

# Chapter 19

## Entanglement and Differentiable Information Gain Maximization

A. Montillo, J. Tu, J. Shotton, J. Winn, J.E. Iglesias, D.N. Metaxas,  
and A. Criminisi

Decision forests can be thought of as a flexible optimization toolbox with many avenues to alter or recombine the underlying architectural components and improve recognition accuracy and efficiency. In this chapter, we present two fundamental approaches for re-architecting decision forests that yield higher prediction accuracy and shortened decision time.

The first is entanglement, *i.e.* using the learned tree structure and intermediate probabilities computed in nodes closer to the root to affect the training of other nodes deeper in the trees. Unlike more conventional classifiers which assume that all data points (even those neighboring in space or time) are IID, the entanglement approach learns semantic correlation in non IID data. To demonstrate, we build an entangled decision forest (EDF) that exploits spatial correlation in human anatomy by simultaneously labeling voxels in computed tomography (CT) scans into 12 anatomical structures.

The second contribution is the formulation of information gain as a function that is differentiable with respect to the parameters of the split node weak learner. This provides increased confidence and accuracy of maximum margin boundary localization and reduces classification time by using a few, shallow trees. We further extend the method to incorporate training label confidence, when available, into the information gain maximization. Due to bagging and random feature subset selection, we can retain decision forest virtues such as resiliency to overfitting. To demon-

---

A. Montillo (✉) · J. Tu  
General Electric Global Research, Niskayuna, NY 12309, USA

J. Shotton · J. Winn · A. Criminisi  
Microsoft Research Ltd, Cambridge, UK

J.E. Iglesias  
Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

D.N. Metaxas  
Rutgers, Piscataway, NJ, USA

strate, we build a gradient ascent decision forest (GADF) that tracks visual objects in videos. For both approaches, superior accuracy and computational efficiency is shown in quantitative comparisons with state of the art algorithms.

## 19.1 Introduction

As discussed in Part I of this book, decision forests are a flexible framework for addressing diverse tasks, with many avenues to alter or recombine the underlying architectural components to improve accuracy and efficiency. In this chapter, we present two fundamental approaches for re-designing the decision forest. These lead to improved prediction accuracy, increased confidence and accuracy of maximum margin boundary localization, and reduced decision time and memory requirements for real world applications including semantic segmentation of 3D medical images and tracking objects in video.

## 19.2 Entangled Decision Forests

Our first approach, a re-architecting of decision forests, is the *entanglement* or sharing of information between the nodes in a decision forest. In entangled decision forests, the result of the binary tests applied at each tree node depends on the results of tests applied earlier during forest growth. This concept was first presented in [252] and later refined with context selectivity [250]. This chapter presents a more general exposition than reported previously, enabling the most broad interpretation and application.

Entanglement is *the use of the learned tree structure and intermediate probabilities associated with nodes in the higher levels of a tree to affect training of split nodes in deeper levels of the forest*. In its simplest incarnation one may think of entanglement as using the class posteriors of previously trained nodes as input feature into the training of subsequent nodes in the same tree.

A traditional assumption of many classifiers is that all data points (*e.g.* pixels in an image) are independent and identically distributed (IID). However, in many applications, this assumption is incorrect; many data points are in fact highly correlated and thus non IID. Entanglement automatically learns the semantic structural pattern of this correlation and encodes it in the features chosen during decision tree training. In practice, this correlation tends to occur over time, space or both. For example, in 3D medical image segmentation, human anatomy defines a canonical 3D configuration (correlation) over 3D space. In other cases, such as 4D medical scans, the correlation can be in both space and time (the fourth dimension). In entanglement, a tree node,  $j$ , at level,  $\ell$ , in the forest is constructed by designing *entanglement* features that exploit the uncertain partial contextual information learned (or at test time, inferred) in a correlation neighborhood by the previous  $\ell - 1$  levels of the forest (already trained). We call a forest that uses such features an entangled decision forest (EDF).

As an additional contribution, we randomly sample feature types and parameters from learned, non-uniform *proposal distributions* rather than from a uniform distribution used (implicitly) in previous decision forest research [5, 44, 77, 128, 212, 341, 411]. With this modification in place, the random draws from the proposal distribution select, with greater probability, the feature types and parameters that tend to be relevant for classification. As we will demonstrate, this allows for higher accuracy for the same number of features evaluated during training. Entanglement and learned proposal distributions allow faster training, and faster, more accurate prediction.

To illustrate entanglement, we discuss an example application where we wish to automatically segment a 3D Computed Tomography (CT) scan into its anatomical components such as the aorta, pelvis, and the lungs. We cast this task as a voxel classification problem which we solve via an EDF. In this case entanglement allows the class posteriors of voxels reaching nodes deep in the tree to depend directly from the intermediate posteriors attained higher up in the same tree. This improves accuracy and captures long-range *semantic* context. Previously, segmentation constraints in the form of semantic (*e.g.* anatomical) context have been applied, but these have required either a separate random field [342] or multi-pass processing [341, 375]; EDFs achieve this in one pass with no additional methods.

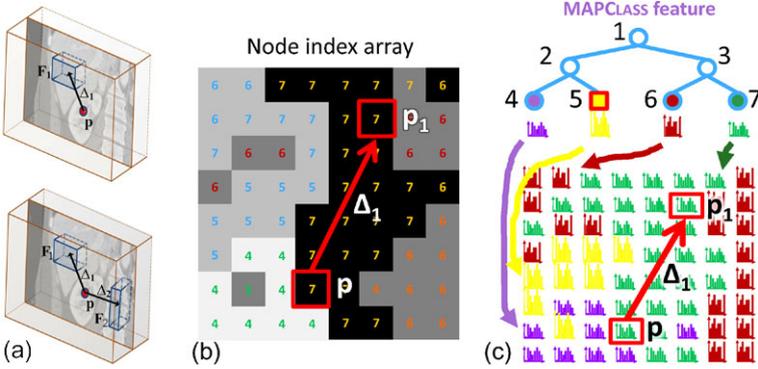
### 19.2.1 Entanglement Feature Design

We assume we are given a set,  $\mathcal{S} = \{(\mathbf{v}, c)\}$ , of voxels,  $\mathbf{v} = (i, \mathbf{p})$ , each consisting of its image intensity,  $i$ , (a measure of tissue density in the case of CT) voxel location  $\mathbf{p}$  and ground truth label,  $c$ . This set is formed from the collection of voxels from a group of training CT scans. Our goal is to infer the probability of each label for each voxel of unseen test scans.

Following the work in [78] we construct two types of long-range, context-aware feature. The first type captures “appearance context”, the latter are entangled and capture “semantic context”. See also Chap. 15. Details are explained next.

#### 19.2.1.1 Appearance Features

Using the intensity image,  $J$ , we construct intensity features for each voxel  $\mathbf{v}$  that are spatially defined by (1) their position,  $\mathbf{p}$ , centered on the voxel to be labeled (Fig. 19.1a), and (2) one or two cuboidal probe regions,  $\mathbf{F}_1$  and  $\mathbf{F}_2$ , offset by displacement vectors,  $\mathbf{\Delta}_1$  and  $\mathbf{\Delta}_2$ , which can be up to 200 mm in each dimension ( $x, y, z$ ). A probe region,  $\mathbf{F}(\mathbf{q}; \mathbf{w})$ , is the set of voxels within the region centered at  $\mathbf{q}$  with side lengths,  $\mathbf{w}$ . We construct two variants of intensity features. The first variant consists of the mean CT intensity at a probed region,  $\mathbf{F}_1$  (Fig. 19.1a, left), while the second consists of the difference in the mean intensity of regions,  $\mathbf{F}_1$  and



**Fig. 19.1** Intensity and entanglement features. (a) Intensity features measure image information from regions offset from the reference voxel at  $\mathbf{p}$ . (b) MAPCLASS feature retrieves the label that the classifier currently predicts at location  $\mathbf{p}_1$  offset from  $\mathbf{p}$ . We maintain a node index array which associates with each voxel the current tree node ID (represented by the number in each voxel). (c, top) The array allows to determine the current label posterior in the tree for the voxel at location  $\mathbf{p}_1$ . (c, bottom) Conceptually, the tree induces a vector image of class posteriors which we use when designing MAPCLASS and TOPNCLASSES features

$\mathbf{F}_2$  (Fig. 19.1a, right). Then split functions are defined from these as follows:

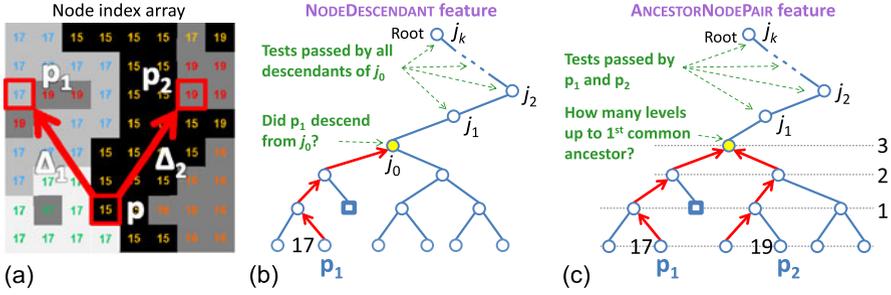
$$h_{\text{INTENSITY}}(\mathbf{v}, \theta_j) = [\bar{J}(\mathbf{F}_1(\mathbf{p} + \Delta_1)) > \tau], \quad (19.1)$$

$$h_{\text{INTENSITYDIFF}}(\mathbf{v}, \theta_j) = [\bar{J}(\mathbf{F}_1(\mathbf{p} + \Delta_1)) - \bar{J}(\mathbf{F}_2(\mathbf{p} + \Delta_2)) > \tau]. \quad (19.2)$$

During training, each type of split function is characterized for node  $j$  by the split parameters  $\theta_j = (\phi, \tau)$ . For  $h_{\text{INTENSITY}}$ ,  $\phi$  includes the parameters of  $\mathbf{F}_1$ : the offset  $\Delta_1$ , the size  $\mathbf{w}_1$  and an intensity threshold  $\tau$ . For  $h_{\text{INTENSITYDIFF}}$ ,  $\phi$  includes the additional parameters  $\Delta_2$  and  $\mathbf{w}_2$ . These parameters are sampled randomly during training for each split node. Once training has finished, the maximum information gain node test along with its optimal features are frozen and stored within the node for later use during testing.

### 19.2.1.2 Semantic Context Entanglement Features

We now describe an instance of our entanglement contribution. During testing on novel images, we exploit the confident voxel label predictions (peaked distributions) that can be found using early levels of the forest to aid the labeling of nearby voxels. This provides semantic context similar to auto-context [341, 375], but does so within a single forest. We define four types of long-range entanglement feature to help train the node currently being grown using knowledge learned in already trained nodes of the forest. Two features (MAPCLASS and TOPNCLASSES) are based on the posterior class distribution of the nodes corresponding to probed voxels, and two (NODEDESCENDANT and ANCESTORNODEPAIR) are based on the location of the nodes within the trees.



**Fig. 19.2** Further entanglement features. (a) Node index array associates voxels with intensity and tree node indices (same format as Fig. 19.1b but for a deeper tree level). (b) NODEDESCENDANT feature tests whether probe voxel at  $\mathbf{p}_1$  descends from a node ( $j_0$  in this case). (c) ANCESTORNODEPAIR feature tests whether the nodes of voxels  $\mathbf{p}_1$  and  $\mathbf{p}_2$  have a common ancestor  $< \tau$  levels away

**MAPCLASS Entanglement Features** As the name suggests, this type of feature uses the maximum a posteriori label of a neighboring voxel at  $\mathbf{p}_1$  to reduce uncertainty about the label at  $\mathbf{p}$  (Fig. 19.1b). When such semantic context is helpful to classify the voxel at  $\mathbf{p}$ , the feature yields high information gain and may become the winning feature for the node during tree growth. The MAPCLASS split function tests whether the MAP class in the posterior of a probed voxel  $\mathbf{p}_1 = \mathbf{p} + \Delta_1$  is equal to a particular class  $c^*$ :

$$h_{\text{MAPCLASS}}(\mathbf{v}, \theta_j) = \left[ \arg \max_c p(c; j(\mathbf{p}_1)) = c^* \right]. \quad (19.3)$$

The parameter  $\theta_j$  includes  $\phi = (\Delta_1, c^*)$  while  $p(c; j(\mathbf{p}_1))$  is the posterior class distribution of the node of  $\mathbf{p}_1$  denoted  $j(\mathbf{p}_1)$ . This posterior can be retrieved from the tree because (1) we train and test voxels in breadth-first fashion, and (2) we maintain an association between voxels and the tree node ID at which they reside while moving down the tree. This association is a node index array (Fig. 19.1b).

**TOPNCLASSES Entanglement Features** Similarly we define features, called TOPNCLASSES, where  $N \in \{2, 3, 4\}$ , that generalize the MAPCLASS feature. A TOPNCLASSES feature tests whether a particular class  $c^*$  is in the top  $N$  classes of the posterior class distribution of the probe voxel at  $\mathbf{p}_1 = \mathbf{p} + \Delta_1$ . The split function, with parameter  $\theta_j$  including  $\phi = (\Delta_1, N, c^*)$  is defined as

$$h_{\text{TOPNCLASSES}}(\mathbf{v}, \theta_j) = \left[ c^* \in \text{top } N \text{ classes of } p(c; j(\mathbf{p}_1)) \right]. \quad (19.4)$$

**NODEDESCENDANT Entanglement Features** This type of feature tests whether a region near voxel  $\mathbf{p}$  has a particular appearance. The neighboring region is centered at voxel  $\mathbf{p}_1$  (Fig. 19.2a, b). The split test is whether the node currently corresponding to  $\mathbf{p}_1$  descends from a particular tree node,  $j_0$ . If it does, then we know  $\mathbf{p}_1$  has satisfied the appearance tests at nodes ( $j_1 \dots j_k$ ) above  $j_0$  in the tree in a particular way to arrive at  $j_0$ .

**ANCESTORNODEPAIR Entanglement Features** This type of feature tests whether two regions near voxel  $\mathbf{p}$  have passed similar appearance and semantic tests. The neighboring regions are centered at voxels  $\mathbf{p}_1$  and  $\mathbf{p}_2$  (Fig. 19.2a). The split test is whether the nodes currently corresponding to  $\mathbf{p}_1$  and  $\mathbf{p}_2$  have their first common ancestor  $< \tau$  tree levels above the current level (Fig. 19.2c). The threshold controls the required degree of similarity: the lower  $\tau$ , the greater the required appearance and context similarity needed to pass the test, because the lower  $\tau$ , the larger the number of tests with identical outcomes above the common ancestor.

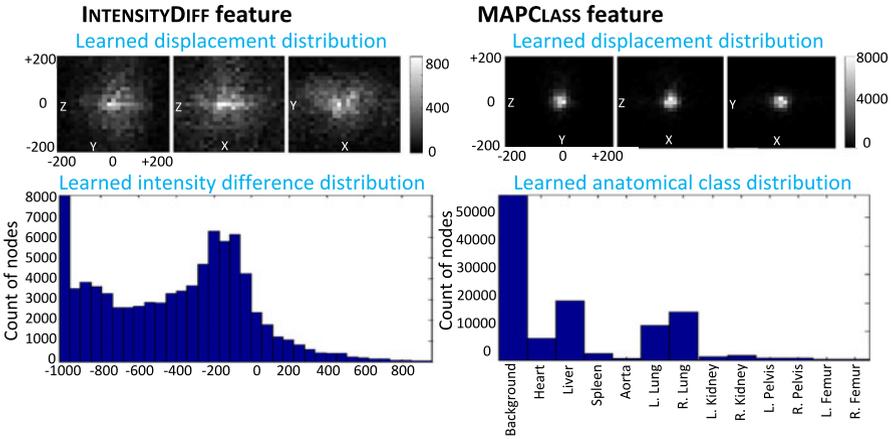
### 19.2.2 Guiding Feature Selection by Learned Proposal Distributions

This section describes the use of learned proposal distributions. These distributions aim to match the feature types and their parameters proposed at each tree node during training to those that have proven to be most useful for classification in a previous training run. The decision forest still chooses the winning feature, but each node chooses from features sets that are likely to be useful based on prior experience. Specifically, we train an initial decision forest,  $F_{\text{temp}}$ , on our training data, using a uniform proposal distribution. We then record (as histograms) the distribution of accepted feature parameters and feature types across all tree nodes in the forest.  $F_{\text{temp}}$  is then discarded, and we then use parameter distributions as the proposal distributions in a subsequent training of the next decision forest. While this requires additional training, it imposes no time penalty for prediction. This process could be repeated, though in practice even just one iteration has proven sufficient for a substantial improvement in accuracy (*e.g.*  $> 5\%$ ).

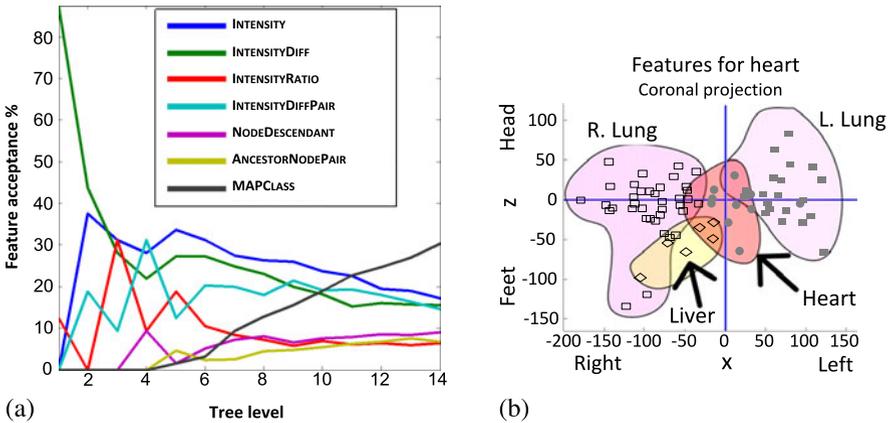
The learned displacements tends to be Gaussian distributed and centered on the reference voxel (Fig. 19.3 top row). Acceptance distributions of the remaining parameters, such as the thresholds  $\tau$  or the choice of the MAPCLASS class  $c^*$ , also have non-uniform distributions (Fig. 19.3 bottom row). Similarly, the distribution of feature types for each tree level is learned. Drawing feature types from this distribution can also improve classifier accuracy. Figure 19.4a shows how the ratio of feature types varies with tree depth. As the tree is grown, entanglement features increasingly dominate the scene over the more conventional intensity features. The entangled features used by the nodes in the lower part of the tree exploit semantic context and neighborhood consistency inferred from appearance features of earlier levels.

### 19.2.3 Results

We evaluate our EDF model on the task of segmenting a database of 250 varying field of view CT scans. Each voxel in each CT scans needs be assigned one of 12 class labels from the following set of anatomical structures of interest {heart, liver,



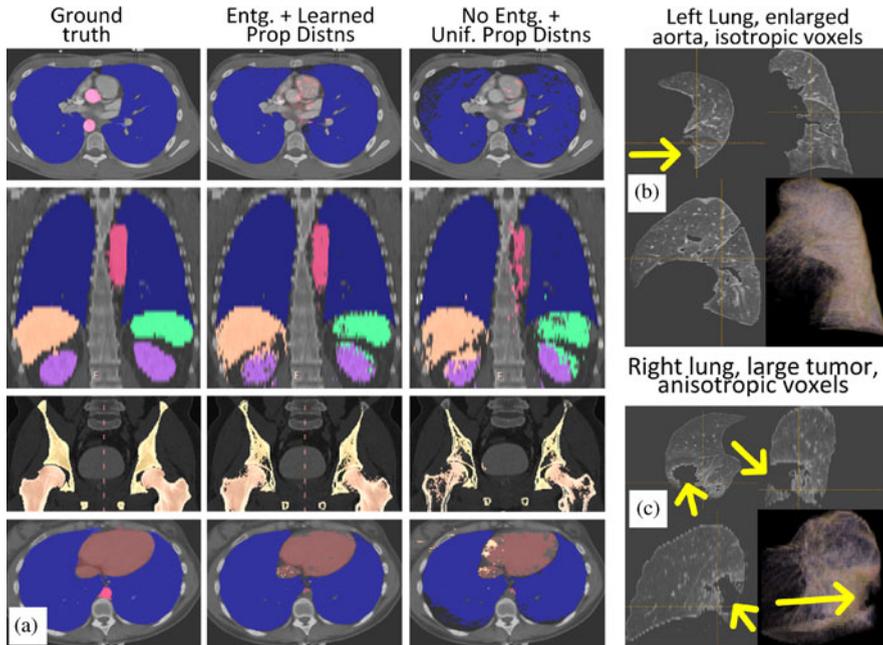
**Fig. 19.3** Learned parameter distributions are clearly non-uniform. (Left) Learned displacement and anatomical class distributions for MAPCLASS feature. (Right) Displacement and intensity difference distributions for INTENSITYDIFF feature



**Fig. 19.4** An EDF reveals how and what it has learned. (a) Learned relative proportion of feature types chosen at each level of forest growth. (b) Location and organ class of the top 50 features used to identify heart voxels. The hand-drawn regions here group these locations for different MAPCLASS classes  $c^*$

spleen, aorta, l./r. lung, l./r. femur, l./r. pelvis, l./r. kidney} or the background class. This database has been designed to include wide variations in patient health status, field of view and scan protocol. We randomly selected 200 volumes for training and 50 for testing.

**Qualitative Results** The EDF achieves a visually accurate segmentation of organs throughout the 50 test volumes. Example segmentations are shown in Fig. 19.5a where the first column is the ground truth segmentation, and the sec-



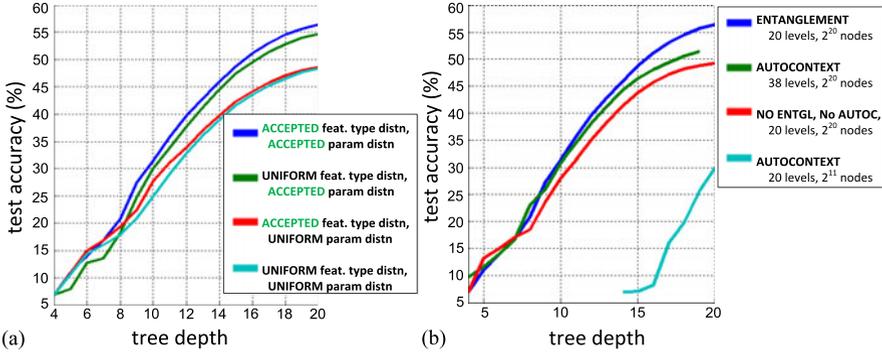
**Fig. 19.5** Qualitative segmentation results. (a) The use of entanglement and learned proposal distributions (*column 2*) improves accuracy compared to not using them (*column 3*). The *rows* show four different subjects. (b) EDF segmented left lung distorted by enlarged aorta; volume rendering in *lower right*. (c) EDF accurately segments a right lung despite a severe tumor

ond column is the EDF result. We see good agreement for the lungs (blue), liver (orange), spleen (green), kidneys (purple), femur (tan), and heart (dark brown). Column 3 shows the result using our decision forest without entanglement and with uniform proposal distributions. Entanglement with proposal distributions noticeably improves the lungs, aorta (red), kidneys, spleen, femur, and heart.

The algorithm handles many complexities commonly found in the clinic. For example, our algorithm correctly segmented the lung despite the case of a severely enlarged aorta (Fig. 19.5b) and another with a tumor (Fig. 19.5c).

**Quantitative Impact of Each Contribution** For a quantitative analysis we measured segmentation accuracy across all 50 test scans using the average class Jaccard similarity coefficient [100]. The metric is the ratio of the intersection size (ground truth and predicted labels) divided by the size of their union. While EDF achieves  $> 97\%$  average voxel accuracy throughout our database, we use the Jaccard metric because we feel it is a more reliable metric of segmentation accuracy.

To understand the impact of using the acceptance distributions as proposal distributions (Sect. 19.2.2), we trained the decision forest in four different ways: (1) using uniform feature type and uniform feature parameter distributions for baseline performance (light blue curve, Fig. 19.6a), (2) using learned feature type distribution



**Fig. 19.6** Quantitative impact of each contribution. (a) Learning proposal distributions for both feature types and feature parameters increases accuracy. (b) Entanglement (*dark blue*) provides greater accuracy and prediction speed than auto-context (*green*). Note: the *green* curve should properly be plotted at depths 20–38, but for ease of comparison we plot it at depths 1–19

with uniform feature parameter distributions (red), (3) using uniform feature type distributions with learned feature parameter distributions (green), (4) using learned feature type and learned parameters distributions (dark blue). Learning only the feature type distribution yields a negligible improvement to baseline (red vs. light blue). Learning feature parameter distribution boosts accuracy significantly (green vs. red). Learning both yields the best performance boosting accuracy over baseline by 8 %.

We compared our method to auto-context [341, 375] by conducting four experiments. First, we trained our decision forest 20 levels deep without entanglement and without auto-context for a baseline (red, Fig. 19.6b). Second, we trained a two-round, auto-context decision forest (ADF) using 10 levels in each round (light blue). Third, we trained another ADF, but this time with an equal *modeling capacity* to the baseline by using two rounds with 19 levels each (green). Fourth, we trained the proposed EDF method as a single, 20 level deep forest using entanglement (dark blue curve). We find considerably better accuracy using the EDF method (dark blue vs. green). In addition to beating the performance of ADF, it reduces the prediction time by 47 % since the EDF requires 18 fewer levels (20 vs. 38).

**Efficiency Considerations** With a parallel implementation, EDF segments volumes ( $512 \times 512 \times 424$ ) in just 12 seconds per volume using an 8 core Xeon 2.4 GHz computer with 16 GB RAM. This speed is equal to or better than state of the art single organ methods [419], yet we segment *multiple* (12) organs simultaneously.

**Inspecting the Chosen Features** Figure 19.4b shows how the MAPCLASS feature learns to segment a heart voxel located at the cross-hair. To find the top contributing semantic context features, we express information gain as a sum of the information gain from each class:

$$I(\mathcal{S}, \theta) = \sum_{c \in \mathcal{C}} \left( -p(c|\mathcal{S}) \log p(c|\mathcal{S}) + \sum_{i \in \{L, R\}} \left( \frac{|S^i|}{|\mathcal{S}|} p(c|S^i) \log p(c|S^i) \right) \right), \quad (19.5)$$

where  $\mathcal{S}$  is the set of voxels being split into partitions  $\mathcal{S}^L$  and  $\mathcal{S}^R$ , and  $c$  is the index over classes. We can then readily rank the learned node features based on how much they contributed to classifying the voxels of a given class (*e.g.* heart) by increasing the information gain for that class.

### 19.3 Differentiable Information Gain Maximization

In Sect. 19.2 we simultaneously increased classification accuracy and reduced decision time using entanglement which propagates knowledge from one part of the forest to another. In this section we achieve a similar result using a complementary approach. This second approach optimizes training by applying gradient ascent to differentiable information gain. Finding the optimum parameters for the split tests has traditionally [302] been achieved via exhaustive discrete search to find those parameters which maximize information gain. For a given computational budget, exhaustive search is limited to a small region of parameter space or a coarse quantization of a wider region.

By making information gain differentiable, we can directly find the optimal data partition for the given input subset and node feature subset. This produces more compact decision trees which in turn reduces classification (test) time and memory requirements. Through the use of random input (bagging) and feature subset selection, decision forest virtues including independent trees and resiliency to overfitting can be retained.

Using non-differentiable information gain, an optimal solution can be found by simulated annealing techniques [158, 260], though this can be computationally impractical in high dimensional feature spaces. Alternative discriminative criteria could be optimized, such as LDA [246], SVM [383], or boosting techniques [374, 413], but these may not provide the optimal data partitions when the data distributions are from many classes.

As discussed in Sect. 3.3, binary tests based on parameterized functionals (such as hyperplanes or conic sections) are stronger learner models than coordinate aligned split functions, and can have greater generalization capabilities. We show below that our differentiable information gain forest both accepts these more powerful split functions and improves their generalization power to approximate the maximum margin decision boundary (see also [80] for a detailed discussion about maximum margin behavior in decision forests). Like [173], we impose a soft split using a sigmoid function; however, we explicitly present the derivation absent in [173], and extend it to include label confidence when training data include label uncertainty. We also incorporate hyperplane and non-linear split tests in the gradient ascent framework.

We call the resulting classifier a gradient ascent decision forest (GADF) and illustrate its power in both synthetic and real-world examples. First, we investigate the impact of the GADF for a synthetic 2D classification task, and demonstrate that gradient ascent reduces classification time, increases prediction accuracy and increases

confidence in the estimation of the maximum margin decision boundary, compared to the standard decision forest. Second, we implement a GADF to solve classification problems in several application domains including mass spectrometry, biomechanics, botany, image classification and 1D signal processing. We demonstrate how the GADF approach increases prediction accuracy across this application spectrum. Third, we cast visual object tracking as an iterative classification task and train a gradient ascent classifier to perform object tracking in public PET videos. We show how the approach avoids tracker drift and handles severe occlusions better than state of the art trackers.

### 19.3.1 Formulating Differentiable Information Gain

Given the labeled dataset  $\mathcal{S} = \{(\mathbf{v}, c)\}$  of size  $N$ , where  $\mathbf{v}$  is the feature data of dimension  $d$ , and  $c = \zeta(\mathbf{v})$  is the ground truth class label of  $\mathbf{v}$  (with  $c \in \{1, \dots, C\}$ ), the Shannon entropy of the class distribution can be computed as

$$H(\mathcal{S}) = - \sum_{c=1}^C p(c|\mathcal{S}) \log p(c|\mathcal{S}), \quad (19.6)$$

where

$$p(c|\mathcal{S}) = \frac{\sum_{\mathbf{v}} [\zeta(\mathbf{v}) - c]}{N} \quad (19.7)$$

defines the data class distribution and  $[\cdot]$  is the indicator function.

Given a binary split function (weak learner)  $h(\cdot, \cdot)$ , we can partition the data into two subsets (see Sect. 3.2.3):

$$\mathcal{S}^L = \{\mathcal{S}|h = 1\} = \{(\mathbf{v}, c) | \mathbf{v} \in \mathcal{S}, h(\mathbf{v}; \boldsymbol{\theta}) = 1\} \quad (19.8)$$

of size  $N^L$  and

$$\mathcal{S}^R = \{\mathcal{S}|h = 0\} = \{(\mathbf{v}, c) | \mathbf{v} \in \mathcal{S}, \bar{h}(\mathbf{v}; \boldsymbol{\theta}) = 1 - h(\mathbf{v}; \boldsymbol{\theta}) = 1\} \quad (19.9)$$

of size  $N^R = N - N^L$ . The information gain defines the entropy change before and after the split  $h$  is applied:

$$I(\mathcal{S}|h) = H(\mathcal{S}) - H(\mathcal{S}|h), \quad (19.10)$$

where the entropy after partitioning is computed as

$$H(\mathcal{S}|h) = \frac{N^L}{N} H(\mathcal{S}^L) + \frac{N^R}{N} H(\mathcal{S}^R). \quad (19.11)$$

Information gain is a typical measure for selecting discriminative weak learners in decision tree training as described in Chap. 3. However, the information gain formulation is not differentiable w.r.t. the parameters  $\boldsymbol{\theta}$  of  $h$ , making analytical optimization problematic.

To make  $I(\mathcal{S}|h)$  in (19.10) differentiable w.r.t. the binary test  $h$ , we first define the split function  $h$  for data point  $\mathbf{v}$  as a parameterized functional:

$$h_{\psi}(\mathbf{v}; \boldsymbol{\theta}) = \begin{cases} 0, & \psi(\mathbf{v}; \boldsymbol{\theta}) < 0, \\ 1, & \psi(\mathbf{v}; \boldsymbol{\theta}) \geq 0, \end{cases} \quad (19.12)$$

where  $\psi(\mathbf{v}; \boldsymbol{\theta})$  is the geometric split function of feature space with parameter set  $\boldsymbol{\theta}$ . The partition occurs at the boundary  $\psi(\mathbf{v}; \boldsymbol{\theta}) = 0$ .

We then define partition integrals for each class for all data w.r.t.  $h$  as follows:

$$U_c^{\mathcal{S}}(h) = \sum_{\mathbf{v}} h(\mathbf{v}; \boldsymbol{\theta}) [\zeta(\mathbf{v}) - c], \quad c \in \{1, \dots, C\}, \quad (19.13)$$

$$U^{\mathcal{S}}(h) = \sum_{\mathbf{v}} h(\mathbf{v}; \boldsymbol{\theta}), \quad (19.14)$$

where  $h(\mathbf{v}; \boldsymbol{\theta})$  can be replaced with  $\bar{h}(\mathbf{v}; \boldsymbol{\theta}) = 1 - h(\mathbf{v}; \boldsymbol{\theta})$  as needed.

We can then define  $N^{\text{L}} = U^{\mathcal{S}}(h)$ ,  $N^{\text{R}} = U^{\mathcal{S}}(\bar{h})$ ,  $N = N^{\text{L}} + N^{\text{R}}$ ,  $p_c(\mathcal{S}^h) = \frac{U_c^{\mathcal{S}}(h)}{U^{\mathcal{S}}(h)}$  and  $p_c(\mathcal{S}^{\bar{h}}) = \frac{U_c^{\mathcal{S}}(\bar{h})}{U^{\mathcal{S}}(\bar{h})}$ , and the entropy after partition by  $h$  is

$$\begin{aligned} H(\mathcal{S}|h) = & -\frac{1}{N} \left( \sum_c U_c^{\mathcal{S}}(h) \log U_c^{\mathcal{S}}(h) - U^{\mathcal{S}}(h) \log U^{\mathcal{S}}(h) \right. \\ & \left. + \sum_c U_c^{\mathcal{S}}(\bar{h}) \log U_c^{\mathcal{S}}(\bar{h}) - U^{\mathcal{S}}(\bar{h}) \log U^{\mathcal{S}}(\bar{h}) \right). \end{aligned} \quad (19.15)$$

Using the chain rule, the derivative of information gain w.r.t.  $\boldsymbol{\theta}$  is

$$\begin{aligned} \frac{\partial I}{\partial \boldsymbol{\theta}} = & -\frac{\partial H(\mathcal{S}|h)}{\partial \boldsymbol{\theta}} \\ = & \frac{1}{N} \left( \sum_c U_c^{\mathcal{S}}(h) (\log U_c^{\mathcal{S}}(h) + 1) - U^{\mathcal{S}}(h) (\log U^{\mathcal{S}}(h) + 1) \right. \\ & \left. + \sum_c U_c^{\mathcal{S}}(\bar{h}) (\log U_c^{\mathcal{S}}(\bar{h}) + 1) - U^{\mathcal{S}}(\bar{h}) (\log U^{\mathcal{S}}(\bar{h}) + 1) \right), \end{aligned} \quad (19.16)$$

where  $U_c^{\mathcal{S}}(h) = \sum_{\mathbf{v}} \frac{\partial h(\mathbf{v}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} [\zeta(\mathbf{v}) - c]$ ,  $c \in \{1, \dots, C\}$ , and  $U^{\mathcal{S}}(h) = \sum_{\mathbf{v}} \frac{\partial h(\mathbf{v}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ .

Information gain is not differentiable w.r.t. the binary test parameter  $\boldsymbol{\theta}$  because  $h_{\psi}(\mathbf{v}; \boldsymbol{\theta})$  in (19.12) is not differentiable. To make it differentiable, we approximate the weak learner  $h$  by a sigmoid function  $h_{\psi}(\mathbf{v}; \boldsymbol{\theta}) = 1/(1 + e^{-\frac{\psi(\mathbf{v}; \boldsymbol{\theta})}{\sigma}})$  to get:

$$\frac{\partial h_{\psi}(\mathbf{v}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{\sigma} h_{\psi}(\mathbf{v}; \boldsymbol{\theta}) (1 - h_{\psi}(\mathbf{v}; \boldsymbol{\theta})) \frac{\partial \psi(\mathbf{v}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \quad (19.17)$$

Combining (19.16) and (19.17) allows us to compute the derivative of information gain w.r.t. the binary test function parameter  $\boldsymbol{\theta}$  using the chain rule. The split

function  $\psi(\mathbf{v}; \theta)$  can be designed according to the purpose of information gain optimization. It is worth noting that the parameter  $\sigma$  defines the fidelity of the binary test, and controls the smoothness of the information gain surface in the decision boundary parametric space. One may apply annealing to  $\sigma$  when doing gradient ascent (*i.e.* letting  $\sigma \rightarrow 0$ ) so that the chance that the optimization reaches global maxima can be increased.

**Soft Label Decision Forests** If each training data point  $\mathbf{v}$  also includes a (training) class label probability measure  $q_c(\mathbf{v})$ , then we define a confidence score  $\gamma(\mathbf{v}) \in [0, 1]$  as a function of the label log-likelihood  $l(\mathbf{v}) = \log(\frac{q_c(\mathbf{v})}{1-q_c(\mathbf{v})})$  as follows:

$$\gamma(\mathbf{v}) = \frac{2}{1 + e^{-\frac{\sqrt{|l(\mathbf{v})|} - t_l}{\sigma_l}}}. \quad (19.18)$$

The intuition is that the label is less confident if the class probability ratio is too close to 1 (and thus  $\sqrt{|l(\mathbf{v})|}$  approaches zero), and  $\sigma_l$  controls how sensitive the confidence score is to the log-likelihood-ratio score. Such class label probability measures occur naturally in tasks such as video processing. In this case an on-line model learning may be applied per frame. Given the classification model trained on-line using the previous frames, the new observations in the current frame may be labeled with likelihood confidence (soft labels), and become the training data for the on-line model updating for the current frame. In [52], such ‘soft label decision forests’ are used for on-line tracking in videos where labels are quantized into a histogram and a standard node training procedure is applied.

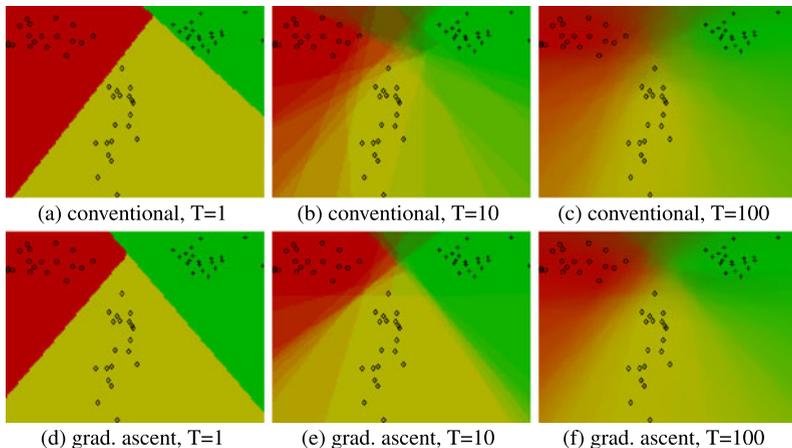
We provide an analytic solution for the soft label decision forests learning problem. By modeling the label confidence measures based on how much the label likelihood deviates from the decision threshold, we derive a differentiable information gain formulation weighted by the label confidence. Our gradient ascent optimization technique can then be applied to find the optimal data split based on the information gain criteria with respect to the known class labels. Specifically, to optimize the information gain with emphasis on data  $\mathbf{v}$  labeled with high confidence, we can simply derive the differentiable information gain by weighting the terms in (19.13) with the confidence measure:

$$U_c^{\mathcal{S}}(h) = \sum_{\mathbf{v}} h(\mathbf{v}; \theta) [\zeta(\mathbf{v}) - c] \gamma(\mathbf{v}), \quad c \in \{1, \dots, C\}, \quad (19.19)$$

$$U^{\mathcal{S}}(h) = \sum_{\mathbf{v} \in \mathcal{S}} h(\mathbf{v}; \theta) \gamma(\mathbf{v}). \quad (19.20)$$

### 19.3.2 Split Function Design and Gradient Ascent Optimization

In training a classification forest, we solve for a decision boundary,  $\psi_{\text{DF}}(\mathcal{S}; \theta_{\text{DF}})$ , optimally partitioning the data  $\mathcal{S}$  with maximal information gain:



**Fig. 19.7** Gradient ascent can improve the location and confidence in the maximum margin decision boundary and reduce classification time. *Top row: discrete optimization* using: (a) 1 tree, (b) 10 trees, (c) 100 trees. *Bottom row: discrete optimization followed by gradient ascent optimization* using: (d) 1 tree, (e) 10 trees, (f) 100 trees

$$\theta_{\text{DF}}^* = \arg \max_{\theta_{\text{DF}}} I(\mathcal{S} | h(\mathbf{v}; \theta_{\text{DF}})), \quad (19.21)$$

where  $\theta_{\text{DF}}$  is the concatenated vector of the binary test parameters  $\theta$  at each tree node.

The classic decision forest [44] uses a univariate split test, which consists of a threshold  $\tau$  of the  $k$ th feature element of  $\mathbf{v}$ . We can denote this partitioning boundary function as  $\psi = g_0(\mathbf{v}; \theta) = v_k - \tau$  with  $\theta = \theta_0 = \{k, \tau\}$ . Such a boundary is well suited for fast discrete search via maximizing  $I$ . However, the boundary coordinate alignment in feature space can require many binary tests to make the joint decision boundary  $\psi_{\text{DF}}(\mathcal{S}; \theta_{\text{DF}})$  approximate the maximum margin class separation boundary (as discussed in Chap. 3). An alternative is to approximate the maximal margin decision boundary using far fewer but stronger weak learner models, such as hyperplanes or conic functions, and this is especially true if differentiable information gain is used. This is vital since maximum margin fidelity is likely to endow the forest with superior generalization properties.

To illustrate, Fig. 19.7 shows synthetic 2D data points from three classes (red, yellow, green) and the resulting partitions of a 2D feature space  $(x_1, x_2)$  by different decision forests architectures. The top row shows the results of applying a conventional classification forest. The bottom row shows the results (for corresponding forest size) of a classification forest using our differentiable information gain. While it has been shown in [80] (and in Chap. 4) that when a large number of trees are used the maximum margin class boundaries can be found, in practice, a small number of trees is typically preferred for either classification runtime efficiency or memory constraints. A comparison between the two rows in Fig. 19.7 shows that our new formulation of information gain allows forests to approximate maximum margin

behavior accurately even with just a few trees. In fact, in each column, showing the results on various number of trees, we see an improvement in the maximum margin boundary.

In detail, Fig. 19.7a shows the result of a decision forest with 1 tree and oriented-line weak learners. We observe that the decision boundary does not approximate well the maximum margin decision boundary. Averaging the output of 10 trees, Fig. 19.7b, starts to improve the location of the class boundary. Using 100 trees Fig. 19.7c provides a reasonable approximation to the maximum margin location and a smooth transition class posterior.

Using gradient ascent optimization yields improved location of the class boundary even for just one tree (Fig. 19.7d). Here, the method is initialized with the result from Fig. 19.7a. Figure 19.7e shows the result when the output from 10 gradient ascent trained trees are averaged. Compared to Fig. 19.7b we can see the confidence in the correct maximum margin boundary location is improved and a smoother posterior. Similarly, when the output from 100 gradient ascent trained trees are averaged in Fig. 19.7f, an improvement in the confidence of the correct maximum margin decision boundary is still observed.

The improvement in maximum margin fidelity obtained by using GADF can provide additional generalization when training data are limited, which is often the case in practice. The use of fewer trees also substantially speeds up classification time, since each gradient ascent trained tree does not require additional time to test yet provides increased accuracy. For example, the gradient ascent based result in Fig. 19.7e has similar maximum margin fidelity to the non-gradient ascent result in Fig. 19.7c yet requires 10 times fewer trees. In additional 2D synthetic tests we have also observed a large capture range (basin of convergence) for both oriented hyperplane and conic weak learners when using GADF.

With this motivation for differentiable information gain, we derive the gradient of two useful binary tests for gradient ascent as follows:

**Hyperplane Partition** Binary test functionals in the form of hyperplanes for the training of decision forest can be defined as  $\psi(\mathbf{v}; \boldsymbol{\theta}) = g_1(\mathbf{v}; \boldsymbol{\theta}) = \boldsymbol{\theta}^\top \begin{bmatrix} \mathbf{v} \\ 1 \end{bmatrix}$  where  $\boldsymbol{\theta} = \theta_1 = [\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(d+1)}] \in \mathbb{R}^{d+1}$  may be directly incorporated into the proposed gradient ascent information gain optimization where we can show that

$$\frac{\partial \psi}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \mathbf{v} \\ 1 \end{bmatrix}. \quad (19.22)$$

**Hyper-Ellipsoid Partition** Binary tests with hyper-ellipsoid split functionals in  $\mathbb{R}^d$  can be defined as  $\psi(\mathbf{v}; \boldsymbol{\theta}) = g_2(\mathbf{v}; \boldsymbol{\theta}) = 1 - (\mathbf{v} - \mathbf{v}_0)^\top \mathbf{Q}(\mathbf{v} - \mathbf{v}_0)$  where  $\mathbf{v}_0$  is the ellipsoid center, and  $\mathbf{Q}$  defines the semi-axes lengths and orientations. To incorporate into our gradient ascent information gain optimization, we have

$$\frac{\partial \psi}{\partial \mathbf{v}_0} = -2(\mathbf{v} - \mathbf{v}_0)^\top \mathbf{Q} \quad \text{and} \quad \frac{\partial \psi}{\partial \mathbf{Q}} = (\mathbf{v} - \mathbf{v}_0)(\mathbf{v} - \mathbf{v}_0)^\top \quad (19.23)$$

with the optimization subject to  $\mathbf{Q} > 0$ .

To train each node using gradient ascent information gain optimization, an initial estimate of the split function parameters can be chosen by random guess or discrete optimization in discrete parameter space. For example, for hyperplanes we find the best simple binary test with parameter  $\theta_0^* = \{k^*, \tau^*\}$  by discrete search and then set the initial guess  $\theta_1 = [0, \dots, 0, \theta_1^{(k^*)} = 1, 0, \dots, -\tau^*]^\top$ . Given the formulation and initial guess, we can conveniently implement the gradient ascent optimization by adopting existing off-the-shelf gradient ascent optimization toolboxes (*i.e.* `fminunc()` in Matlab with default options).

### 19.3.3 Object Tracking via Information Gain Maximization

Reliable visual tracking of a target object is difficult as many confusing factors need to be addressed including: occlusions, distractions from background clutter, and object appearance variations. We cast object tracking as an iterative classification problem [9, 11, 139, 178], and model the on-line appearance model update and tracking as a sequential process of information gain maximization that partitions the pixels in feature space (image features) and in image space (incorporating pixel coordinates as additional features) iteratively. Various decision forest based visual trackers have been proposed in the literature. A popular approach is to use classification forests to construct an appearance likelihood model that is updated on-line in the current frame and evaluated for the next frame. Tracking is achieved by finding the maxima of the confidence map for the next frame by picking the centroid. As shown in [317], the on-line decision forest model based visual tracker consistently outperformed that based on an on-line Adaboost model. In [120], the concept of Hough forests is proposed. With Hough forests, the target center is detected and tracked by the fusion of generalized Hough transforms that are based on the codebook classification of local image patches. Also, Chap. 12 and Chap. 16 provide further forest-based video tracking algorithms.

Our approach is substantially different from the previous approaches. We consider the tracking as an *information gain maximization process* (Gain-Max tracking) in pixel XY-coordinate space. By parameterizing the target shape with an ellipsoid, the tracking of the target location and scale can be achieved by maximizing the differentiable information gain via gradient ascent techniques. We note that while we train a forest for each frame, the amount of data is small and thus a forest can be trained quickly. Realtime computation can be achieved by sacrificing some model optimality (*i.e.* limiting the number of trees), or by adopting on-line decision forest learning techniques [317].

Given an image  $J$ , we define a region of interest as  $\Omega(J)$  and denote the target region as  $\Omega^+(J)$  and background region  $\Omega^-(J)$  such that  $\Omega(J) = \Omega^+(J) \cup \Omega^-(J)$ . The pixel feature vector at location  $\mathbf{p}$  is a  $d$  dimensional vector denoted  $J(\mathbf{p}) \in \mathbb{R}^d$  where  $J$  has channels for textures, RGB colors, gradients, and wavelets. We denote target and background pixel labels as  $\{F, B\}$ . We use information gain maximization in two ways: first, to learn to discriminate foreground from background based

on pixel appearance (feature space) and second, to track the target in the pixel XY-coordinate space (image space). These are explained next.

**GainMax in Feature Space: Updating the Appearance Model** To discriminate foreground from background pixels, we train a two-category pixel classifier that assigns pixels with a label from  $\{F, B\}$ . When information gain maximization is achieved, solving (19.21) the classifier learns the image features that best separate the training data:  $\mathcal{S}^J = \{(J(\mathbf{p}), \zeta(\mathbf{p})) \mid \mathbf{p} \in \Omega(J), J(\mathbf{p}) \in \mathbb{R}^d, \zeta(\mathbf{p}) \in \{F, B\}\}$  into foreground and background. The features are computed directly from the image while the target and background labels come from a prior frame (the initial frame is assumed manually labeled). For example, we can obtain the label of the pixel at location  $\mathbf{p}$  as

$$\zeta(\mathbf{p}) = \begin{cases} F & \text{if } \mathbf{p} \in \Omega^+(J), \\ B & \text{if } \mathbf{p} \in \Omega^-(J). \end{cases} \quad (19.24)$$

**GainMax in Image Space: Tracking the Target** Further optimization of the foreground and background is possible if we take into consideration each pixel's XY-coordinates in image space in addition to the pixel foreground and background labels output from the previous step's two-category classifier. We denote such a training dataset as  $\mathcal{S}^{\Omega(J)} = \{(\mathbf{p}, \zeta_{DF}(J(\mathbf{p}))) \mid \mathbf{p} \in \Omega(J)\}$  where  $\{\zeta_{DF}(J(\mathbf{p})) \in \{F, B\} \mid \mathbf{p} \in \Omega(J)\}$ . To solve, we find the optimal partition boundary  $h_\psi(\mathbf{p}; \theta^*)$  that achieves maximal information gain. Intuitively, the optimal split function  $\psi(\mathbf{p}; \theta^*) = 0$  should match the target region boundary. The solution can again be found by solving (19.21) using gradient ascent, but with the target boundary function being parameterized based on the predefined target shape model, *i.e.* a 2D ellipsoid. As  $\{\zeta_{DF}(J(\mathbf{p}))\}$  is estimated by the on-line trained classification forest in feature space with probability  $q_\zeta(J(\mathbf{p}))$ , we can perform tracking by optimizing the confidence weighted information gain formulation derived in (19.15) and (19.19).

## 19.3.4 Results

### 19.3.4.1 Classification of Public Machine Learning Datasets

To evaluate the gradient ascent decision forest (GADF), we compare its performance to those of commonly used classifiers including a reference standard Adaboost implementation and a decision forest with oblique hyperplanes. We denote these classifiers as follows:

**Adaboost** A standard Adaboost classifier that uses axis-aligned stumps as decision functions. This is used as a baseline reference for comparison.

**StumpDF** A standard decision forest classifier with an oblique hyperplane for the binary test. The optimal binary test is searched by randomly drawing 20 hyperplane samples in the feature space.

**Table 19.1** Comparison of classification equal-error rate for **Adaboost**, **StumpDF** and **GADF** on public datasets

Dataset Name	#sample	# fea.	#Train:#Test	Adaboost	StumpDF	GADF
<i>Arcene</i>	200	10000	1:1	0.25	0.318	<b>0.240</b>
<i>Vertebral Column</i>	310	6	3:7	<b>0.157</b>	0.175	0.170
<i>Iris</i>	150	2	1:4	0.281	0.010	<b>0.000</b>
<i>Cardiotocography</i>	2126	23	3:7	0.021	0.022	<b>0.019</b>
<i>Breast Cancer Wisconsin</i>	569	32	1:1	0.056	0.048	<b>0.043</b>

**GADF** Similar to the **StumpDF**, but gradient ascent information gain optimization is also used during training, using the hyperplane with best performance of the randomly drawn 20 planes. Assuming the data are always normalized into standard deviation along each dimension, we do gradient ascent information gain optimization by gradually reducing the annealing parameter  $\sigma$  starting from  $0.03 \frac{1}{\sqrt{d}}$  (where  $d$  is the feature dimension of the data) with multiplicative scaling factor 0.7. The annealing stops when the optimization converges.

We train both **StumpDF** and **GADF** with 10 trees and we train the trees by randomly sampling 90 % of the original training set. When training each tree node, we search for the optimal split parameters in a randomly sampled feature subspace with dimension number  $\text{ceil}(\sqrt{d})$ . We limit the maximal tree depth to 15. To evaluate the three methods, we compare their performance over a variety of different application domains using publicly available standard datasets used throughout the machine learning community [111, 377]. We select five datasets, including: *Arcene* (where the application is mass spectrometry), *Vertebral Column* (biomechanics), *Iris* (botany), *Cardiotocography* (1D signal processing), and *Breast Cancer Wisconsin* (cell image classification). We used the given train and test datasets when they are explicitly provided and divide the dataset into different ratios to evaluate generalization, as shown in Table 19.1, when they are not given. We also reduced *Iris* to only the first two features to make the test more challenging. The table summarizes the classification equal-error rates for the three methods, averaged over five experimental runs for each method. We observe that in nearly every case, **GADF** outperforms **StumpDF** as well as the reference standard **Adaboost**.

#### 19.3.4.2 Object Tracking in Videos

In Sect. 19.3.3 we embedded GADF into a full-fledged tracking application. Here we compare its performance to the mean-shift tracker [69]. For the video data we used a standard object tracking PET video as the evaluation task [290]. We evaluated three variants of our GADF based tracker. For all our variants the tracker begins by learning a two-category foreground/background pixel classifier (an appearance model) using the manually delineated first frame. Characteristics of the trackers we compare are as follows:



**Fig. 19.8** GainMax trackers using gradient ascent information gain maximization can handle distraction and occlusions well. For the mean-shift tracker (column 1), the red box is the tracking result; for GainMax trackers 0–2 (columns 2–4), red indicates pixels correctly labeled in the ellipsoid as target (true positive), blue indicate false positive pixels outside of the target ellipsoid

**MeanShift** The standard mean-shift tracker with histogram of size 9 by 9 by 9 bins in RGB space.

**GainMax0** The appearance model from the first frame is reused on all frames. Gradient ascent is used to refine the tracking boundary in image space.

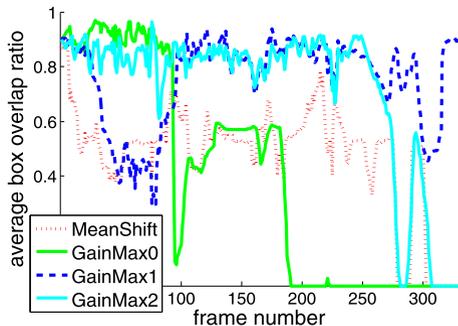
**GainMax1** Training data are updated to use background pixels from the previous frame and target pixels from the first frame. Then the two-category pixel classifier is retrained and the tracking boundary is refined using gradient ascent for tracking in both image and pixel feature space.

**GainMax2** Training data are updated to use the previous frame’s target and background pixels. Then the two-category pixel classifier is retrained and the tracking boundary is refined using gradient ascent for tracking in both image and pixel feature space.

For the three different variants of GainMax tracking, we fix the number of trees to 20, and the depth of the trees to be 10. We train the decision forest by randomly sampling 80 % of the pixels as training data for each tree, and randomly choose four features from  $[r, g, b, \Delta X_x, \Delta X_y, \|\Delta X\|, \angle \Delta X]$  for the training of the binary test functionals.

Some qualitative results are shown in Fig. 19.8, in which a woman turns her head around in an office environment, and then a second person enters and occludes the

**Fig. 19.9** The average tracking/ground truth box overlap ratio for the lady video



woman. Tracking is also challenging because the wall has nearly human skin color. We observe that **Meanshift** and **GainMax0** cannot handle the substantial variations in target appearance as they do not do model updating. They are eventually distracted and fail to track. **GainMax1** and **GainMax2** both maintain correct tracking of the lady even over the wall because their appearance models are updated on-line in the pixel feature space and learn to distinguish the face color from wall color. By comparison, **GainMax2** works better when there is no occlusions. However, it is easily distracted by the occlusion. However, **GainMax1** can resist this distraction successfully because it does not update the target pixel data, and can maintain tracking to the end of the video.

Given the ground truth target bounding box  $B_g$  and the tracking bounding box  $B_t$  in a frame, we can evaluate the tracking performance by their average box overlapping ratio w.r.t. the boxes:  $R(B_g, B_t) = (\frac{\Omega(B_g \cap B_t)}{\Omega(B_g)} + \frac{\Omega(B_g \cap B_t)}{\Omega(B_t)})/2$ . In Fig. 19.9 we plot the average overlap ratio of the four trackers on the *lady* video.

We can further summarize the tracker’s accuracy by the percentage of frames in which  $R(B_t, B_g) > 0.5$ . Figure 19.10 summarizes the performance of the four evaluated trackers on three videos commonly used for evaluation in visual tracking. Due to substantial appearance variations, occlusions and background distractions, **Mean-shift** and **GainMax0** get distracted easily as they do not perform model updating. **GainMax2** can achieve high tracking accuracy when there is gradual appearance variations (as shown in Fig. 19.9), but fails to track when occlusions exist. Over all,

	 <i>lady</i> [135]	 <i>view5</i> [291]	 <i>mall</i> [57]
% ( $R > 0.5$ )			
Meanshift	0.66	0.46	0.55
GainMax0	0.43	0.47	0.42
GainMax1	0.85	0.92	1.00
GainMax2	0.82	0.47	1.00

**Fig. 19.10** Robustness comparison of the visual trackers by the percentage of frames with  $R(B_t, B_g) > 0.5$

**GainMax1** achieves the best robustness as it does model updating while avoiding drifting.

## 19.4 Discussion and Conclusion

This chapter has presented three complementary improvements to the decision forest framework presented in this book.

*Entanglement* propagates knowledge from one part of the forest to another which speeds learning, improves classifier generalization, and exploits long-range spatial correlations in the input data.

*Differentiable information gain maximization* allows the optimal data partitioning functional to be found directly through gradient ascent rather than through an exhaustive search over discrete functional parameter space.

Entanglement and differentiable information gain maximization enhance different aspects of decision forests: the use of semantic contextual features and the node optimization function, respectively; they are mutually compatible and may be combined to further enhance the forest accuracy.

The *learned proposal distributions* (Sect. 19.2.2) and differentiable information gain maximization both tackle the problem of node optimization in the presence of a high dimensional feature space. The former increases the effectiveness of brute-force feature search. The latter optimizes the information gain more directly. Since differentiable information gain requires initialization, these two methods can be combined effectively.

The fundamental enhancements presented here may be directly applied to improve results in other applications that use classification forests, including multiple sclerosis lesion segmentation [125], brain segmentation [411], myocardium delineation [212], and more generic object class segmentation tasks [342]. Here we have applied entanglement only to the task of anatomy segmentation, but it is a generic concept and may be adapted to exploit other correlations (*e.g.* over time or space). Likewise our differential information gain approach can form the basis for gradient ascent optimization with more complicated data partitioning functionals (*e.g.* differentiable shape models) based on a-priori heuristics for specific applications.

# References

1. Agarwal S, Awan A, Roth D (2004) Learning to detect objects in images via a sparse, part-based representation. *IEEE Trans Pattern Anal Mach Intell* 26(11)
2. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Ageing-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement* 7(3)
3. Amit Y (2002) 2D object detection and recognition. MIT Press, Cambridge
4. Amit Y, Geman D (1994) Randomized inquiries about shape; an application to handwritten digit recognition. Technical Report 401, Dept of Statistics, University of Chicago, IL, Nov 1994
5. Amit Y, Geman D (1997) Shape quantization and recognition with randomized trees. *Neural Comput* 9(7)
6. Amit Y, Geman D, Wilder K (1997) Joint induction of shape features and tree classifiers. *IEEE Trans Pattern Anal Mach Intell* 19
7. Andriluka M, Roth S, Schiele B (2008) People-tracking-by-detection and people-detection-by-tracking. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
8. Anguelov D, Taskar B, Chatalbashev V, Koller D, Gupta D, Ng A (2005) Discriminative learning of Markov random fields for segmentation of 3D scan data. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
9. Avidan S (2001) Support vector tracking. In: Proc IEEE conf computer vision and pattern recognition (CVPR), vol 1
10. Avidan S (2005) Ensemble tracking. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
11. Avidan S (2007) Ensemble tracking. *IEEE Trans Pattern Anal Mach Intell* 29(2)
12. Babenko B, Yang M-H, Belongie S (2011) Robust object tracking with online multiple instance learning. *IEEE Trans Pattern Anal Mach Intell*
13. Badrinarayanan V, Galasso F, Cipolla R (2010) Label propagation in video sequences. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
14. Bai X, Wang J, Simons D, Sapiro G (2009) Video SnapCut: robust video object cutout using localized classifiers. In: ACM SIGGRAPH
15. Ballard DH (1981) Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognit* 13(2)
16. Barinova O, Lempitsky VS, Kohli P (2010) On detection of multiple object instances using Hough transforms. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
17. Barinova O, Lempitsky VS, Kohli P (2012) On detection of multiple object instances using Hough transforms. *IEEE Trans Pattern Anal Mach Intell*

18. Barnett GH (ed) (2007) High-grade gliomas. Springer, Berlin
19. Batra D, Sukthakar R, Chen T (2008) Learning class-specific affinities for image labelling. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
20. Bauer S, Nolte L-P, Reyes M (2011) Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In: Fichtinger G, Martel A, Peters T (eds) Proc medical image computing and computer assisted intervention (MICCAI). LNCS, vol 6893. Springer, Berlin
21. Beis J, Lowe DG (1997) Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
22. Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput*
23. Belkin M, Niyogi P (2008) Towards a theoretical foundation for Laplacian-based manifold methods. *J Comput Syst Sci* 74(8)
24. Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. *IEEE Trans Pattern Anal Mach Intell* 24
25. Benfold B, Reid I (2011) Unsupervised learning of a scene-specific coarse gaze estimator. In: Proc IEEE intl conf on computer vision (ICCV), Barcelona, Spain
26. Besag J (1977) Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrika*
27. Bibby C, Reid I (2008) Robust real-time visual tracking using pixel-wise posteriors. In: Proc European conf on computer vision (ECCV). Springer, Berlin
28. Bishop CM (2006) Pattern recognition and machine learning. Springer, New York
29. Bishop CM, Svensen M, Williams CKI (1998) GTM: the generative topographic mapping. *Neural Comput*
30. Bjoerck A (1996) Numerical methods for least squares problems. Society for Industrial and Applied Mathematics (SIAM), Philadelphia
31. Blake A, Rother C, Brown M, Perez P, Torr PHS (2004) Interactive image segmentation using an adaptive GMMRF model. In: Pajdla T, Matas J (eds) Proc European conf on computer vision (ECCV), Prague, Czech Republic, May 2004. LNCS, vol 3021. Springer, Berlin
32. Blockeel H, De Raedt L, Ramon J (1998) Top-down induction of clustering trees. In: Proc intl conf on machine learning (ICML)
33. Boland MV, Murphy RF (2001) A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. *Bioinformatics* 17
34. Boland MV, Markey MK, Murphy RF (1998) Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images. *Cytometry* 33
35. Borenstein E, Ullman S (2002) Class-specific, top-down segmentation. In: Proc European conf on computer vision (ECCV). LNCS, vol 2351. Springer, Berlin
36. Bosch A, Zisermann A, Muñoz X (2007) Image classification using random forests and ferns. In: Proc IEEE intl conf on computer vision (ICCV)
37. Bourdev L, Malik J (2009) Poselets: body part detectors trained using 3D human pose annotations. In: Proc IEEE intl conf on computer vision (ICCV)
38. Boykov Y, Jolly M-P (2001) Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In: Proc IEEE intl conf on computer vision (ICCV), Vancouver, Canada, July 2001, vol 1
39. Boykov Y, Veksler O, Jolly M-P (1999) Fast approximate energy minimization via graph cuts. In: Proc IEEE intl conf on computer vision (ICCV), Kerkyra, Corfu, Greece, September 1999, vol 1
40. Braak H, Braak E (1998) Evolution of neuronal changes in the course of Alzheimer's disease. *J Neural Transm Suppl* 53
41. Bregler C, Malik J (1998) Tracking people with twists and exponential maps. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
42. Breiman L (1996) Bagging predictors. *Mach Learn* 24(2)
43. Breiman L (1999) Random forests. Technical Report TR567, UC Berkeley

44. Breiman L (2001) Random forests. *Mach Learn* 45(1)
45. Breiman L, Friedman J, Stone CJ, Olshen RA (1984) *Classification and regression trees*. Chapman and Hall/CRC, London
46. Brook A, El-Yaniv R, Isler E, Kimmel R, Meir R, Peleg D (2008) Breast cancer diagnosis from biopsy images using generic features and SVMs. Technical Report CS-2008-07, Technion, Israel
47. Brookmeyer R, Johnson E, Ziegler-Grahamm K, Arrighi HM (2007) Forecasting the global burden of Alzheimer's disease. *Alzheimer's Dement* 3(3)
48. Brostow GJ, Shotton J, Fauqueur J, Cipolla R (2008) Segmentation and recognition using structure from motion point clouds. In: *Proc European conf on computer vision (ECCV)*. Springer, Berlin
49. Brox T, Malik J (2010) Object segmentation by long term analysis of point trajectories. In: *Proc European conf on computer vision (ECCV)*. Springer, Berlin
50. Brubaker MA, Fleet DJ, Hertzmann A (2010) Physics-based person tracking using the anthropomorphic walker. *Int J Comput Vis*
51. Budiu M, Shotton J, Murray D, Finocchio M (2011) Parallelizing the training of the Kinect body parts labeling algorithm. In: *Advances in neural information processing systems (NIPS)*. BigLearn workshop
52. Budvytis I, Badrinarayanan V, Cipolla R (2010) Label propagation in complex video sequences using semi-supervised learning. In: *Proc British machine vision conference (BMVC)*
53. Budvytis I, Badrinarayanan V, Cipolla R (2011) Semi-supervised video segmentation using tree structured graphical models. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
54. Burr A (2010) Active learning literature survey. Technical Report 2010-09-14, Univ Wisconsin Madison, Computer Sciences Technical Report
55. Calabresi P (2007) Multiple sclerosis and demyelinating conditions of the central nervous system. In: *Cecil medicine*. Saunders Elsevier, Philadelphia
56. Caruana R, Karampatziakis N, Yessenalina A (2008) An empirical evaluation of supervised learning in high dimensions. In: *Proc intl conf on machine learning (ICML)*
57. CAVIAR04. <http://homepages.inf.ed.ac.uk/rbf/CAVIAR04/>
58. Cayton L (2005) Algorithms for manifold learning. Technical Report CS2008-0923, University of California, San Diego
59. Cehovin L, Kristan M, Leonardis A (2011) An adaptive coupled-layer visual model for robust visual tracking. In: *Proc IEEE intl conf on computer vision (ICCV)*
60. Chandna P, Deswal S, Pal M (2010) Semi-supervised learning based prediction of musculoskeletal disorder risk. *J Ind Syst Eng*
61. Chapelle O, Schölkopf B, Zien A (2006) *Semi-supervised learning*. MIT Press, Cambridge
62. Chen Y, Kim T-K, Cipolla R (2011) Silhouette-based object phenotype recognition using 3D shape priors. In: *Proc IEEE intl conf on computer vision (ICCV)*, Barcelona, Spain
63. Cheung V, Frey BJ, Jovic N (2005) Video epitomes. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*, June 2005, vol 1
64. Chipman H, George EI, McCulloch RE (1997) Bayesian CART model search. *J Am Stat Assoc* 93
65. Cho TS, Joshi N, Zitnick CL, Kang SB, Szeliski R, Freeman WT (2010) A content-aware image prior. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
66. Chockalingam P, Pradeep N, Birchfield S (2009) Adaptive fragments-based tracking of non-rigid objects using level sets. In: *Proc IEEE intl conf on computer vision (ICCV)*
67. Chum O, Zisserman A (2007) An exemplar model for learning object classes. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
68. Cohn DA, Ghahramani Z, Jordan MI (1996) Active learning with statistical models. *J Artif Intell Res* 4
69. Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. *IEEE Trans Pattern Anal Mach Intell* 24(5)

70. Cootes TF, Ionita MC, Lindner C, Sauer P (2012) Robust and accurate shape model fitting using random forest regression voting. In: Proc European conf on computer vision (ECCV)
71. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261(5123)
72. Corso JJ, Sharon E, Dube S, El-saden S, Sinha U, Yuille A (2008) Efficient multilevel brain tumor segmentation with integrated Bayesian model classification. *Trans Med Imaging* 27(5)
73. Cox TF, Cox MAA (2001) Multidimensional scaling. Chapman and Hall, London
74. Crammer K, Singer Y (2001) On the algorithmic implementation of multi-class SVMs. *J Mach Learn Res*
75. Cremers D, Funka-Lea G (2006) Dynamical statistical shape priors for level set based tracking. *IEEE Trans Pattern Anal Mach Intell*
76. Criminisi A, Sharp T, Blake A (2008) GeoS: geodesic image segmentation. In: Proc European conf on computer vision (ECCV). Springer, Berlin
77. Criminisi A, Shotton J, Bucciarelli S (2009) Decision forests with long-range spatial context for organ localization in CT volumes. In: MICCAI workshop on probabilistic models for medical image analysis (PMMIA)
78. Criminisi A, Shotton J, Robertson D, Konukoglu E (2010) Regression forests for efficient anatomy detection and localization in CT studies. In: MICCAI workshop on medical computer vision: recognition techniques and applications in medical imaging, Beijing. Springer, Berlin
79. Criminisi A, Shotton J, Konukoglu E (2011) Online tutorial on decision forests. <http://research.microsoft.com/projects/decisionforests>
80. Criminisi A, Shotton J, Konukoglu E (2012) Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Found Trends Comput Graph Vis* 7(2–3)
81. Csurka G, Dance CR, Fan L, Willamowski J, Bray C (2004) Visual categorization with bags of keypoints. In: ECCV intl workshop on statistical learning in computer vision
82. Cuingnet R, Prevost R, Lesage D, Cohen L, Mory B, Ardon R (2012) Automatic detection and segmentation of kidneys in 3D CT images using random forests. In: Proc medical image computing and computer assisted intervention (MICCAI)
83. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proc IEEE conf computer vision and pattern recognition (CVPR), June 2005, vol 2
84. Dantone M, Gall J, Fanelli G, van Gool L (2012) Real-time facial feature detection using conditional regression forests. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
85. Danuser G (2011) Computer vision in cell biology. *Cell* 147(5)
86. Dawbarn D, Allen SJ (eds) (2007) *Neurobiology of Alzheimer's disease*, 3rd edn. Oxford University Press, New York
87. De Porte J, Herbst BM, Hereman W, van Der Walt SJ (2008) An introduction to diffusion maps. *Techniques*
88. Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B Methodol* 39
89. Denison DGT, Mallick BK, Smith AFM (1998) A Bayesian CART algorithm. *Biometrika* 85
90. Deselaers T, Moller H, Clough P, Ney H, Lehmann TM (2007) The CLEF 2005 automatic medical image annotation task. *Int J Comput Vis* 74
91. Devroye L (1986) *Non-uniform random variate generation*. Springer, New York
92. Domingos P, Pazzani M, Provan G (1997) On the optimality of the simple Bayesian classifier under zero-one loss. *Mach Learn*
93. Donida Labati R, Piuri V, Scotti F (2011) ALL-IDB: the acute lymphoblastic leukemia image database for image processing. In: Proc IEEE intl conference on image processing (ICIP), September 2011

94. Driessens K, Reutemann P, Pfahringer B, Leschi C (2010) Using weighted nearest neighbour to benefit from unlabelled data. In: Advances in knowledge discovery and data mining. 10th Pacific-Asia conference
95. Duchateau N, De Craene M, Piella G, Frangi AF (2011) Characterizing pathological deviations from normality using constrained manifold learning. In: Proc medical image computing and computer assisted intervention (MICCAI)
96. Dumont M, Marée R, Geurts P, Wehenkel L (2009) Fast multi-class image annotation with random subwindows and multiple output randomized trees. In: Proc intl conference on computer vision theory and applications (VISAPP)
97. Elkan C (2003) Using the triangle inequality to accelerate  $k$ -means. In: Proc intl conf on machine learning (ICML)
98. Evans AC, Collins DL, Mills SR, Brown ED, Kelly RL, Peters TM (1993) 3D statistical neuroanatomical models from 305 MRI volumes. In: IEEE-nuclear science symposium and medical imaging conference
99. Everingham M, van Gool L, Williams C, Winn J, Zisserman A (2007) The Pascal visual object classes (VOC) challenge. <http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html>
100. Everingham M, van Gool L, Williams C, Winn J, Zisserman A (2010) The Pascal visual object classes (VOC) challenge 2010. Int J Comput Vis 88
101. Fan J, Shen X, Wu Y (2010) Closed-loop adaptation for robust tracking. In: Proc European conf on computer vision (ECCV). Springer, Berlin
102. Fanelli G, Gall J (2011) Real time head pose estimation with random regression forests. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
103. Fathi A, Balcan M, Ren X, Rehg JM (2011) Combining self training and active learning for video segmentation. In: Proc British machine vision conference (BMVC)
104. Fei-Fei L, Perona P (2005) A Bayesian hierarchical model for learning natural scene categories. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
105. Felzenszwalb P, Girshick R, McAllester D, Ramanan D (2010) Object detection with discriminatively trained part based models. IEEE Trans Pattern Anal Mach Intell
106. Fenchel M, Thesen S, Schilling A (2008) Automatic labeling of anatomical structures in MR FastView images using a statistical atlas. In: Proc medical image computing and computer assisted intervention (MICCAI)
107. Fergus R, Perona P, Zisserman A (2003) Object class recognition by unsupervised scale-invariant learning. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
108. Feulner J, Zhou SK, Seifert S, Cavallaro A, Hornegger J, Comaniciu D (2009) Estimating the body portion of CT volumes by matching histograms of visual words. In: Pluim JPW, Dawant BM (eds) Proc Intl society for optical engineering (SPIE) medical imaging
109. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24
110. Floyd RW (1962) Algorithm 97: shortest path. Commun ACM 5(6)
111. Frank A, Asuncion A (2010) UCI machine learning repository
112. Freund Y, Schapire RE (1997) A decision theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55(1)
113. Freund Y, Schapire RE (1998) Discussion of the paper "Arcing classifiers" by Leo Breiman. Processing 26(3)
114. Freund Y, Dasgupta S, Kabra M, Verma N (2007) Learning the structure of manifolds using random projections. In: Advances in neural information processing systems (NIPS)
115. Friedman J (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 2(28)
116. Friedman JH, Fayyad U (1997) On bias, variance, 0/1-loss, and the curse-of-dimensionality. Data Min Knowl Discov 1
117. Gall J, Lempitsky V (2009) Class-specific Hough forests for object detection. IEEE Trans Pattern Anal Mach Intell

118. Gall J, Lempitsky VS (2009) Class-specific Hough forests for object detection. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
119. Gall J, Razavi N, van Gool L (2010) On-line adaption of class-specific codebooks for instance tracking. In: Proc British machine vision conference (BMVC)
120. Gall J, Yao A, Razavi N, van Gool LJ, Lempitsky VS (2011) Hough forests for object detection, tracking, and action recognition. *IEEE Trans Pattern Anal Mach Intell* 33(11)
121. Ganapathi V, Plagemann C, Koller D, Thrun S (2010) Real time motion capture using a single time-of-flight camera. In: Proc IEEE conf computer vision and pattern recognition (CVPR). IEEE, New York
122. Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell* 6
123. Geman G, Jedinak B (1996) An active testing model for tracking roads from satellite images. *IEEE Trans Pattern Anal Mach Intell* 18(1)
124. Gerber S, Tasdizen T, Joshi S, Whitaker R (2009) On the manifold structure of the space of brain images. In: Proc medical image computing and computer assisted intervention (MICCAI)
125. Geremia E, Menze B, Clatz O, Konukoglu E, Criminisi A, Ayache N (2010) Spatial decision forests for MS lesion segmentation in multi-channel MR images. In: Proc medical image computing and computer assisted intervention (MICCAI). Springer, Berlin
126. Geremia E, Clatz O, Menze BH, Konukoglu E, Criminisi A, Ayache N (2011) Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance. *NeuroImage*
127. Geurts P (2002) Contributions to decision tree induction: bias/variance tradeoff and time series classification. PhD thesis, University of Liège, Belgium, May
128. Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. *Mach Learn* 36(1)
129. Geurts P, Marée R, Wehenkel L (2006) Segment and combine: a generic approach for supervised learning of invariant classifiers from topologically structured data. In: Proc of the machine learning conference of Belgium and The Netherlands (Benelearn)
130. Geurts P, Wehenkel L, d Alché-Buc F (2006) Kernelizing the output of tree-based methods. In: Proc intl conf on machine learning (ICML)
131. Girshick R, Shotton J, Kohli P, Criminisi A, Fitzgibbon A (2011) Efficient regression of general-activity human poses from depth images. In: Proc IEEE intl conf on computer vision (ICCV)
132. Glesner S, Koller D (1995) Constructing flexible dynamic belief networks from first-order probabilistic knowledge bases. In: ECSQARU
133. Glocker B, Feulner J, Criminisi A, Haynor DR, Konukoglu E (2012) Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. In: Proc medical image computing and computer assisted intervention (MICCAI)
134. Glocker B, Pauly O, Konukoglu E, Criminisi A (2012) Joint classification-regression forests for spatially structured multi-object segmentation. In: Proc European conf on computer vision (ECCV). Springer, Berlin
135. Godec M, Roth PM, Bischof H (2011) Hough-based tracking of non-rigid objects. In: Proc IEEE intl conf on computer vision (ICCV)
136. Gooya A, Pohl KM, Bilello M, Biros G, Davatzikos C (2011) Joint segmentation and deformable registration of brain scans guided by a tumor growth model. In: Proc medical image computing and computer assisted intervention (MICCAI)
137. Gorriz L, Menze BH, Weber M-A, Kelm BM, Hamprecht FA (2007) Semi-supervised tumor detection in magnetic resonance spectroscopic images using discriminative random fields. In: Proc annual symposium of the German association for pattern recognition (DAGM)
138. Gould S, Fulton R, Koller D (2009) Decomposing a scene into geometric and semantically consistent regions. In: Proc IEEE intl conf on computer vision (ICCV)
139. Grabner H, Grabner M, Bischof H (2006) Real-time tracking via on-line boosting. In: Proc British machine vision conference (BMVC)
140. Grabner H, Leistner C, Bischof H (2008) Semi-supervised on-line boosting for robust tracking. In: Proc European conf on computer vision (ECCV). Springer, Berlin

141. Grauman K, Darrell T (2005) The pyramid match kernel: discriminative classification with sets of image features. In: Proc IEEE intl conf on computer vision (ICCV)
142. Gray KR, Aljabar P, Heckeman RA, Hammers A, Rueckert D (2011) Random forest-based manifold learning for classification of imaging data in dementia. In: Proc medical image computing and computer assisted intervention (MICCAI)
143. Gray KR, Aljabar P, Heckemann RA, Hammers A, Rueckert D (2013) Random forest-based similarity measures for multi-modal classification of Alzheimer's Disease. *Neuroimage* 65:167–175
144. Grest D, Woetzel J, Koch R (2005) Nonlinear body pose estimation from depth images. In: Proc annual symposium of the German association for pattern recognition (DAGM)
145. Grossberg S (1987) Competitive learning: from interactive activation to adaptive resonance. *Cogn Sci*
146. Grundmann M, Kwatra V, Han M, Essa I (2010) Efficient hierarchical graph based video segmentation. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
147. Guedl MO, Kohnen M, Keysers D, Schubert H, Wein BB, Bredno J, Lehmann TM (2002) Quality of DICOM header information for image categorization. In: SPIE storage and retrieval for image and video databases, San Diego
148. Gupta SS (1963) Probability integrals of multivariate normal and multivariate  $t$ . *Ann Math Stat* 34(3)
149. Hamm J, Ye DH, Verma R, Davatzikos C (2010) GRAM: a framework for geodesic registration on anatomical manifolds. *Med Image Anal* 14(5)
150. Hampel H, Burger K, Teipel SJ, Bokde AL, Zetterberg H, Blennow K (2008) Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. *Alzheimer's Dement* 4(1)
151. Hansson O, Zetterberg H, Buchlave P, Londos E, Blennow K, Minthon L (2006) Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study. *Lancet Neurol* 5(3)
152. Hardle W (1990) Applied non-parametric regression. Cambridge University Press, Cambridge
153. Hartley R, Zisserman A (2003) Multiple view geometry in computer vision, 2nd edn. Cambridge University Press, Cambridge
154. Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning. Springer, Berlin
155. Hastie T, Tibshirani R, Friedman J, Franklin J (2005) The elements of statistical learning: data mining, inference and prediction. *Math Intell* 27(2)
156. Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57
157. He X, Zemel RS, Carreira-Perpiñán MÁ (2004) Multiscale conditional random fields for image labeling. In: Proc IEEE conf computer vision and pattern recognition (CVPR), June 2004, vol 2
158. Heath D, Kasif S, Salzberg S (1993) Induction of oblique decision trees. *J Artif Intell Res* 2(2)
159. Heckemann RA, Keihaninejad S, Aljabar P, Gray KR, Nielsen C, Rueckert D, Hajnal JV, Hammers A, The Alzheimer's Disease Neuroimaging Initiative (2011) Automatic morphometry in Alzheimer's disease and mild cognitive impairment. *NeuroImage* 56(4)
160. Hegde C, Wakin MB, Baraniuk RG (2007) Random projections for manifold learning—proofs and analysis. In: Advances in neural information processing systems (NIPS)
161. Herholz K, Salmon E, Perani D, Baron JC, Holthoff V, Frolich L, Schonknecht P, Ito K, Mielke R, Kalbe R, Zundorf G, Delbeuck X, Pelati O, Anchisi D, Fazio F, Kerrouche N, Desgranges B, Eustache F, Beuthien-Baumann B, Menzel C, Schroder J, Kato T, Arahata Y, Henze M, Heiss WD (2002) Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. *NeuroImage* 17(1)
162. Hinrichs C, Singh V, Xu G, Johnson SC, The Alzheimer's Disease Neuroimaging Initiative (2011) Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. *NeuroImage* 55(2)

163. Hinton GE (2002) Training products of experts by minimizing contrastive divergence. *Neural Comput* 14
164. Hinton GE (2010) Learning to represent visual input. *Philos Trans R Soc B* 365
165. Ho TK (1995) Random decision forests. In: Proc intl conf on document analysis and recognition
166. Ho TK (1998) The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 20(8)
167. Ho S, Bullitt E, Gerig G (2002) Level-set evolution with region competition: automatic 3-D segmentation of brain tumors. In: Proc intl conf on pattern recognition (ICPR)
168. Hoiem D, Sukthankar R, Schneiderman H, Huston L (2004) Object-based image retrieval using the statistical structure of images. *J Mach Learn Res* 02
169. IEEE-IMS Workshop on Information Theory and VA Statistics, Alexandria, Oct 1994
170. Isgum I, Staring M, Rutten A, Prokop M, Viergever MA, van Ginneken B (2009) Multi-atlas-based segmentation with local decision fusion: application to cardiac and aortic segmentation in CT scans. *Trans Med Imaging* 28(7)
171. Javed O, Ali S, Shah M (2005) Online detection and classification of moving objects using progressively improving detectors. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
172. Joachims T (1999) Making large-scale SVM learning practical. In: Schölkopf B, Burges C, Smola A (eds) *Advances in kernel methods—support vector learning*. MIT Press, Cambridge
173. John GH (1995) Robust linear discriminant trees. In: Fifth intl workshop on artificial intelligence and statistics
174. Jovic N, Frey BJ, Kannan A (2003) Epitomic analysis of appearance and shape. In: Proc IEEE intl conf on computer vision (ICCV), Nice, France, October 2003, vol 1
175. Jolliffe IT (1986) *Principal component analysis*. Springer, Berlin
176. Julesz B (1981) Textons, the elements of texture perception, and their interactions. *Nature* 290(5802)
177. Jurie F, Triggs B (2005) Creating efficient codebooks for visual recognition. In: Proc IEEE intl conf on computer vision (ICCV), vol 1
178. Kalal Z, Matas J, Mikolajczyk K (2010) P-N learning: bootstrapping binary classifiers by structural constraints. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
179. Kannan A, Winn J, Rother C (2006) Clustering appearance and shape by learning jigsaws. In: *Advances in neural information processing systems (NIPS)*
180. Kelm BM, Mittal S, Zheng Y, Tsybal A, Bernhardt D, Vega-Higuera F, Zhou KS, Meer P, Comaniciu D (2011) Detection, grading and classification of coronary stenoses in computed tomography angiography. In: Proc medical image computing and computer assisted intervention (MICCAI)
181. Keyser D, Dahmen J, Ney H (2001) Invariant classification of red blood cells: a comparison of different approaches. In: *Bildverarbeitung für die Medizin'01*
182. Kim TK, Stenger B, Woodley T, Cipolla R (2010) Online multiple classifier boosting for object tracking. In: *Online learning for computer vision workshop*
183. Klafki H-W, Staufenbiel M, Kornhuber J, Wiltfang J (2006) Therapeutic approaches to Alzheimer's disease. *Brain* 129(11)
184. Klein S, Staring M, Murphy K, Viergever MA, Pluim JP (2010) Elastix: a toolbox for intensity-based medical image registration. *Trans Med Imaging* 29
185. Klunk WE, Engler H, Nordberg A, Wang YM, Blomqvist G, Holt DP, Bergstrom M, Savitcheva I, Huang GF, Estrada S, Ausén B, Debnath ML, Barletta J, Price JC, Sandell J, Lopresti BJ, Wall A, Koivisto P, Antoni G, Mathis CA, Långström B (2004) Imaging brain amyloid in Alzheimer's disease with Pittsburgh compound-B. *Ann Neurol* 55(3)
186. Knoop S, Vacek S, Dillmann R (2006) Sensor fusion for 3D human body tracking with an articulated 3D body model. In: Proc IEEE intl conf on robotics and automation (ICRA)
187. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: 14th intl joint conference on artificial intelligence (IJCAI), vol 2

188. Kohli P, Torr PHS (2005) Efficiently solving dynamic Markov random fields using graph cuts. In: Proc IEEE intl conf on computer vision (ICCV), Beijing, China, October 2005, vol 2
189. Koller D, Friedman N (2009) Probabilistic graphical models: principles and techniques. MIT Press, Cambridge
190. Kolmogorov V (2006) Convergent tree-reweighted message passing for energy minimization. *IEEE Trans Pattern Anal Mach Intell* 28(10)
191. Kolmogorov V, Boykov Y (2005) What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. In: Proc IEEE intl conf on computer vision (ICCV)
192. Kotschieder P, Rota Buló S, Bischof H, Pelillo M (2011) Structured class-labels in random forests for semantic image labelling. In: Proc IEEE intl conf on computer vision (ICCV), Barcelona, Spain
193. Konukoglu E, Criminisi A, Pathak S, Robertson D, White S, Haynor D, Siddiqui K (2011) Robust linear registration of CT images using random regression forests. In: Proc intl society for optical engineering (SPIE) medical imaging
194. Konukoglu E, Glocker B, Zikic D, Criminisi A (2012) Neighborhood approximation forests. In: Proc medical image computing and computer assisted intervention (MICCAI)
195. Kristan M, Skocaj D, Leonardis A (2008) Incremental learning with Gaussian mixture models. In: Computer vision winter workshop (CVWW), Moravske Toplice, Slovenia
196. Kumar S, Hebert M (2003) Discriminative random fields: a discriminative framework for contextual interaction in classification. In: Proc IEEE intl conf on computer vision (ICCV), October 2003, vol 2
197. Kwon JS, Lee KM (2009) Tracking of a non-rigid object via patch-based dynamic appearance modeling and adaptive Basin Hopping Monte Carlo sampling. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
198. Lafferty J, McCallum A, Pereira F (2001) Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proc intl conf on machine learning (ICML)
199. Lampert CH (2008) Kernel methods in computer vision. *Found Trends Comput Graph Vis* 4(3)
200. Lampert C, Blaschko M, Hofmann T (2008) Beyond sliding windows: object localization by efficient subwindow search. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
201. Landau SM, Harvey D, Madison CM, Reiman EM, Foster NL, Aisen PS, Petersen RC, Shaw LM, Trojanowski JQ, Jack CR Jr., Weiner MW, Jagust WJ (2010) Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology* 75(3)
202. Langbaum JBS, Chen K, Lee W, Reschke C, Bandy D, Fleisher AS, Alexander GE, Foster NL, Weiner MW, Koeppe RA, Jagust WJ, Reiman EM, The Alzheimer's Disease Neuroimaging Initiative (2009) Categorical and correlational analyses of baseline fluoro-deoxyglucose positron emission tomography images from the Alzheimer's disease neuroimaging initiative (ADNI). *NeuroImage* 45(4)
203. Laptev I, Marszalek M, Schmid C, Rozenfeld B (2008) Learning realistic human actions from movies. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
204. Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
205. Lee CH, Wang S, Murtha A, Brown M, Greiner R (2008) Segmenting brain tumors using pseudo-conditional random fields. In: Proc medical image computing and computer assisted intervention (MICCAI)
206. Lee H, Grosse R, Ranganath R, Ng AY (2009) Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proc intl conf on machine learning (ICML)
207. Lee YJ, Kim J, Grauman K (2011) Key-segments for video object segmentation. In: Proc IEEE intl conf on computer vision (ICCV)

208. Leibe B, Schiele B (2003) Interleaved object categorization and segmentation. In: Proc British machine vision conference (BMVC), vol II
209. Leibe B, Leonardis A, Schiele B (2004) Combined object categorization and segmentation with an implicit shape model. In: ECCV'04 workshop on statistical learning in computer vision, May 2004
210. Leibe B, Leonardis A, Schiele B (2008) Robust object detection with interleaved categorization and segmentation. *Int J Comput Vis* 77(1–3)
211. Leistner C, Saffari A, Santner J, Bischoff H (2009) Semi-supervised random forests. In: Proc IEEE intl conf on computer vision (ICCV)
212. Lempitsky V, Verhoeck M, Noble A, Blake A (2009) Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography. In: Workshop on functional imaging and modelling of the heart (FIMH). Springer, Berlin
213. Lepetit V, Fua P (2006) Keypoint recognition using randomized trees. *IEEE Trans Pattern Anal Mach Intell*
214. Lepetit V, Lagger P, Fua P (2005) Randomized trees for real-time keypoint recognition. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
215. Lezama J, Alahari K, Sivic J, Laptev I (2011) Track to the future: spatio-temporal video segmentation with long-range motion cues. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
216. Lezoray O, Elmoataz A, Cardot H (2003) A color object recognition scheme: application to cellular sorting. *Mach Vis Appl* 14
217. Li SZ (1995) Markov random field modeling in computer vision. Springer, Berlin
218. Li J, Allinson N (2008) A comprehensive review of current local features for computer vision. *Neurocomputing* 71
219. Li L-J, Fei-Fei L (2007) What, where and who? Classifying events by scene and object recognition. In: Proc IEEE intl conf on computer vision (ICCV)
220. Li Y, Sun J, Shum H-Y (2005) Video object cut and paste. *ACM Trans Graph* 24
221. Liang Z-P, Lauterbur PC (1999) Principles of magnetic resonance imaging: a signal processing perspective. IEEE Press/Wiley, New York
222. Liaw A, Wiener M (2002) Classification and regression by random forest. *R News* 2
223. Lin Y, Jeon Y (2002) Random forests and adaptive nearest neighbors. *J Am Stat Assoc*
224. Lindner C, Thiagarajah S, Wilkinson JM, arcOGEN Consortium, Wallis GA, Cootes TF (2012) Accurate fully automatic femur segmentation in pelvic radiographs using regression voting. In: Proc medical image computing and computer assisted intervention (MICCAI)
225. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60(2)
226. Ma J (2008) Dixon techniques for water and fat imaging. *J Magn Reson Imaging*
227. MacQueen JB (1967) Some methods for classification and analysis of multivariate observations. In: Proc of 5th Berkeley symposium on mathematical statistics and probability. University of California Press, Berkeley
228. Maji S, Malik J (2009) Object detection using a max-margin Hough transform. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
229. Malik J, Belongie S, Leung T, Shi J (2001) Contour and texture analysis for image segmentation. *Int J Comput Vis* 43(1)
230. Malisiewicz T, Gupta A, Efros AA (2011) Ensemble of exemplar-SVMs for object detection and beyond. In: Proc IEEE intl conf on computer vision (ICCV), Barcelona, Spain
231. Mardia KV, Kent JT, Bibby JM (1979) Multivariate analysis, 5th edn. Academic Press, London
232. Marée R (2012) Towards generic image classification an extensive empirical study. Technical report, University of Liège
233. Marée R, Geurts P, Visimberga G, Piater J, Wehenkel L (2003) An empirical comparison of machine learning algorithms for generic image classification. In: Coenen F, Preece A, Macintosh AL (eds) Proc of the 23rd SGAI intl conference on innovative techniques and

- applications of artificial intelligence, research and development in intelligent systems XX. Springer, Berlin
234. Marée R, Geurts P, Piater J, Wehenkel L (2004) A generic approach for image classification based on decision tree ensembles and local sub-windows. In: Proc Asian conf on computer vision (ACCV), vol 2
  235. Marée R, Geurts P, Piater J, Wehenkel L (2005) Random subwindows for robust image classification. In: Proc IEEE conf computer vision and pattern recognition (CVPR), vol 1. IEEE, New York
  236. Marée R, Geurts P, Wehenkel L (2007) Content-based image retrieval by indexing random subwindows with randomized trees. In: Proc Asian conf on computer vision (ACCV). LNCS, vol 4844. Springer, Berlin
  237. Marée R, Geurts P, Wehenkel L (2007) Random subwindows and extremely randomized trees for image classification in cell biology. *Data Mining Inf* 8(S1). BMC Cell Biology supplement on Workshop of Multiscale Biological Imaging, July 2007
  238. Marée R, Geurts P, Wehenkel L (2009) Content-based image retrieval by indexing random subwindows with randomized trees. *IPSJ Trans Comput Vis Appl* 1(1) (open-access)
  239. Marée R, Stevens B, Geurts P, Guern Y, Mack P (2009) A machine learning approach for material detection in hyperspectral images. In: Proc 6th IEEE workshop on object tracking and classification beyond and in the visible spectrum (CVPR09). IEEE, New York
  240. Marée R, Denis P, Wehenkel L, Geurts P (2010) Incremental indexing and distributed image search using shared randomized vocabularies. In: Proc 11th ACM intl conference on multimedia information retrieval (MIR), March 2010. ACM Press, New York
  241. Marée R, Stevens B, Rollus L, Rocks N, Moles-Lopez X, Salmon I, Cataldo D, Wehenkel L (2012) A rich Internet application for remote visualization and collaborative annotation of digital slide images in histology and cytology. In: BMC diagnostic pathology, proc 12th European congress on telepathology and 5th intl congress on virtual microscopy
  242. Matthews L, Ishikawa T, Baker S (2004) The template update problem. *IEEE Trans Pattern Anal Mach Intell*
  243. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease—report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology* 34(7)
  244. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr., Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH (2011) The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement* 7(3)
  245. Menze BH, Leemput KV, Lashkari D, Weber M-A, Ayache N, Golland P (2010) A generative model for brain tumor segmentation in multi-modal images. In: Proc medical image computing and computer assisted intervention (MICCAI)
  246. Menze B, Kelm BM, Splittthoff DN, Koethe U, Hamprecht FA (2011) On oblique random forests. In: Proc European conf on machine learning (ECML/PKDD)
  247. Microsoft Corporation Kinect for Windows and Xbox 360
  248. Mikolajczyk K, Schmid C (2004) Scale and affine invariant interest point detectors. *Int J Comput Vis* 60(1)
  249. Mikolajczyk K, Tuytelaars T, Schmid C, Zisserman A, Matas J, Schaffalitzky F, Kadir T, van Gool L (2005) A comparison of affine region detectors. *Int J Comput Vis* 65(1/2)
  250. Montillo A (2011) Context selective decision forests and their application to lung segmentation in CT images. In: MICCAI workshop on pulmonary image analysis
  251. Montillo A, Ling H (2009) Age regression from faces using random forests. In: Proc intl conf on image processing (ICIP)
  252. Montillo A, Shotton J, Winn J, Iglesias J, Metaxas D, Criminisi A (2011) Entangled decision forests and their application for semantic segmentation of CT images. In: Proc information

- processing in medical imaging (IPMI). Springer, Berlin
253. Moosmann F, Triggs B, Jurie F (2006) Fast discriminative visual codebooks using randomized clustering forests. In: *Advances in neural information processing systems (NIPS)*
  254. Moosmann F, Nowak E, Jurie F (2008) Randomized clustering forests for image classification. *IEEE Trans Pattern Anal Mach Intell* 30(9)
  255. Moreno-Noguer F, Sanfeliu A, Samaras D (2008) Dependent multiple cue integration for robust tracking. *IEEE Trans Pattern Anal Mach Intell*
  256. Motter R, Vigopelfrey C, Kholodenko D, Barbour R, Johnsonwood K, Galasko D, Chang L, Miller B, Clark C, Green R, Olson D, Southwick P, Wolfert R, Munroe B, Lieberburg I, Seubert P, Schenk D (1995) Reduction of beta-amyloid peptide(42) in the cerebrospinal fluid of patients with Alzheimer's disease. *Ann Neurol* 38(4)
  257. Müller J, Arens M (2010) Human pose estimation with implicit shape models. In: *ARTEMIS*
  258. Müller A, Nowozin S, Lampert CH (2012) Information theoretic clustering using minimum spanning trees. In: *Proc annual symposium of the German association for pattern recognition (DAGM)*
  259. Murphy RF (2011) An active role for machine learning in drug development. *Nat Chem Biol* 7
  260. Murthy SK, Kasif S, Salzberg S (1994) A system for induction of oblique decision trees. [arXiv:cs/9408103](https://arxiv.org/abs/cs/9408103)
  261. Mutch J, Lowe DG (2006) Multiclass object recognition with sparse, localized features. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
  262. Nadler B, Lafon S, Coifman RR, Kevrekidis IG (2005) Diffusion maps, spectral clustering and eigenfunctions of Fokker-Plank operators. In: *Advances in neural information processing systems (NIPS)*
  263. Neal RM (2001) Annealed importance sampling. *Stat Comput* 11
  264. Nejhun SM, Ho J, Yang M-H (2008) Visual tracking with histograms and articulating blocks. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
  265. Nelder JA, Mead R (1965) A simplex method for function minimization. *Comput J* 7(4)
  266. Nistér D, Stewénus H (2006) Scalable recognition with a vocabulary tree. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
  267. Nowak E, Jurie F, Triggs B (2006) Sampling strategies for bag-of-features image classification. In: *Proc European conf on computer vision (ECCV)*. Springer, Berlin
  268. Nowozin S (2012) Improved information gain estimates for decision tree induction. In: *Proc intl conf on machine learning (ICML)*
  269. Nowozin S, Lampert CH (2009) Global connectivity potentials for random field models. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
  270. Nowozin S, Lampert CH (2011) Structured learning and prediction in computer vision. *Found Trends Comput Graph Vis* 6(3–4)
  271. Nowozin S, Rother C, Bagon S, Sharp T, Yao B, Kohli P (2011) Decision tree fields. In: *Proc IEEE intl conf on computer vision (ICCV)*
  272. Oberg J, Eguro K, Bittner R, Forin A (2012) Random decision tree body part recognition using FPGAs. In: *Proc 22nd int conf on field programmable logic and applications (FPL)*
  273. O'Hara S, Draper BA (2012) Scalable action recognition with a subspace forest. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
  274. Ojala T, Pietikainen M, Harwood D (1996) A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit* 29
  275. Okada R (2009) Discriminative generalized Hough transform for object detection. In: *Proc IEEE intl conf on computer vision (ICCV)*
  276. Oliva A, Torralba A (2006) Building the gist of a scene: the role of global image features in recognition. *Vis Percept Prog Brain Res* 155(1)
  277. Opelt A, Pinz A, Zisserman A (2008) Learning an alphabet of shape and appearance for multi-class object detection. *Int J Comput Vis*
  278. Orlov N, Shamir L, Macura T, Johnston J, Eckley DM, Goldberg I (2008) WND-CHARM: multi-purpose image classification using compound transforms. *Pattern Recognit Lett* 29(11)

279. Ozuysal M, Fua P, Lepetit V (2007) Fast keypoint recognition in ten lines of code. In: Proc IEEE conf computer vision and pattern recognition (CVPR), June 2007
280. Ozuysal M, Calonder M, Lepetit V, Fua P (2010) Fast keypoint recognition using random ferns. *IEEE Trans Pattern Anal Mach Intell* 32(3)
281. Pang J, Huang Q, Jiang S (2008) Multiple instance boost using graph embedding based decision stump for pedestrian detection. In: Proc European conf on computer vision (ECCV). Springer, Berlin
282. Parzen E (1962) On estimation of a probability density function and mode. *Ann Math Stat* 33
283. Pathak S, Criminisi A, White S, Munasinghe I, Sparks B, Robertson D, Siddiqui K (2011) Automatic semantic annotation and validation of anatomy in DICOM CT images. In: Proc intl society for optical engineering (SPIE) medical imaging
284. Patwardhan MB, McCrory DC, Matchar DB, Samsa GP, Rutschmann OT (2004) Alzheimer disease: operating characteristics of PET—a meta-analysis. *Radiology* 231(1)
285. Pauly O, Mateus D, Navab N (2010) ImageCLEF 2010 working notes on the modality classification subtask. Technical report, Technische Universitat Munchen
286. Pauly O, Glocker B, Criminisi A, Mateus D, Martinez Möller A, Nekolla S, Navab N (2011) Fast multiple organs detection and localization in whole-body MR Dixon sequences. In: Proc medical image computing and computer assisted intervention (MICCAI), Toronto
287. Payet N, Todorovic S (2010)  $(RF)^2$ —random forest random field. In: Advances in neural information processing systems (NIPS)
288. Percannella G, Foggia P, Soda P Contest on HEP-2 cells classification. <http://mivia.unisa.it/hep2contest/index.shtml>
289. Petersen RC (2004) Mild cognitive impairment as a diagnostic entity. *J Intern Med* 256(3)
290. PETS. <http://www.cvg.rdg.ac.uk/slides/pets.html>
291. PETS10. <http://www.cvg.rdg.ac.uk/PETS2010/a.html>
292. Plackett RL (1954) A reduction formula for normal multivariate integrals. *Biometrika* 41
293. Plagemann C, Ganapathi V, Koller D, Thrun S (2010) Real-time identification and localization of body parts from depth images. In: Proc IEEE intl conf on robotics and automation (ICRA)
294. Popuri K, Cobzas D, Murtha A, Jägersand M (2011) 3D variational brain tumor segmentation using Dirichlet priors on a clustered feature set. *Int J Comput Assisted Radiol Surg*
295. Porikli FM (2005) Integral histogram: a fast way to extract histograms in Cartesian spaces. In: Proc IEEE conf computer vision and pattern recognition (CVPR), vol 1
296. Prasad M, Zisserman A, Fitzgibbon AW, Kumar MP, Torr PHS (2006) Learning class-specific edges for object detection and segmentation. In: ICVGIP
297. Prastawa M, Bullitt E, Ho S, Gerig G (2004) A brain tumor segmentation framework based on outlier detection. *Med Image Anal*
298. Predictions Workshop. Models, Computing. 31st Symp on the Interface: Computing Science, and IL. Statistics. Schaumburg, Jun 1999
299. Price SJ, Peña A, Burnet NG, Jena R, Green HAL, Carpenter TA, Pickard JD, Gillard JH (2004) Tissue signature characterisation of diffusion tensor abnormalities in cerebral gliomas. *Eur Radiol* 14
300. Prima S, Ayache N, Barrick T, Roberts N (2001) Maximum likelihood estimation of the bias field in MR brain images: investigating different modelings of the imaging process. In: Proc medical image computing and computer assisted intervention (MICCAI). LNCS, vol 2208. Springer, Berlin
301. Prima S, Ourselin S, Ayache N (2002) Computation of the mid-sagittal plane in 3D brain images. *Trans Med Imaging* 21(2)
302. Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann, San Mateo
303. Rabinovich A, Vedaldi A, Galleguillos C, Wiewiora E, Belongie S (2007) Objects in context. In: Proc IEEE intl conf on computer vision (ICCV)
304. Ram P, Gray AG (2011) Density estimation trees. In: Proc ACM SIGKDD intl conf on knowledge discovery and data mining (KDD)

305. Rasmussen CE, Williams C (2006) Gaussian processes for machine learning. MIT Press, Cambridge
306. Razavi N, Gall J, Van Gool L (2010) Backprojection revisited: scalable multi-view object detection and similarity metrics for detections. In: Proc European conf on computer vision (ECCV). Springer, Berlin
307. Ren X, Malik J (2007) Tracking as repeated figure/ground segmentation. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
308. Rey D (2002) Détection et quantification de processus évolutifs dans des images médicales tridimensionnelles : application à la sclérose en plaques. Thèse de sciences, Université de Nice Sophia-Antipolis, October (in French)
309. Roberts MG, Cootes TF, Adams JE (2012) Automatic location of vertebrae on DXA images using random forest regression. In: Proc medical image computing and computer assisted intervention (MICCAI)
310. Rocca WA, Hofman A, Brayne C, Breteler MMB, Clarke ML, Copeland JRM, Dartigues J-F, Engedal K, Hagnell O, Heeren TJ, Jonker C, Lindsay J, Lobo A, Mann AH, Mls PK, Morgan K, O'Connor DLW, da Silva Droux A, Sulkava R, Kay DWK, Amaducci L (1991) Frequency and distribution of Alzheimer's disease in Europe: a collaborative study of 1980–1990 prevalence findings. *Ann Neurol* 30(3)
311. Rogez G, Rihan J, Ramalingam S, Orrite C, Torr PHS (2008) Randomized trees for human pose detection. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
312. Rosenberg C, Hebert M, Schneiderman H (2005) Semi-supervised self-training of object detection models. In: 17-th IEEE workshop on applications of computer vision
313. Roses AD, Saunders AM (1997) ApoE, Alzheimer's disease, and recovery from brain stress. *Cerebrovasc Pathol Alzheimer's Dis* 826
314. Roth S, Black MJ (2007) Steerable random fields. In: Proc IEEE intl conf on computer vision (ICCV)
315. Rother C, Kolmogorov V, Blake A (2004) GrabCut—interactive foreground extraction using iterated graph cuts. *ACM Trans Graph* 23(3)
316. Rudin M (2005) Molecular imaging: basic principles and applications in biomedical research. Imperial College Press, London
317. Saffari A, Leistner C, Santner J, Godec M, Bischoff H (2009) On-line random forests. In: ICCV workshop on on-line learning for computer vision
318. Santner J, Leistner C, Saffari A, Pock T, Bischof H (2010) PROST: parallel robust online simple tracking. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
319. Saul LK, Jordan MI (1996) Exploiting tractable substructures in intractable networks. In: Advances in neural information processing systems (NIPS)
320. Schapire RE (1990) The strength of weak learnability. *Mach Learn* 5(2)
321. Schindler G, Brown M, Szeliski R (2007) City-scale location recognition. In: Proc IEEE conf computer vision and pattern recognition (CVPR), Minneapolis, June 2007
322. Schmidt M, Levner I, Greiner R, Murtha A, Bistriz A (2005) Segmenting brain tumors using alignment-based features. In: ICMLA
323. Schnitzspan P, Roth S, Schiele B (2010) Automatic discovery of meaningful object parts with latent CRFs. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
324. Schroff F, Criminisi A, Zisserman A (2008) Object class segmentation using random forests. In: Proc British machine vision conference (BMVC)
325. Schuster S, Leistner C, Roth PM, van Gool L, Bischof H (2011) On-line Hough forests. In: Proc British machine vision conference (BMVC)
326. Seber GAF, Wild CJ (1989) Non linear regression. Wiley, New York
327. Seemann E, Schiele B (2006) Cross-articulation learning for robust detection of pedestrians. In: Proc annual symposium of the German association for pattern recognition (DAGM)
328. Seifert S, Barbu A, Zhou SK, Liu D, Feulner J, Huber M, Sühling M, Cavallaro A, Comaniciu D (2009) Hierarchical parsing and semantic navigation of full body CT data. In: Plum JPW, Dawant BM (eds) Proc intl society for optical engineering (SPIE) medical imaging
329. Selkoe DJ (1991) The molecular pathology of Alzheimer's disease. *Neuron* 6(4)

330. Settles B (2010) Active learning literature survey. Technical report, Computer Sciences Technical Report 1648, University of Wisconsin Madison
331. Shamir L, Macura T, Orlov N, Eckley DM, Goldberg IG (2008) IICBU 2008—a benchmark suite for biological imaging. In: 3rd workshop on bio-image informatics: biological imaging, computer vision and data mining
332. Shamir L, Delaney J, Orlov N, Eckley DM, Goldberg IG (2010) Pattern recognition software and techniques for biological image analysis. *PLoS Comput Biol* 6(11)
333. Sharp T (2008) Implementing decision trees and forests on a GPU. In: Proc European conf on computer vision (ECCV). Springer, Berlin
334. Shawe-Taylor J, Cristianini N (2004) Kernel methods for pattern analysis. Cambridge University Press, Cambridge
335. Shi T, Horvath S (2006) Unsupervised learning with random forest predictors. *J Comput Graph Stat* 15
336. Shi J, Malik J (1997) Normalized cuts and image segmentation. In: Proc IEEE conf computer vision and pattern recognition (CVPR), Washington, DC, USA
337. Shimizu A, Ohno R, Ikegami T, Kobatake H (2006) Multi-organ segmentation in three-dimensional abdominal CT images. *Int J Comput Assisted Radiol Surg* 1
338. Shotton J, Winn J, Rother C, Criminisi A (2006) TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: Proc European conf on computer vision (ECCV). Springer, Berlin
339. Shotton J, Blake A, Cipolla R (2008) Efficiently combining contour and texture cues for object recognition. In: Proc British machine vision conference (BMVC)
340. Shotton J, Blake A, Cipolla R (2008) Multiscale categorical object recognition using contour fragments. *IEEE Trans Pattern Anal Mach Intell* 30(7)
341. Shotton J, Johnson M, Cipolla R (2008) Semantic texton forests for image categorization and segmentation. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
342. Shotton J, Winn JM, Rother C, Criminisi A (2009) TextonBoost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *Int J Comput Vis* 81(1)
343. Shotton J, Fitzgibbon AW, Cook M, Sharp T, Finocchio M, Moore R, Kipman A, Blake A (2011) Real-time human pose recognition in parts from a single depth image. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
344. Shotton J, Girshick R, Fitzgibbon A, Sharp T, Cook M, Finocchio M, Moore R, Kohli P, Criminisi A, Kipman A, Blake A (2012) Efficient human pose estimation from single depth images. *IEEE Trans Pattern Anal Mach Intell*
345. Siddiqui M, Medioni G (2010) Human pose estimation from a single view point, real-time range sensor. In: CVCG at CVPR
346. Sigal L, Bhatia S, Roth S, Black MJ, Isard M (2004) Tracking loose-limbed people. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
347. Silverman BW (1986) Density estimation. Chapman and Hall, London
348. Sivic J, Zisserman A (2003) Video Google: a text retrieval approach to object matching in videos. In: Proc IEEE intl conf on computer vision (ICCV)
349. Skilling J (2010) Maximum entropy and Bayesian methods. Kluwer Academic, Dordrecht
350. Smith SM (2002) Fast robust automated brain extraction. *Hum Brain Mapp*
351. Smola AJ, Schölkopf B (2003) A tutorial on support vector regression. Technical report, Statistics and Computing
352. Sonnenburg S, Rätsch G, Schäfer C, Schölkopf B (2006) Large scale multiple kernel learning. *J Mach Learn Res* 7
353. Souplet J-C, Lebrun C, Ayache N, Malandain G (2008) An automatic segmentation of T2-FLAIR multiple sclerosis lesions. In: The MIDAS journal—MS lesion segmentation (MICCAI 2008 workshop)
354. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR Jr., Kaye J, Montine TJ, Park DC, Reiman EM, Rowe CC, Siemers E, Stern Y, Yaffe K, Carrillo MC, Thies B, Morrison-Bogorad M, Wagster MV, Phelps CH (2011) Toward defining

- the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement* 7(3)
355. Statnikov A, Wang L, Aliferis CA (2008) A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinf*
  356. Stern O, Marée R, Aceto J, Jeanray N, Muller M, Wehenkel L, Geurts P (2011) Automatic localization of interest points in zebrafish images with tree-based methods. In: *Proc 6th IAPR intl conference on pattern recognition in bioinformatics. Lecture notes in bioinformatics*. Springer, Berlin
  357. Strecha C, Fransens R, van Gool L (2006) Combined depth and outlier estimation in multi-view stereo. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
  358. Styner M, Lee J, Chin B, Chin MS, Commowick O, Tran H, Markovic-Plese S, Jewells V, Warfield SK (2008) 3D segmentation in the clinic: a grand challenge II: MS lesion segmentation. *MIDAS J*
  359. Sudderth EB, Jordan MI (2008) Shared segmentation of natural scenes using dependent Pitman-Yor processes. In: *Advances in neural information processing systems (NIPS)*
  360. Sun M, Kohli P, Shotton J (2012) Conditional regression forests for human pose estimation. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
  361. Sutton C, McCallum A (2006) *An introduction to conditional random fields for relational learning*. MIT Press, Cambridge. Chap 4
  362. Szekely GJ, Rizzo ML (2004) Testing for equal distributions in high dimensions. *Interstat*, Nov 2004.
  363. Szeliski R, Zabih R, Scharstein D, Veksler O, Kolmogorov V, Agarwala A, Tappen ML, Rother C (2008) A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *IEEE Trans Pattern Anal Mach Intell* 30(7)
  364. Szummer M, Jaakkola T (2001) Partially labelled classification with Markov random walks. In: *Advances in neural information processing systems (NIPS)*
  365. Szummer M, Kohli P, Hoiem D (2008) Learning CRFs using graph cuts. In: *Proc European conf on computer vision (ECCV)*. Springer, Berlin
  366. Taskar B, Chatalbashev V, Koller D, Guestrin C (2005) Learning structured prediction models: a large margin approach. In: *Proc intl conf on machine learning (ICML)*
  367. Tenenbaum JB, deSilva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500)
  368. Torgerson WS (1952) Multidimensional scaling: I. Theory and method. *Psychometrika* 17(4)
  369. Torralba A, Murphy KP, Freeman WT, Rubin MA (2003) Context-based vision system for place and object recognition. In: *Proc IEEE intl conf on computer vision (ICCV)*, Nice, France, October 2003, vol 2
  370. Torralba A, Murphy KP, Freeman WT (2004) Sharing features: efficient boosting procedures for multiclass object detection. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*, vol 2
  371. Torralba A, Murphy KP, Freeman WT (2007) Sharing visual features for multiclass and multiview object detection. *IEEE Trans Pattern Anal Mach Intell* 19(5)
  372. Trojanowski JQ, Vandeersticbele H, Korecka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P, Dean R, Siemers E, Potter WZ, Weiner MW, Jack CR Jr., Jagust W, Toga AW, Lee VM-Y, Shaw LM (2010) Update on the biomarker core of the Alzheimer's disease neuroimaging initiative subjects. *Alzheimer's Dement* 6(3)
  373. Tsai D, Flagg M, Rehg JM (2010) Motion coherent tracking with multi-label MRF optimization. In: *Proc British machine vision conference (BMVC)*
  374. Tu Z (2005) Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. In: *Proc IEEE intl conf on computer vision (ICCV)*, Beijing, China, October 2005, vol 2
  375. Tu Z, Bai X (2010) Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *IEEE Trans Pattern Anal Mach Intell* 32(10)

376. Tuytelaars T, Schmid C (2007) Vector quantizing feature space with a regular lattice. In: Proc IEEE intl conf on computer vision (ICCV)
377. UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml/datasets.html>
378. Urtasun R, Darrell T (2008) Local probabilistic regression for activity-independent human pose inference. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
379. Vandermeeren M, Mercken M, Vanmechelen E, Six J, Vandevoorde A, Martin JJ, Cras P (1993) Detection of tau proteins in normal and Alzheimer's disease cerebrospinal fluid with a sensitive sandwich enzyme-linked immunosorbent assay. *J Neurochem* 61(5)
380. Vapnik V (2000) *The nature of statistical learning theory*. Springer, Berlin
381. Varma M, Zisserman A (2005) A statistical approach to texture classification from single images. *Int J Comput Vis* 62(1–2)
382. Vazquez-Reina A, Avidan S, Pfister H, Miller E (2010) Multiple hypothesis video segmentation from superpixel flows. In: Proc European conf on computer vision (ECCV). Springer, Berlin
383. Vedaldi A, Blaschko M, Zisserman A (2011) Learning equivariant structured output SVM regressors. In: Proc IEEE intl conf on computer vision (ICCV)
384. Verbeek J, Triggs B (2007) Region classification with Markov field aspect models. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
385. Verma R, Zacharaki EI, Ou Y, Cai H, Chawla S, Lee A-K, Melhem ER, Wolf R, Davatzikos C (2008) Multi-parametric tissue characterization of brain neoplasm and their recurrence using pattern classification of MR images. *Acad Radiol* 15(8)
386. Vezhnevets A, Ferrari V, Buhmann JM (2012) Weakly supervised structured output learning for semantic segmentation. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
387. Villamizar M, Moreno-Noguer F, Andrade-Cetto J, Sanfeliu A (2010) Efficient rotation invariant object detection using boosted random ferns. In: Proc IEEE conf computer vision and pattern recognition (CVPR)
388. Viola P, Jones MJ (2001) Rapid object detection using a boosted cascade of simple features. In: Proc IEEE conf computer vision and pattern recognition (CVPR), December 2001, vol 1
389. Viola P, Jones MJ (2004) Robust real-time face detection. *Int J Comput Vis* 57(2)
390. Viola P, Jones MJ, Snow D (2003) Detecting pedestrians using patterns of motion and appearance. In: Proc IEEE intl conf on computer vision (ICCV)
391. Vishwanathan SVN, Schraudolph NN, Schmidt MW, Murphy KP (2006) Accelerated training of conditional random fields with stochastic gradient methods. In: Proc intl conf on machine learning (ICML)
392. Vitter JS (1985) Random sampling with a reservoir. *ACM Trans Math Softw* 11(1)
393. Wainwright MJ, Jordan MI (2008) Graphical models, exponential families, and variational inference. *Found Trends Mach Learn* 1(1–2)
394. Walhovd KB, Fjell AM, Dale AM, McEvoy LK, Brewer J, Karow DS, Salmon DP, Fennema-Notestine C (2010) Multi-modal imaging predicts memory performance in normal aging and cognitive decline. *Neurobiol Aging* 31(7)
395. Wang J (2007) *On transductive support vector machines*. In: Prediction and discovery. American Mathematical Society, Providence
396. Wang RY, Popović J (2009) Real-time hand-tracking with a color glove. In: Proc ACM SIGGRAPH
397. Wang C, Gorce M, Paragios N (2009) Segmentation, ordering and multi-object tracking using graphical models. In: Proc IEEE intl conf on computer vision (ICCV)
398. Wang Y, Fan Y, Bhatt P, Davatzikos C (2010) High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables. *NeuroImage*
399. Wels M, Carneiro G, Aplas A, Huber M, Comaniciu D, Hornegger J (2008) A discriminative model-constrained graph-cuts approach to fully automated pediatric brain tumor segmentation in 3D MRI. In: Proc medical image computing and computer assisted intervention (MICCAI)

400. Wen PY, Macdonald DR, Reardon DA, Cloughesy TF, Sorensen AG, Galanis E, Degroot J, Wick W, Gilbert MR, Lassman AB, Tsien C, Mikkelsen T, Wong ET, Chamberlain MC, Stupp R, Lamborn KR, Vogelbaum MA, van den Bent MJ, Chang SM (2010) Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *Am J Neuroradiol*
401. Williams B, Klein G, Reid I (2007) Real-time SLAM relocalisation. In: *Proc IEEE intl conf on computer vision (ICCV)*
402. Winder S, Brown M (2007) Learning local image descriptors. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
403. Winn J, Shotton J (2006) The layout consistent random field for recognizing and segmenting partially occluded objects. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
404. Winn J, Criminisi A, Minka T (2005) Categorization by learned universal visual dictionary. In: *Proc IEEE intl conf on computer vision (ICCV)*, Beijing, China, October 2005, vol 2
405. Xiong C, Johnson D, Xu R, Corso JJ (2012) Random forests for metric learning with implicit pairwise position dependence. In: *Proc of ACM SIGKDD intl conf on knowledge discovery and data mining*
406. Yakushev I, Hammers A, Fellgiebel A, Schmidtman I, Scheurich A, Buchholz HG, Peters J, Bartenstein P, Lieb K, Schreckenberger M (2009) SPM-based count normalization provides excellent discrimination of mild Alzheimer's disease and amnesic mild cognitive impairment from healthy aging. *NeuroImage* 44(1)
407. Yan R, Yang J, Hauptmann A (2003) Automatically labeling video data using multi-class active learning. In: *Proc IEEE intl conf on computer vision (ICCV)*
408. Yao C, Wada T, Shimizu A, Kobatake H, Nawano S (2006) Simultaneous location detection of multi-organ by atlas-guided eigen-organ method in volumetric medical images. *Int J Comput Assisted Radiol Surg* 1
409. Yao A, Gall J, van Gool L (2010) A Hough transform-based voting framework for action recognition. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
410. Yao B, Khosla K, Fei-Fei L (2011) Combining randomization and discrimination for fine-grained image categorization. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*, Springs, USA, June 2011
411. Yi Z, Criminisi A, Shotton J, Blake A (2009) Discriminative, semantic segmentation of brain tissue in MR images. In: *Proc medical image computing and computer assisted intervention (MICCAI)*. Springer, Berlin
412. Yin Z, Collins R (2009) Shape constrained figure-ground segmentation and tracking. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
413. Yin P, Criminisi A, Winn J, Essa I (2007) Tree based classifiers for bilayer video segmentation. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
414. Zhan Y, Zhou X-S, Peng Z, Krishnan A (2008) Active scheduling of organ detection and segmentation in whole-body medical images. In: *Proc medical image computing and computer assisted intervention (MICCAI)*
415. Zhang Q, Souvenir R, Pless R (2006) On manifold structure of cardiac MRI data: application to segmentation. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*, Los Alamitos, CA, USA
416. Zhang J, Marszałek M, Lazebnik S, Schmid C (2007) Local features and kernels for classification of texture and object categories: a comprehensive study. *Int J Comput Vis* 73(2)
417. Zhang D, Wang Y, Zhou L, Yuan H, Shen D, The Alzheimer's Disease Neuroimaging Initiative (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage* 55(3)
418. Zheng F, Webb GI (2005) A comparative study of semi-naïve Bayes methods in classification learning. In: *Australasian data mining conference*
419. Zheng Y, Georgescu B, Comaniciu D (2009) Marginal space learning for efficient detection of 2D/3D anatomical structures in medical images. In: *Proc information processing in medical imaging (IPMI)*. Springer, Berlin

420. Zhou SK, Comaniciu D (2010) Shape regression machine and efficient segmentation of left ventricle endocardium from 2D B-mode echocardiogram. *Med Image Anal*
421. Zhou SK, Georgescu B, Zhou X, Comaniciu D (2005) Image-based regression using boosting method. In: *Proc IEEE intl conf on computer vision (ICCV)*
422. Zhou SK, Zhou J, Comaniciu D (2007) A boosting regression approach to medical anatomy detection. In: *Proc IEEE conf computer vision and pattern recognition (CVPR)*
423. Zhu Y, Fujimura K (2007) Constrained optimization for human pose estimation from depth sequences. In: *Proc Asian conf on computer vision (ACCV)*
424. Zhu X, Ghahramani Z (2002) Learning from labeled and unlabeled data with label propagation. Technical Report CMU-CALD-02-107, Carnegie Mellon University
425. Zhu X, Goldberg A (2009) Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning*. Morgan and Claypool Publishers, San Rafael
426. Zhu C, Byrd RH, Lu P, Nocedal J (1997) Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans Math Softw* 23(4)
427. Zien A, Ong CS (2007) Multiclass multiple kernel learning. In: *Proc intl conf on machine learning (ICML)*
428. Zikic D, Glocker B, Konukoglu E, Criminisi A, Demiralp C, Shotton J, Thomas OM, Das T, Jena R, Price SJ (2012) Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR. In: *Proc medical image computing and computer assisted intervention (MICCAI)*