Dynamic Belief Fusion for Object Detection

Abstract

A novel approach for the fusion of detection scores from disparate object detection methods is proposed. In order to effectively integrate the outputs of multiple detectors, the level of ambiguity in each individual detection score (called “uncertainty”) is estimated using the precision/recall relationship of the corresponding detector. The proposed fusion method, called Dynamic Belief Fusion (DBF), dynamically assigns basic probabilities to propositions (target, non-target, uncertain) based on confidence levels in the detection results of individual approaches. A joint basic probability assignment, containing information from all detectors, is determined using Dempster’s combination rule, and is easily reduced to a single fused detection score. Experiments on ARL and PASCAL VOC 07 datasets demonstrate that the detection accuracy of DBF is considerably greater than conventional fusion approaches as well as state-of-the-art individual detectors.

1. Introduction

“The more information, the better.” This saying is true so long as the additional information is complementary to the existing knowledge base. Making sense of large amounts of disparate data from different sources often requires a complex filtering process to eliminate redundancy, ignore noise, resolve conflicting information, uncover patterns, and ultimately obtain an awareness of the data that is greater than the sum of its parts. This leads to a critical question: How does one optimally combine information?

We bring this question to the task of object detection/recognition, one of the leading directions of study in the computer vision community. Over the past several years, a number of advanced object detection algorithms have been introduced. However, few attempts have been made to adaptively combine existing algorithms in such a way that best exploits information about the objects of interest and prior performance of the detector. We intend to effectively integrate multiple state-of-the-art object detection approaches into a single fusion system; however, this is very challenging due to the unique structures of individual algorithms. To address this issue, a high-level fusion framework is needed; one that can effectively integrate detections regardless of the structural differences of the individual algorithms. When fusing the results of multiple approaches, the concept of uncertainty is critical. Uncertainty is the condition in which a precise answer is not ascertainable due to limited knowledge, such as data incompleteness, sensor ambiguity, etc. Bayesian fusion, a widely used approach, measures uncertainty with regard to a set of given hypotheses by associating a probability distribution (called “likelihood”) with each hypothesis. However, one major drawback of Bayesian fusion is that it inherently cannot provide a way to express ignorance about observations. In other words, the Bayesian method is forced to provide basic probabilistic estimates even for the cases where the estimate should not be made due to lack of information. In order to interpret this type of uncertainty, G. Shafer de-
developed a general theory of evidential reasoning, now known as Dempster-Shafer theory (DST) [3, 16]. DST introduces the concept of compound hypotheses by using a power set consisting of all subsets of hypotheses. A basic probability is assigned to individual compound hypotheses, allowing the expression of a level of ignorance about observations. DST also employs Dempster’s combination rule [3], which calculates the joint belief of a certain hypothesis by combining the basic probability distributions of the compound hypotheses from different information sources.

In the present work, we introduce a novel approach to enable the advanced fusion of multiple object detection approaches under the DST framework. We refer to this method as Dynamic Belief Fusion (DBF) as opposed to the conventional DST, which assigns fixed belief to the hypotheses. As in Figure 1, DBF can dynamically assign the basic probability to individual hypotheses as a function of confidence levels of the decisions about observations from individual approaches. These basic probabilities are estimated from the precision-recall (PR) curve (a common evaluation tool in object detection) of each approach in a validation step, whose values vary dynamically with the detection scores of given detections. In this work, the constituent hypotheses include target, non-target, and uncertainty (target or non-target) under the DST framework. In estimating basic probabilities using the precision/recall relationship, we can reasonably assume that the false detections are caused by two factors: (i) the lack of the background representation in training and (ii) inherent similarity between target and background. We hypothesize that the first factor is associated with the uncertainty hypotheses and the second is linked to the non-target hypotheses. We then use the notion of a best possible detector, whose detections do not contain false positives caused by the first factor to effectively estimate the basic probabilities associated with the two factors and the two corresponding hypotheses (see Figure 3). The proposed DBF approach is described in section 3.3.

The proposed DBF method is evaluated using ARL [17] and PASCAL VOC 07 [4]. DBF is compared to other well-known fusion methods such as Bayesian fusion, Weighted Sum, and DST with fixed basic probability assignments. Results demonstrate that DBF outperforms all individual computer vision methods, as well as the other fusion methods. Moreover, DBF is the only approach able to consistently outperform the best individual method.

2. Related Works

In this section, we focus on the fusion of multiple approaches rather than the fusion of different features. Lan et al. [10], Wang et al. [13], Fernando et al. [6] and Natarajan et al. [15] proposed the fusion of different types of features but their proposed methods cannot solve the fusion of separate approaches with different principles or structures.

Work based on the fusion of multiple approaches can be split into two categories; (i) building a joint model by integrating multiple approaches and (ii) fusing the output of multiple approaches. Kwon and Lee [8] integrated multiple sample-based tracking approaches into one compound tracker through an interactive Markov Chain Monte Carlo (iMCMC) framework. The compound tracker achieved high efficiency by exchanging information among the approaches and, thus reducing the number of samples. The authors proposed another novel tracking framework called a “visual tracker sampler” [9] to fuse the multiple trackers which are sampled from tracker space (appearance models, motion models, state representation types, and observation types) using the Markov Chain Monte Carlo (MCMC) method. For event detection, Wu et al. [20] employed detectors of different modality (concept, text, speech) and construct relationships among the modes and the event by utilizing natural language processing. In each of the multiple approach integration methods presented above, either the approaches must be conditionally independent, or their dependencies must be explicitly modeled. However, in general, modeling the dependencies among multiple approaches built on different principles is infeasible.

In cases where modeling dependency among multiple approaches is not possible, fusion can be performed over their outputs. Bailer et al. [11] collected trajectories from multiple tracking algorithms and computed one fused trajectory using dynamic programming to improve accuracy, trajectory continuity, and smoothness. However, this fusion method is only appropriate for multiple approaches aiming at the same goal and based on the same principle. Kim et al. [7] and Liu et al. [12] used weighted-sum methods to fuse multiple types of data. The approach by Kim et al. learned the weights by employing the quadratic relaxation method to address the non-convex problem. Liu et al. argued that existing methods generally used fixed fusion weights for constituent classifiers, and thus failed to optimally determine the fusion weight for the individual classifiers. They constructed a graph with labeled data and with (unlabeled) test data, and propagated fusion weights from the weights of the labeled data. A large amount of labeled data is critical to taking advantage of these weight propagations, and the computational complexity of constructing the graph and performing the propagation process increases with data size. Nevertheless, their assertion that fusion should be optimally applied to the specific test data is important. Ma and Yuen [13] and Liu et al. [11] used likelihood functions to model the prior information of multiple approaches and employed Bayesian fusion based on these likelihoods. However, Bayesian fusion does not quantify detector ignorance, as discussed in the previous section. As opposed to Bayesian fusion, DST can handle output ambiguity to ignore potentially misleading information from
making decision based on ambiguous observations. DBF also resolves the issue of fixed detector performance, argued by Liu et al. [12], by assigning basic probabilities that vary as a function of the confidence scores of each test detection.

3. Dynamic Belief Fusion

3.1. Overview of the Fusion of Detectors

The objective of object detection is to localize each target object in an image. Image databases were split into three sets: training, validation, and test. Figure 2 illustrates the fusion process of the proposed DBF algorithm. First, individual object detectors are learned over the training set. Subsequently, prior information in the form of precision/recall used for dynamic belief fusion is computed using the validation set. Lastly, evaluation of the trained detectors is performed with the test set. The complete fusion procedure is performed in the following three steps:

Assign labels to each window and compute basic probability distribution (validation set): In the validation step, detection windows are compared with the actual ground truth, and basic probabilities are assigned according to the PR performance of each individual detector. Note that each of the detectors examines multiple windows, and the windows coming from different detectors are normally different in size, location, and number. The windows provided by each detection algorithm are first compared with ground truth (annotated windows containing the objects of interest) and categorized as positive/negative/undecided. Any window that has an intersection-over-union overlap (PASCAL VOC criteria [4]) of greater than 0.5 with a ground truth window is assigned a positive label. If there is no overlap between a detection window and a ground truth window, the detection is assigned a negative label. The remaining detection windows, which have minor overlaps with ground truth data are labeled “undecided.” By matching these labels to the window’s confidence scores (provided by the detector), precision and recall are computed. The reliability of each detector will be taken into account in the final fusion step by using precision and recall to create a basic probability assignment (BPA).

Collect overlapped detection windows and construct their detection vector (test set): For each window in the test set, we collect the windows from other detectors that significantly overlap that window (see Figure 2). Let $d^i_j, i = 1, 2, \cdots, K, j = 1, 2, \cdots, W_i$ be the $j^{th}$ window of the $i^{th}$ detector and be associated with its bounding box $b^i_j$ and detection confidence $c^i_j$. $K$ is the number of detectors and $W_i$ is the number of windows of the $i^{th}$ detector. Two windows $d^i_j$ and $d^k_l$, $i \neq k$ are considered significantly overlapped if the intersection-over-union overlap of their bounding boxes is greater than 0.5. A $K$-dimensional detection vector $c = [c^1_{j_1}, c^2_{j_2}, \cdots, c^K_{j_K}]$ is constructed, consisting of the window’s detection confidence score and those of the
overlapped windows. If multiple windows from the same
detector overlap the candidate window, the maximum conﬁdence
score between them is used. If no overlaps exist for a
particular detector, the corresponding conﬁdence score bin
is ﬁlled by a value of negative inﬁnity.

Fusion of the windows (test set): In the test step, fusion is
performed over the results of the detectors using the DBF as
described in section 3.3 (DST is described in section 3.2). After
rescoring all windows through fusion (darker windows in the
“fusion” step of Figure 2 indicate higher conﬁdence), non-maximum suppression is applied to merge windows
whose intersection over union overlap is greater than 0.5. The ﬁnal output of the fusion procedure is a conso-
dilated set of windows, each with a fused conﬁdence score.

3.2. Dempster-Shafer Theory

Dempster-Shafer theory is a general framework for rea-
soning about evidence with uncertainty that incorporates
aspects of probability and possibility theory. It has
been widely used in the context of information fusion.

Let \( X \) be a universal set consisting of \( M \) exhaustive and
mutually exclusive propositions, i.e. \( X = \{1, 2, \ldots, M\} \).
The power set \( 2^X \) is the set of all subsets of \( X \), including
the empty set \( \emptyset \). DST assigns a basic probability in the range
\([0, 1]\) to each element of the power set \( 2^X \). A function
defined as \( m : 2^X \rightarrow [0, 1] \) is called a basic probability assign-
ment (BPA). A BPA has two properties; (i) \( m(\emptyset) = 0 \) (the
mass of the empty set is zero) and (ii) \( \sum_{A \in 2^X} m(A) = 1 \)
(the BPA values of the members of the power set sum to
one).

From the BPAs, the belief function \( bel(A) \) for a set \( A \)
can be deﬁned as the sum of all masses which are subsets
of the set of interest:

\[
bel(A) = \sum_{B | B \subseteq A} m(B). \tag{1}
\]

This represents the information in direct support of \( A \). Let
\((bel_1, m_1)\) and \((bel_2, m_2)\) denote two pairs of a belief func-
tion and its corresponding BPA. Dempster’s combination
rule can be applied to calculate a joint BPA from separate
BPAs. Under the condition that the evidence from each pair
is independent of the other, Dempster’s combination rule
defines a joint BPA \( m = m_1 \oplus m_2 \), which represents the
combined effect of \( m_1 \) and \( m_2 \), i.e.,

\[
m(A) = m_1 \oplus m_2(A) = \frac{1}{N} \sum_{X \cap Y = A, A \neq \emptyset} m_1(X)m_2(Y), \tag{2}
\]

where \( N = \sum_{X \cap Y \neq \emptyset} m_1(X)m_2(Y) \) and \( X \) and \( Y \) are
subsets of \( 2^X \). \( N \) is a measure of the amount of any mass whose
common evidence is not the null set. Dempster’s rule can
be extended for multiple pieces of evidence (e.g., multiple
detectors) using the associative and commutative properties
of BPAs (i.e. \( m = m_1 \oplus m_2 \oplus \cdots \oplus m_K \)) with the following
formula:

\[
m(A) = \frac{1}{N} \sum_{X_1 \cap X_2 \cap \cdots \cap X_K = A} \prod_{i=1}^{K} m_i(X_i), \tag{3}
\]

where \( N = \sum_{X_1 \cap X_2 \cap \cdots \cap X_K \neq \emptyset} \prod_{i=1}^{K} m_i(X_i) \).

3.3. Dynamic Belief Fusion based on Dempster-
Shafer Theory

In object detection of a two-class problem, the universal
set \( X \) is deﬁned as \( \{T, \neg T\} \) and thus its power set is ex-
pressed as \( \{\emptyset, T, \neg T, \{T, \neg T\}\} \), where \( T \) is a target
proposition and \( \neg T = X - T \) is the non-target proposition, a
compound hypothesis. \( \{T, \neg T\} \) in the power set represents
a detection ambiguity denoted by \( U \) (uncertainty). In order
to assign the belief based on prior detection performance of
individual approaches, we calculate precision and recall,
widely used metrics for detection performance in object
detection, on the validation set and use them to compute BPAs
for \( T, \neg T, \) and \( U \). Precision and recall are deﬁned as

- Precision: \( p = \frac{N_{TP}}{N_{TP} + N_{FP}} \)
- Recall: \( r = \frac{N_{TP}}{N_{tobj}} \)

where \( N_{TP} \) and \( N_{FP} \) are a number of true and false posi-
tive, respectively. \( N_{tobj} \) is a number of target objects in a
given test set. Note that \( N_{TP} \) and \( N_{FP} \) vary according
to a classiﬁcation threshold. We can reasonably assume
that false positives are caused by two factors: (i) the ﬁrst
factor comprises false positives occurring due to the lack
of the representation of background images in training and
(ii) the second factor contains false positives arising from
inherent similarity between target and background. If the
negative training imageset covers all background images in
the world completely, the ﬁrst factor will disappear while
the second factor still remains regardless of the training pro-
cess. If the detector is trained over the complete negative set
representing all background images, its precision \( p_{pf} \) can be
expressed as

\[
p_{pf} = \frac{N_{TP}}{N_{TP} + N_{FP,II}} > \frac{N_{TP}}{N_{TP} + N_{FP}} = p, \tag{4}
\]

where \( N_{FP,II} \) is a number of the false positives caused by
the second factor. The difference between the precision of
\( p \) and \( p_{pf} \) measures a degree of the incompleteness of
the training process, which is equivalent to the ambiguity of
its performance. To calculate this difference, we introduce the
ﬁctional best possible detector whose precision is modeled by
\( \hat{p}_{pf}(r) \) deﬁned as

\[
\hat{p}_{pf}(r) = 1 - r^n, \tag{5}
\]
where $r$ is a recall. As $n$ approaches infinity, it becomes the perfect detector (i.e. no false positives). Thus, the value of $n$ for the best possible detector should be much less than infinity, which is supported by Figure 6. As shown in Figure 3 for a given level of recall $r$, $m(T)$ is assigned as the precision, while $m(U)$ is defined by the difference between $\hat{p}_T(r)$ and the $m(T)$. The remaining fraction of precision $(1 - \hat{p}_T(r))$ which represents the second type of false positives, evidence that only supports the $\neg T$ proposition, is assigned to $m(\neg T)$. The basic probability distribution varies as a function of the recall $r$ so we call this dynamic basic probability assignment incorporating the dynamic belief fusion in which basic probabilities are dynamically assigned to the corresponding hypotheses. Fusion of the detections of multiple individual detectors is achieved by collecting the basic probability distribution of individual detectors and computing joint belief of “target” and “non-target” proposition, $\text{bel}(T)$ and $\text{bel}(\neg T)$, via Dempster’s rule to combine the distributions. The overall fusion confidence score is given by $s = \text{bel}(T) - \text{bel}(\neg T)$.

4. Experiments

4.1. Evaluation Setting

Individual Detectors: Five object detectors with unique detection structures whose codes are readily available online were selected: two SVM-based detectors incorporating both HOG [2] and Dense SIFT [11], Deformable-Part Model (DPM) [5], Transductive Annotation by Graph (TAG) [19], and exemplar SVM [13]. All of these detectors are built on an SVM formulation used as a final classification tool. Given data, the sign of a detector’s SVM confidence is the decision (i.e. observation is a target object if the score is positive) while its value indicates a degree of confidence about the decision. The five selected detectors use different feature extraction methods and different principles of detecting objects of interest.

Baselines (Fusion): As a baseline, we used two conventional fusion approaches: Weighted Sum (WS) and Bayesian fusion. The WS approach finds weights of detection scores that maximize the product of a weight vector $w$ and the detection vector of detector scores $c$, $f_{WS}(c) = w^T c$. $w$ is learned through linear SVM optimization. For Bayesian fusion, we use a naive Bayesian model assuming that all the approaches are independent of each other. In other words, the joint likelihood can be decomposed as the product of the likelihoods of each detector, while the posterior is expressed as the product of the prior and the joint likelihood (i.e. Bayes’ rule).

To demonstrate the advantages of dynamic basic probability assignment in the proposed DBF, we also implement a regular DST fusion method that employs only static basic probability assignment [21] using the precision value in each detector’s precision-recall curve that corresponded to a recall of 0.2.

4.2. Evaluation of ARL Dataset

The Army Research Lab (ARL) image dataset was originally created for the purpose of analyzing human performance in Rapid Serial Visual Presentation (RSVP) [17] tasks, but is also applicable to object detection tasks. (In future work, we plan to integrate computer vision-based object detection with human decisions.) The dataset contains 3000 images of both indoor and outdoor scenes, 1438 images of which contain at least one object-of-interest. The target objects include chair, container, door, poster, and stair. Figure 4 displays several example images of all five objects as well as background images. The number of images in the ARL dataset is relatively small compared to that of other benchmark datasets, such as PASCAL VOC 07 (0.239 for DPM) and ImageNet, but, with regard to the mean average precision (mAP), the ARL dataset (0.253 for DPM) is not considerably less challenging compared to the benchmark datasets.

The proposed DBF algorithm was evaluated on the ARL dataset and its average precision (AP) was compared to that of four individual detection algorithms (the HOG-SVM detector was not utilized on the ARL dataset.) and three other “baseline” fusion methods (WS, Bayes, and DST) for each object class. Results are shown in Table 1.
Figure 6. Comparison of fusion performance with respect to the various ideal best possible detector.

Table 1. AP on the ARL dataset.

|               | chair | contr | door | postr | stair | mAP |
|---------------|-------|-------|------|-------|-------|-----|
| DSIFT         | .143  | .037  | .073 | .143  | .061  | .091|
| TAG           | .045  | .123  | .159 | .066  | .008  | .080|
| ESVM          | .125  | .318  | .150 | .236  | .122  | .190|
| DPM           | .188  | .396  | .194 | .342  | .143  | .253|
| WS            | .029  | .108  | .034 | .009  | .006  | .037|
| Bayes         | .095  | .154  | .132 | .086  | .016  | .096|
| DST           | .195  | .307  | .217 | .288  | .129  | .227|
| DBF           | .292  | .418  | .313 | .352  | .160  | .307|

curves for four of the object classes are also shown (see Figure 5). Both mAP and ROC performance metrics show that DBF outperformed the baseline fusion algorithms as well as the individual detectors. From these results, we can infer the limits of the baseline metrics: for WS, the low-dimensional weight vector cannot classify objects perfectly, and in Bayesian fusion, the results are significantly degraded by poorly performing detectors.

Figure 5 shows that mAP varies as a function of \( n \), a parameter related to the best possible detector for every object categories in the ARL dataset. It is clear that the best value of \( n \) in terms of mAP, depends on the particular object categories. Note that the notional perfect detector \( (n = \text{inf}) \) underperforms other best possible detectors in every object category. This result shows that splitting the false positive images into “uncertain” and “non-target” categories is actually beneficial, supporting the assumption that modeling false detections based on the aforementioned second factor is needed. Figure 8 shows some examples from the proposed DBF in three object categories, chair, poster, and stair on ARL dataset.

4.3. Evaluation of PASCAL VOC 07 Dataset

The fusion and individual detection methods were also evaluated on the PASCAL VOC 07 dataset. Examples of fusion in three object categories—person, aeroplane, and car—are shown in Figure 9.

The mAP of each individual detector and fusion method is reported in Table 2. To evaluate fusion on the PASCAL VOC 07 dataset, five individual detectors (DSIFT-SVM, HOG-SVM, TAG, Exemplar SVM, and DPM) were selected and fusion of their detection results was conducted. In terms of mAP, the proposed DBF outperforms all individual approaches and fusion baselines. Unlike the ARL dataset, the results of DPM are comparable to DBF. DPM shows the best results in 8 out of 20 object categories while DBF results in the best performance for the rest. In the categories which DPM outperforms DBF, the mAP of the categories (0.375) was relatively high compared to that of other categories (0.152). Because the mAP values of other detectors were similar, we believe that the weak performance in those categories may be caused by insufficient complementary information among the five individual approaches. Figure 7 shows ROC curves of four object categories out of 20: bird, cat, diningtable, and train. In the ‘train’ ROC plot, DBF exhibits better performance than DPM after approximately 0.5 recall, despite the AP of DBF being lower. The DST-based fusion approaches generally perform well at high levels of false positive per images (FPPI).

In order to investigate whether complementary information is provided by each detector, mAP was calculated while varying the number of individual detectors used in the fusion. For each combination number \( m \), detectors with the \( m \) highest mAP were selected. Results, shown in Table 3 for both the ARL and PASCAL datasets, illustrate that performance consistently increases with number of detectors. The final row corresponds to the maximum number of combined
Figure 7. ROC curves on the PASCAL VOC07 dataset [4]: The ROC curves for bird, cat, diningtable, and horse categories are plotted. DBF is evaluated with $n$ of 4.

|       | aero | bike | bird | boat | bottle | bus | cat | cat | chair | cow | table | dog | horse | mbik | pers | plant | sheep | sofa | train | tv | mAP |
|-------|------|------|------|------|--------|-----|-----|-----|-------|-----|-------|-----|-------|------|------|-------|-------|------|------|----|-----|
| DSHIFT | .079 | .023 | .017 | .004 | .002   | .078| .113| .141| .005  | .097 | .108  | .126| .038  | .037 | .076 | .002  | .057  | .105 | .122 | .026| .063 |
| HOG    | .093 | .005 | .001 | .061 | .001   | .002| .001 | .001| .001  | .001 | .002  | .038| .092  | .005 | .001 | .091  | .092  | .092 | .024 | .024| .030 |
| TAG    | .019 | .050 | .009 | .002 | .002   | .027| .023 | .079| .002  | .006 | .056  | .030| .019  | .085 | .050 | .002  | .017  | .020 | .014 | .026| .026 |
| ESVM   | .048 | .167 | .002 | .092 | .047   | .145| .155| .094| .024  | .068 | .007  | .093| .200  | .182 | .106 | .005  | .096  | .158 | .115 | .093| .093 |
| DPM    | .231 | .500 | .036 | .099 | .162   | .388| .451| .153| .120  | .172 | .129  | .106| .463  | .375| .346 | .109  | .109  | .144 | .353 | .333| .239 |
| WS     | .002 | .107 | .006 | .002 | .001   | .003| .042 | .011| .005  | .004 | .007  | .011| .008  | .020 | .019 | .003  | .004  | .003 | .005 | .013| .013 |
| Bayes  | .076 | .197 | .016 | .024 | .019   | .041| .195| .059| .008  | .010 | .043  | .028| .092  | .083| .186 | .011  | .067  | .034 | .066 | .096| .068 |
| DST    | .196 | .196 | .012 | .099 | .095   | .193| .269| .089| .100  | .075 | .148  | .088| .198  | .187| .232 | .020  | .117  | .117 | .187 | .129| .137 |
| DBF    | .277 | .453 | .040 | .114 | .151   | .382| .375| .207| .120  | .175 | .189  | .146| .440  | .376| .303 | .116  | .115  | .145 | .325 | .301| .250 |

Table 2. AP on the PASCAL VOC 07 dataset.

Table 3. Comparison of fusion performance with respect to the combination of multiple detectors.

| # of detectors | ARL | PASCAL |
|----------------|-----|--------|
| 2              | .260| .226   |
| 3              | .305| .238   |
| 4              | .248|        |
| Maximum        | .307| .250   |

detectors (4 for ARL, 5 for PASCAL). The increased performance supports the use of the proposed fusion method, as, on average, even the weakest detector provides information complementary to others.

5. Conclusions

An effective fusion method, referred to as Dynamic Belief Fusion (DBF), is proposed to integrate more accurately detection scores from multiple object detectors. This method defines uncertainty as the ambiguity of each detection output, measured using the corresponding detector’s precision/recall estimated based on a validation set. Most existing fusion approaches based on Dempster-Shafer Theory (DST) incorporate a fixed level of uncertainty of certain states (e.g. target and non-target) in basic probability assignments and estimates the joint belief for target and non-target by combining the basic probability distributions of multiple detectors. In contrast, the proposed DBF incorporates the prior information of different detectors in the form of the precision/recall relationship in such a way that varying degrees of belief for individual hypotheses are estimated from the dynamic combination of the prior performance of different approaches and the current confidence levels of the detection outputs obtained from given observations. This dynamic combination of the prior information and the current detector confidences for given observations make the proposed approach a powerful way of fusing information from individual detectors.

In terms of mAP, DBF is the only fusion approach that consistently outperforms all the individual detectors. In contrast, the other three baseline fusion approaches could not provide good performance when the number of false positives is very small, as opposed to DBF. This is due mainly to their inherent limitation of effectively exploiting the uncertainty associated with individual hypotheses.

As shown in Figures 5 and 7, most of the baseline fusion methods, however, begin to outperform the individual detectors when the number of false positives per image becomes larger than about 4. Overall, the evaluation demonstrates the robustness of the proposed dynamic belief estimation being able to consistently generate superior performance over the entire range of the false positives.
Figure 8. Examples of objects for which high fusion scores are generated on the ARL dataset. Windows colored by red, blue, green, and yellow are detected by DPM, DSIFT-SVM, TAG, and exemplar SVM, respectively. For each object category, two false detections marked by a thick red box are shown.

Figure 9. Examples of objects for which high fusion scores are generated on the PASCAL VOC 07 dataset. Windows colored by red, blue, pink, green, and yellow are detected by DPM, DSIFT-SVM, HOG-SVM, TAG, and exemplar SVM, respectively. For each object category, two false detections marked by a thick red box are shown.
References

[1] C. Bailer, A. Pagani, and D. Stricker. A superior tracking approach: Building a strong tracker through fusion. In ECCV, 2014.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005.

[3] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. The Annals of Mathematical Statistics, 38(2), 1967.

[4] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html.

[5] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. PAMI, 32(9), 2010.

[6] B. Fernando, E. Fromont, D. Muselet, and M. Sebban. Discriminative feature fusion for image classification. In CVPR, 2012.

[7] T. H. Kim, H. S. Lee, and K. M. Lee. Optical flow via locally adaptive fusion of complementary data costs. In ICCV, 2013.

[8] J. Kwon and K. M. Lee. Visual tracking decomposition. In CVPR, 2010.

[9] J. Kwon and K. M. Lee. Tracking by sampling trackers. In ICCV, 2011.

[10] X. Lan, A. J. Ma, and P. C. Yuen. Multi-cue visual tracking using robust feature-level fusion based on joint sparse representation. In CVPR, 2014.

[11] C. Liu, J. Yuem, and A. Torralba. SIFT flow: Dense correspondence across scenes and its applications. PAMI, 33(5), 2011.

[12] D. Liu, K. Lai, G. Ye, M. Chen, and S.-F. Chang. Sample-specific late fusion for visual category recognition. In CVPR, 2013.

[13] A. J. Ma and P. C. Yuen. Reduced analytical dependency modeling for classifier fusion. In ECCV, 2012.

[14] T. Malisiewicz, A. Gupta, and E. A. A. Ensemble of exemplar-SVMs for object detection and beyond. In ICCV, 2011.

[15] P. Natarajan, S. Vitaladevuni, U. Park, V. Wu, S. and-Manohar, X. Zhuang, S. Tsakalidis, R. Prasad, and P. Natarajan. Multimodal feature fusion for robust event detection in web videos. In CVPR, 2012.

[16] G. Shafer. A mathematical theory of evidence. Princeton University Press, 1976.

[17] J. Touryan, G. Apker, B. Lance, S. Kerick, A. Ries, and K. McDowell. Estimating endogenous changes in task performance from eeg. Neuroprosthetics, 8(15), 2014.

[18] H. Wang, F. Nie, and H. Huang. Heterogeneous visual features fusion via sparse multimodal machine. In CVPR, 2013.

[19] J. Wang, T. Jebara, and S.-F. Chang. Graph transduction via alternating minimization. In ICML, 2008.

[20] S. Wu, S. Bondugula, F. Luisier, X. Zhuang, and P. Natarajan. Zero-shot event detection using multi-modal fusion of weakly supervised concepts aerial reconstructions via probabilistic data fusion. In CVPR, 2014.

[21] L. Xu, A. Kryzak, and C. Y. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. IEEE Trans. on Sys., man, and cyb, 22(3), 1992.