithms to adapt their individual results to the needs of the combination procedure allowing a reduction in runtime.

Keywords: object detection, combining, fusion, image processing.

1 Introduction

Object detection in digital images is an important task in many application fields such as microscopy, remote sensing, robot vision, tracking in surveillance applications, and autonomous navigation. While good solutions exist for some applications such as face recognition, it is still difficult to achieve sufficient detection rates for many other problems to fully automate the analysis of images. The main reason is the variety of the appearance due to changes of camera perspective, illumination, observation of deformable objects, and occlusion of objects. The combination of different algorithms has the potential to overcome some of these problems. First, related work on the combination of object detectors and classifiers is introduced. Second, a novel approach to late fusion is described. After introducing the dataset, the results of combination are presented. The paper concludes with a summary of the potentials of the new algorithm.

Z.-H. Zhou, F. Roli, and J. Kittler (Eds.): MCS 2013, LNCS 7872, pp. 121–133, 2013.

2 Related Work

The combination of object detection algorithms is closely related to the field of multiple classifier systems, since most methods can be applied to combine object detection results. Successful approaches to classifier fusion are based on Bayes theorem [11], the Dempster-Shafer theory of evidence [2], fuzzy logic [4,5,13], voting [15], analysis of correspondence [8,18], and on machine learning approaches [9]. However, modifications are required because object detection results differ from classifier results. First, object detectors output positions and dimensions of objects in addition to the measurement, rank, or abstract level outputs [26]. Consequently, the matching of the result sets of different detectors is needed. Second, object detectors do not output a label or a measurement for each image patch, therefore, the combination algorithms must deal with missing information from some of the detectors. Third, since image patches can overlap or have different scales, matching of objects between detectors may not be unique. Fourth, the cardinal number of the result sets of different detectors for the same image can be different. Additionally, if a measurement level output is provided then a threshold on this value can be used to further reduce the number of elements in the result set of detected objects.

Current research on combining object detection approaches focuses on early or intermediate level fusion [23,21,14]. Fusion of feature vectors at an early stage has the advantage that the spatial relation of the different features are preserved at the pixel or image patch level. Its disadvantage is the curse of dimensionality, using more dimensions requires much more training data for most problems to achieve a good generalization performance for a detector. Other approaches, which claim to be late level fusion methods, still produce extended feature sets [10]. Some classifier fusion approaches have been published for the fusion of different sensor data [1], the boosting of object detection by inclusion of contextual information from classification [20], and the fusion of the decisions of multiple experts for different object categories [16]; however, it is difficult to find published work on combining the final detections of different algorithms for the same object class. In object detection, late fusion is much more related to the abstraction level of the features than in classifier fusion. The reason is obvious, because information processing in human perception works analogously and has always inspired the computer vision community.

Moreover, combining object detectors at the level of detected objects allows not only to fuse different successful approaches to achieve a better performance, but also a deeper understanding of the strengths and weaknesses of individual approaches. It also adds more levels of freedom to the selection of the combined detection methods because they can be handled as a black box.

3 Proposed Method

The basic concepts of the approach are the utilization of operating points and the construction of improved result sets. The consequences of using these concepts are summarized in section 3.1. In the following *operating points* correspond to the thresholds which are used to control the elements contained in the result set of each detector. An operating point is used to convert the measurement level output into an abstract level output. This concept is well-known from analysis of classifier performance in receiver operating characteristic (ROC) and precision recall (PR) space where it is used for the construction of characteristic curves.

3.1 Preliminaries

The operations cut ($\cap$) and union ($\cup$) are used for the fusion of result sets of different detectors. Therefore, correspondences between the elements of the result sets have to be calculated. This must be based on the positions and extends of the objects. The objective is to minimize the error related costs EC:

$$EC = (1 - TPR) \cdot p(+) \cdot C(-|+) + FPR \cdot p(-) \cdot C(+|-) \tag{1}$$

$p(\cdot)$ denotes the apriori probability of the classes and C the costs of false positive and false negative detections. The true positive rate (TPR) and the false positive rate (FPR) depend on the operating points. These are set by thresholds t_A and t_B for the measurement level output of object detectors A and B respectively. Minimization of $EC(t_A, t_B)$ as a function of the two thresholds t_A and t_B is computationally expensive. The number of different threshold values is only bounded by the number of objects in the training dataset. For each pair (t_A, t_B) the costs EC as well as the gradient of the error cost function can be calculated by application of the set operations only.

To solve that problem, the threshold values can be quantized, such that only a limited number of k thresholds is used for each object detector. By setting $k = 3$ we achieve:

1. reduction of complexity for minimization
2. good interpretability

For the last reason the three thresholds are set such that they correspond to operating points with the attributes precise, optimal, and sensitive.

Combining object detectors is repeated until EC stops to decrease. Since the associative law does not hold for arbitrary sets:

$$(A \cap B) \cup C \neq A \cap (B \cup C) \tag{2}$$

the order of object detectors matters. The fusion rule for the object detectors $OD^1, \ldots, OD^M$ has the form:

$$((OD^1_{o_1} \otimes OD^2_{o_2}) \ldots \otimes OD^M_{o_M}) \tag{3}$$

where $o_1, \ldots o_M \in \{1, ..., k\}$ denote the operating points. The brackets denote the cascade and $\otimes$ denotes the selected fusion operation.

3.2 Matching Positions and Shapes

The matching of detection results is a necessary condition for fusion. In the learning phase the object boundaries of a reference dataset are used. In the application phase pairwise matching of the different result sets is needed.

A commonly used measure for performance evaluation of object detectors is the matching of upright rectangles [7]. The overlap of two rectangles can be checked quickly [24]. Two rectangles defined by their centers $\mathbf{M}_1$,$\mathbf{M}_2$ and the distances d^x and d^y to the corner points do not overlap, if any of the following inequations are true:

$$\left|(\mathbf{M}_1 - \mathbf{M}_2)\begin{pmatrix}1\\0\end{pmatrix}\right| > d_1^x + d_2^x \tag{4}$$

$$\left|(\mathbf{M}_1 - \mathbf{M}_2)\begin{pmatrix}0\\1\end{pmatrix}\right| > d_1^y + d_2^y \tag{5}$$

otherwise, the size of the overlap can be calculated easily. An efficient overlapping test for arbitrary convex polygons is known as separating axis test [19]. If a more detailed representation of the object boundaries is needed then this algorithm can be used.

If object detection is compared to other classification problems then an obvious difference is the meaning of false positives. The set of image patches which do not contain an object is typically large and not part of the reference dataset. However, for comparison and combination of different object detectors such a reference is needed. Therefore, the false positives of all methods are combined into a single reference set.

3.3 Selecting Operating Points

Selecting candidate operating points is a preprocessing step. Later, the learning algorithm selects those candidate operating points which provide the maximum error cost reduction. Operating points correspond to points on the precision recall curves of the individual object detectors. It is assumed that each object detector provides a measurement level output. The PR curves are obtained by thresholding the output values. However, if one or more object detectors provide only abstract level output then only a single operating point exist for those methods.

Fig. 1 shows the precision recall curve of an object detector. The positions i of the smallest and the largest threshold are highlighted. If the operating point corresponds to the threshold at position $i = 1$ then the cardinal number of the result set is maximized. This means that the objects of the reference dataset are reproduced best. However, at this operating point the precision has its smallest value because all false detections of the object detector also satisfy the condition that the output value is larger than the selected threshold. Therefore, in this operating range the detector is called sensitive for the presence of objects.

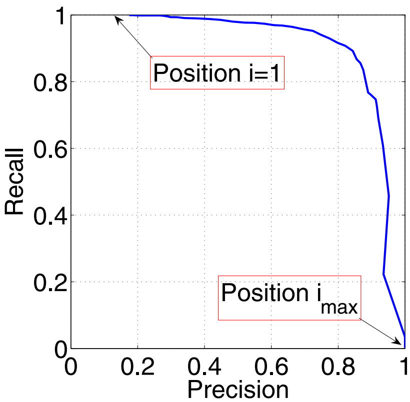


Fig. 1. Precision recall curve for which $k = 3$ operating points have to be selected

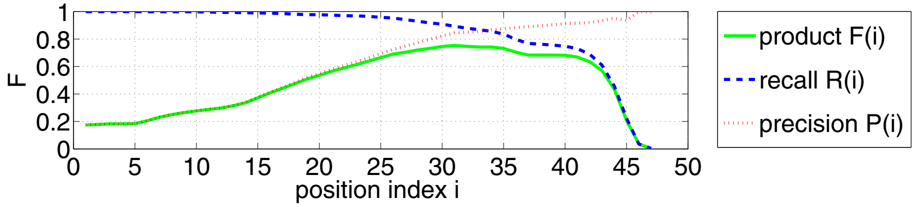


Fig. 2. Values of P,R, and F as a function of position index i

The precise or specific operating points correspond to large threshold values. Only few objects are detected, but the detectors output high confidence values for these detections. Therefore, the recall of the reference dataset is small and the precision is high.

Eq.(6) provides the measure F which is used for the selection of operating points on the precision recall curves:

$$F_i = (P_i - P_{min})(P_{max} - P_{min})^{-1} \cdot (R_i - R_{min})(R_{max} - R_{min})^{-1} \qquad (6)$$

where F_i is the product of the precision P_i and the recall R_i for the operating point with index value i. Fig. 2 shows P and R as a function of the position index i. It is one of the difference to classification problems that the recall must not reach the value 1. Hence, eq.(6) includes offset corrections for both dimensions. Fig. 3 illustrates how the sensitive operating point is found by a threshold on the value of F. Since the recall converges to one in this operating range:

$$F_1 \approx (P_1 - P_{min})(P_{max} - P_{min})^{-1} \qquad (7)$$

With decreasing recall the difference $\Delta = P - F$ grows. In Fig. 3 the threshold $\Delta \leq 0.03$ is plotted. If Δ exceeds the threshold then the operating point is determined by the corresponding position index i.

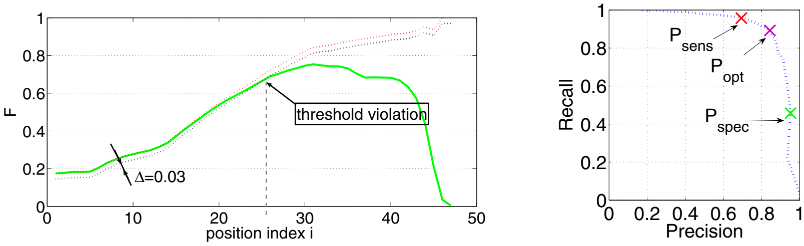


Fig. 3. Position index i of a sensitive operating point (left). Operating points P_{sens} and P_{spec} for $\Delta \leq 0.03$ (right).

Using the same approach an operating point within the high precision operating range is determined. Here, F converges to the value of the normalized recall $(R_i - R_{min})(R_{max} - R_{min})^{-1}$. Fig. 3 shows both operating points P_{spec} und P_{sens} for $\Delta \leq 0.03$. The third operating point P_{opt} was found by minimizing the Euclidian distance between the precision recall curve and the perfect operating point $P_{ideal} = (1, 1)^T$. Search space is limited to the range between P_{sens} and P_{spec}.

3.4 Cascaded Reduction and Growing of Result Sets

Fig. 4 shows the search for combining rules. First, a random order of the object detectors is set. This step is repeated to avoid local minima of the error cost function. A complete search can be done for small numbers n of object detectors. Since the number of possible arrangements is $n!$, with growing n a fixed number of randomly chosen permutations is validated.

The operating point with minimum costs EC is selected for the first object detector OD^1 of the current permutation. The choice depends on the ratio of the error costs which has to be provided as an input value. Selection of the set operation and the operating point is repeated for the next object detector $OD^2, \ldots, OD^n$. By default a ratio of $1:1$ is assumed.

4 Datasets

The proposed algorithm has been tested with several datasets. In this paper, results for a dataset from a surveillance application and for object detection in microscopy images of plant samples are presented. The tasks are to detect individuals in surveillance camera images and the analysis of spatiotemporal fungal patterns [3], Fig. 5 shows the camera field of view as well as a leaf with two colonies.

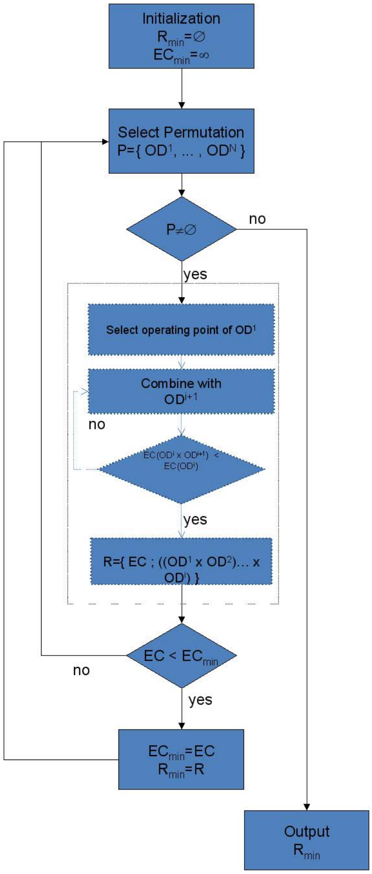


Fig. 4. Flow chart of the search for a combining rule

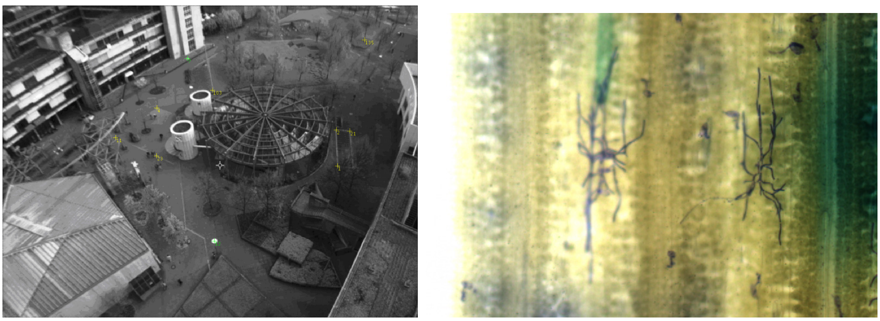


Fig. 5. Field of view of the surveillance camera with ground reference points and image of two phytopathogenic fungi

While the proposed method has been developed for combination of object detection results, it can deal with standard classification problems. The WDBC, Statlog Heart, and SPECTF datasets from the UCI machine learning repository have been selected to test the performance of the algorithm. For all comparisons 10-fold cross-validation is used.

5 Results and Discussion

For detection of people in the surveillance camera images the following algorithms were used:

1. Fixed background subtraction (BS)
2. Difference image (DIFF)
3. Lucas-Kanade tracking (LKT [17])
4. Mixture of Gaussians (MoG [22])
5. Running Average subtraction (RA [6])

Most of the algorithms belong to the class of background subtraction methods while Lucas-Kanade tracking provides a motion based object detection. To estimate the performance of the object detectors a validation set is used. Next, the performance of the combining algorithm is calculated based on 10-fold cross-validation of the outputs of the detectors for this validation set.

The following combination rule was found by cascaded reduction and growing of the result set:

$$(((\mathrm{LKT}_1 \cup \mathrm{BS}_2) \cup \mathrm{DIFF}_2) \cup \mathrm{MoG}_3) \tag{8}$$

For the given dataset the obtained rule uses only a cascade of $\cup$ operations. This indicates that the false detections of the individual object detectors are highly correlated while diversity is given for the true detections. The background subtraction with a Running Average based method does not contribute to the result set, hence, it can be excluded from the multiple object detector system. The indices of the method names are the preselected operating points (1=sensitive, 2=average, 3=specific).

Fig. 6 shows the performance of the individual and the combined detectors in more detail. The diagram shows the precision recall curves of the methods. The PR-curve of the LKT algorithm shows its great contribution to the result. The high overall precision of this detection algorithm allows the multiple detector system to operate the method with a low false positive rate at its sensitive operating point.

Fig. 7 shows a comparison between the approach of cascaded reduction and growing of result sets (CRAGORS) and a number of combiners (e.g. AdaBoost, Random Forest). The proposed method ranks second best and outperforms most of the other methods. The ranking is based on the Euclidian distance between the ideal operating point and the best operating point of each combination method. The red line marks the distance of the operating point of the best performing individual detector. For the detection of fungal patterns a good segmentation into image foreground and background is required. For each pixel 36 features from the

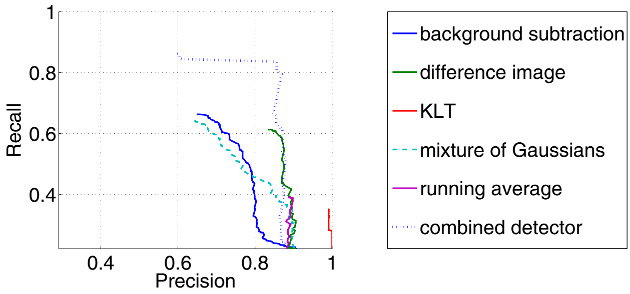


Fig. 6. Precision Recall curves for object detection in surveillance scene

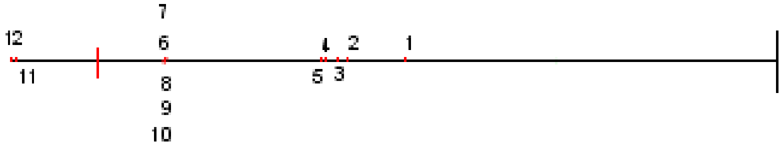


Fig. 7. Ranking of different combining methods for detection of individuals: (1) Random Forest, (2) CRAGORS, (3) CRAGORS (ROC), (4) SCANN, (5) AdaBoost, (6) Sum rule, (7) Max rule, (8) Median rule, (9) Min rule, (10) Product rule, (11) Fuzzy Templates, and (12) Voting

input RGB image are considered, which can be calculated quickly with integral images [12]. The following algorithms were used to find good segmentations of the microscopy images:

1. AdaBoost with J4.8 as weak learner (ADA, 10 stages)
2. Random Forest (RF, 10 trees)
3. Bagging classifier with kNN (BAG, 10 bags)

Finally, the different segmentation results were combined by the proposed algorithm. Only the abstract level outputs of the three segmentation approaches were used. Hence only a single operating point for each method was considered, yielding the following fusion result:

$$((ADA \cup BAG) \cup RF)$$

This combining rule can be easily implemented with binary AND and OR operations per pixel. Fig. 8 shows the improvements of combination and a ranking of different combination algorithms as well. The improved segmentation allows a better detection of objects in subsequent processing steps as well as an improved estimation of important object features such as the area of fungal patterns. The red line marks the performance of the best individual segmentation method. A number of combining algorithms failed to improve the performance (ranks 5-12). Only four algorithms including two variants of the proposed one (ranks 3,4) were capable of improving the results.

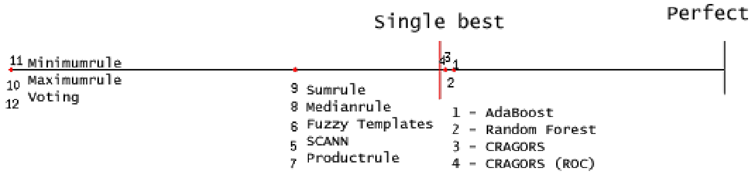


Fig. 8. Ranking of different algorithms for combining segmentation results of microscopy images

The good performance of the proposed method for the combination of detection results raises the question, whether other typical classification problems from the UCI machine learning repository can benefit or not. The following classifiers have been selected and the results of their combination have been evaluated:

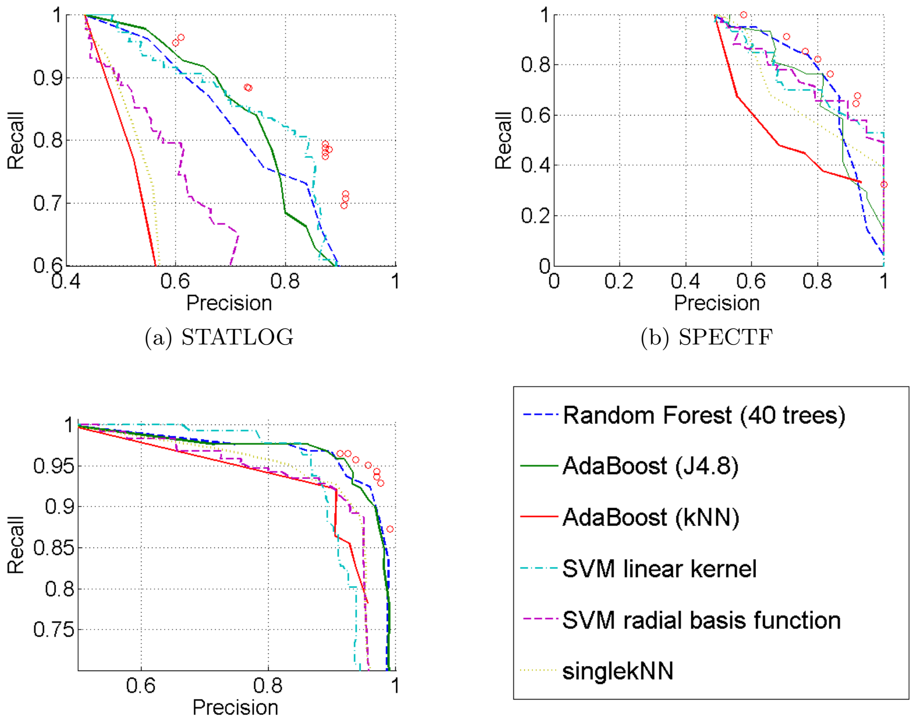
1. Linear kernel SVM
2. Radial basis function kernel SVM
3. AdaBoost with decision trees (J4.8)
4. AdaBoost with kNN (k=1)
5. Random forest classifier
6. kNN classifier (k=10)

The obtained combining rules as well as corresponding precision and recall values are listed in Tab. 1.

Table 1. Combining rules, precision P, and recall R for selected datasets from UCI Machine Learning Repository

Dataset	Combining Rule	P	R
Statlog 1	$(((\text{linSVM}_2 \cap \text{RF40}_1) \cap \text{boostedJ4.8}_2) \cup \text{rbfSVM}_3)$	0.87	0.79
Statlog 2	$((((\text{linSVM}_1 \cup \text{RF40}_2) \cap \text{boostedJ48}_1) \cup \text{boostedkNN}_3) \cup \text{rbfSVM}_1)$	0.96	0.6
SPECTF	$(((\text{rbfSVM}_2 \cap \text{RF40}_2) \cup \text{linSVM}_2) \cap \text{boostedJ48}_1)$	0.86	0.74
WDBC	$(((\text{rbfSVM}_2 \cap \text{singlekNN}_2) \cup \text{RF40}_2) \cap \text{linSVM}_1)$	0.96	0.95

Fig. 9 shows the precision recall curves of the tested classifiers for the Statlog, WDBC, and SPECTF datasets. The operating points of the combined methods are indicated by circles. Each circle corresponds to a different weighting of the false positive and false negative errors. Therefore, for each circle a different combination rule for the classifiers has been calculated.



(a) STATLOG

(b) SPECTF

(c) WDBC

Fig. 9. Precision/recall curves for classification performance on UCI dataset and operating points of combined classifiers (circles)

For all tested datasets the fusion algorithm improves the classification performance. This is worth mentioning because the individual classifiers are classifier ensembles such as random forests as well as adaptive boosted classifiers. It shows that the method is not limited to fusion of object detection algorithms. Its basic principle is the careful selection of operating points prior to fusion to ensure a sufficient level of diversity or compliance of the individual detectors or classifiers.

6 Conclusion

In this paper, a novel approach to the combination of object detection algorithms has been presented. The selection of operating points allows the reconfiguration of individual object detectors. Additionally, redundant detectors can be excluded automatically, as a result the time for the matching step between the result sets is greatly reduced. This is an important feature because matching is required for all combining methods. Depending on the detection algorithm, additional reductions of runtime are possible. Methods such as Viola and Jones cascade of boosted features [25] benefit from sensitive operating points because only few levels of the cascade must be evaluated. The processing time of other algorithms such as LKT [17] decreases, if only the precisely trackable objects have to be detected. Segmentation is a common preprocessing step for object detection. Showing that the proposed algorithm improves the segmentation of microscopy images of phytopathogenic fungi illustrates that combining different algorithms is beneficial at the different processing levels of an object detection chain.

References

1. Aeberhard, M.: Object existence probability fusion using dempster-shafer theory in a high-level sensor data fusion architecture. In: IEEE Intelligent Vehicles Symposium, pp. 770–775 (2011)
2. Al-Ani, A., Deriche, M.: A new technique for combining multiple classifiers using the dempster-shafer theory of evidence. Journal of Artificial Intelligence Research 17, 333–361 (2002)
3. Baum, T., Navarro-Quezada, A., Knogge, W., Douchkov, D., Schweizer, P., Seiffert, U.: Hypharea - automated analysis of spatiotemporal fungal patterns. Journal of Plant Physiology 168(1), 72–78 (2011)
4. Chen, X., Harrison, R., Zhang, Y.-Q.: Genetic fuzzy fusion of svm classifiers for biomedical data. In: IEEE Congress on Evolutionary Computing (2005)
5. Chen, X., Li, Y., Harrison, R., Zhang, Y.-Q.: Type-2 fuzzy logic-based classifier fusion for support vector machines. Applied Soft Computing 8(3), 1222–1231 (2008); Forging the Frontiers - - Soft Computing
6. Cheng, F., Huang, S., Ruan, S.: Advanced motion detection for intelligent video surveillance systems. In: ACM Symposium on Applied Computing (2010)
7. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes (VOC) Challenge. International Journal of Computer Vision 88, 303–338 (2010)
8. Greenacre, M.J.: Theory and application of correspondence analysis. Academic Press, London (1984)
9. Kang, S., Park, S.: A fusion neural network classifier for image classification. Pattern Recogn. Lett. 30(9), 789–793 (2009)
10. Khan, F.S., Anwer, R.M., van de Weijer, H., Bagdanov, A.D., Vanrell, M., Lopez, A.M.: Color Attributes for Object Detection. In: CVPR (2012)
11. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(3), 226–239 (1998)
12. Knauer, U., Meffert, B.: Fast computation of region homogeneity with application in a surveillance task. In: ISPRS Technical Commission V Symposium. ISPRS (2010)
13. Kuncheva, L.I.: 'Fuzzy' vs 'non-fuzzy' in combining classifiers designed by boosting. IEEE Transactions on Fuzzy Systems 11(6), 729–741 (2003)
14. Lee, D.: Multisensor fusion-based object detection and tracking using active shape model. In: Digital Information Management, pp. 108–114 (2011)
15. Leung, K.T., Parker, D.S.: Empirical comparisons of various voting methods in bagging. In: 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2003)

16. Liu, N., Dellandrea, E., Zhu, C., Bichot, C.-E., Chen, L.: A Selective Weighted Late Fusion for Visual Concept Recognition. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part III. LNCS, vol. 7585, pp. 426–435. Springer, Heidelberg (2012)
17. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision, pp. 674–679 (1981)
18. Merz, C.J.: Using correspondence analysis to combine classifiers. Machine Learning 36(1-2), 33–58 (1999)
19. Sevenson, A.: Separating Axis Theorem (SAT) Explanation. Internet (2009)
20. Song, Z.: Contextualizing object detection and classification. In: IEEE Conference on Computer Vision and Pattern Recognition (2011)
21. Spinello, L., Arras, K.O.: Leveraging rgb-d data: Adaptive fusion and domain adaption for object detection. In: IEEE Conference on Robotics and Automation (2012)
22. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: IEEE Computer Vision and Pattern Recognition, vol. 2, p. 200. IEEE (1999)
23. Suja, T.B.: Fusion based object detection. In: National Conference on Communications, pp. 1–3 (2010)
24. Teschner, M.: Algorithmen und Datenstrukturen Bereichsbume. Foliensatz (2011)
25. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 511–518 (2001)
26. Xu, L., Kryzak, A., Suen, C.V.: Methods of Combining Multiple Classifiers and Their Application to Handwriting Recognition. IEEE Transactions on Systems, Man, and Cybernetics 22(3), 418–435 (1992)