

Computer-Aided Detection of Polyps in CT Colonography Using Logistic Regression

Vincent F. van Ravesteijn*, Cees van Wijk, Frans M. Vos, Roel Truyen, Joost F. Peters, Jaap Stoker, and Lucas J. van Vliet, *Member, IEEE*

Abstract—We present a computer-aided detection (CAD) system for computed tomography colonography that orders the polyps according to clinical relevance. The CAD system consists of two steps: candidate detection and supervised classification. The characteristics of the detection step lead to specific choices for the classification system. The candidates are ordered by a linear logistic classifier (logistic regression) based on only three features: the protrusion of the colon wall, the mean internal intensity, and a feature to discard detections on the rectal enema tube. This classifier can cope with a small number of polyps available for training, a large imbalance between polyps and non-polyp candidates, a truncated feature space, unbalanced and unknown misclassification costs, and an exponential distribution with respect to candidate size in feature space. Our CAD system was evaluated with data sets from four different medical centers. For polyps larger than or equal to 6 mm we achieved sensitivities of respectively 95%, 85%, 85%, and 100% with 5, 4, 5, and 6 false positives per scan over 86, 48, 141, and 32 patients. A cross-center evaluation in which the system is trained and tested with data from different sources showed that the trained CAD system generalizes to data from different medical centers and with different patient preparations. This is essential to application in large-scale screening for colorectal polyps.

Index Terms—Computed tomography (CT) colonography, computer aided diagnosis, logistic regression, pattern recognition, polyp detection.

I. INTRODUCTION

COLORECTAL cancer is the second leading cause of mortality due to cancer in the western world [1]. Paradoxically, perhaps, is that it is preventable to a large part or at least curable, if detected early. Adenomatous colorectal polyps are considered important precursors to colon cancer [2]–[4]. It has been shown that screening for such polyps can significantly

reduce the incidence of colon cancer [5], [6]. Computed tomography (CT) colonography (CTC) is a rapidly evolving technique for screening, but the interpretation of the data sets is still time-consuming. Computer-aided detection (CAD) of polyps may enhance the efficiency and also increase the sensitivity. This is specifically important for large-scale screening. Recent studies show that the sensitivity of CAD systems is already comparable to the sensitivity of optical colonoscopy [7]–[9] and radiologists using CTC [10].

The best indicator of the risk that a polyp is malignant or turns malignant over time is size [11]. The consensus [12] is that patients with a polyp of at least 10 mm must be referred to optical colonoscopy for polypectomy and it is advised that diminutive polyps (≤ 5 mm) should not even be reported [13], [14]. There is still debate over the need for polypectomy for 6–9 mm polyps. Surveillance for growth with CT colonography has also been suggested.

A. Related Work

CAD algorithms for polyp detection in CT colonography usually consist of candidate detection followed by supervised classification. Candidate detection aims at 100% sensitivity for polyps larger than 6 mm which goes at the expense of hundreds of false positives (FPs) per scan. The task of supervised classification is to reduce the number of detections to about a handful without sacrificing the sensitivity too much.

For the detection of polyp candidates, Summers *et al.* [19], [20] proposed to use methods from differential geometry in which the principal curvatures were computed by fitting a fourth-order B-spline to local neighborhoods with a 5 mm radius. Candidates were generated by selecting regions of elliptic curvature with a positive mean curvature [19]. Yoshida *et al.* [21], [22] used the shape index and curvedness to find candidate objects on the colon wall. The shape index and curvedness are functions of the principal curvatures of the surface, which were computed in a Gaussian-shaped window (aperture). Alternatively, Kiss *et al.* [23] generated candidates by searching for convex regions on the colon wall. Their method fitted a sphere to the surface normal field. The type of material in which the center of the fitted sphere was found (in tissue or in air) determined the classification of the surface as either convex or concave. As a result, roughly 90% of the colon wall was labeled as concave, that is “normal.” Subsequently, a generalized Hough transformation using a spherical model was applied to the convex surface regions. Candidate objects were generated by searching for local maxima in the parameter space of the Hough transformation. Kiss *et al.* characterized the

Manuscript received May 26, 2009; revised July 13, 2009; accepted July 13, 2009. First published August 07, 2009; current version published January 04, 2010. Asterisk indicates corresponding author.

*V. F. van Ravesteijn and C. van Wijk are with the Quantitative Imaging Group, Delft University of Technology, NL-2628 CJ Delft, The Netherlands (e-mail: v.f.vanravesteijn@tudelft.nl).

R. Truyen and J. F. Peters are with Philips Healthcare, Healthcare Informatics, NL-5684 PC Best, The Netherlands (e-mail: roel.truyen@philips.com).

F. M. Vos is with the Quantitative Imaging Group, Delft University of Technology, NL-2628 CJ Delft, The Netherlands and also with the Department of Radiology, Academic Medical Center, NL-1100 DD Amsterdam, The Netherlands (e-mail: f.m.vos@tudelft.nl).

J. Stoker is with the Department of Radiology, Academic Medical Center, NL-1100 DD Amsterdam, The Netherlands (e-mail: j.stoker@amc.uva.nl).

L. J. van Vliet is with the Quantitative Imaging Group, Delft University of Technology, NL-2628 CJ Delft, The Netherlands (e-mail: l.j.vanvliet@tudelft.nl).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2009.2028576

candidate's shape by comparing the spherical harmonics with those of the polypoid models in a database [24].

Apart from the different candidate detection algorithms, there is a wide variety in the design of the pattern recognition system, ranging from low-complex systems like linear discriminant classifiers to classification systems using multiple neural networks. Yoshida and Näppi used linear and quadratic discriminant classifiers [21], [22], [25] as well as Jerebko *et al.* [26]. Wang *et al.* [27] used a two-level classifier with a further unspecified linear discriminant classifier in the second level. The first level of this classifier consisted of a normalization procedure, which was specially designed and had four parameters. Sundaram *et al.* [28] classified the candidates based on a single heuristically designed score using curvature information of the candidate patches. Göktürk *et al.* [29] employed a support vector machine for classification, in which it was assumed that after a transformation by the kernel function, the data were linearly separable. This implicitly required minimal mixing between polyps and false detections. Jerebko *et al.* [30] and Zheng *et al.* [31] used a committee of support vector machines. Neural networks were also used by Jerebko *et al.* [30] and Näppi *et al.* [8], [32] for classification, and by Suzuki *et al.* [33] for the reduction of false detections on the rectal enema tube.

To conclude, many different proposals for a classification system for CAD of polyps have been presented. However, the motivation for a specific design of the classification system is often unclear. Moreover, proper comparison between classification systems is difficult due to the different candidate detection systems and feature extraction methods. One may reason that the optimization of complex classification systems (with large number of parameters or features) may be complicated by the limited availability of training examples. This could lead to overtraining to a specific patient population or patient preparation.

A steadily growing number of papers (e.g., [7], [21], [23]–[27], [29], and [34]–[37]) reported on the performance of polyp detection algorithms (see Yoshida and Näppi [10] for a review on CAD systems for CTC). However, the results can not easily be compared due to large differences in the data sets used for evaluation (see also Section II-A).

B. Objective

Candidate detection typically renders a lot of candidates to sustain maximum sensitivity. Hence, the number of objects from the target class (polyps) is relatively low. This large imbalance of the prevailing classes typically hampers classifier design and training. A further complication is that the misclassification costs for objects from the two classes are unknown and certainly very different. This paper discusses the consequences of these characteristics for the design of the classification system.

We aim to design a novel, low-complex, classification system that orders the polyps according to clinical relevance. It implicitly takes into account that the misclassification costs of polyps increase with lesion size. In other words, larger polyps are more important than smaller ones and the problem is not considered as a mere two-class classification task, but rather as a regression

problem. With this in mind, we distinguish two types of features in the design of the classification system. First, there are features that facilitate an ordering of the candidates. These are the features that directly relate to the lesion size. Second, there are features which will be shown to render a Gaussian distribution. In order to keep the classifier simple and to prevent the use of complex combination strategies, these features are mapped into features of the first type by a Mahalanobis distance (MD) mapping. This strategy is used to discard outliers and mimics the use of a Gaussian one-class classifier [38]. It will be shown that this two-level classification system is effective over data from various sources.

The technical novelty of our paper is to approach the classification task as a regression problem. Such a strategy requires that features are ordered according to relevance. A mechanism is introduced to map features that are not ordered as such into features that do have the ordering property. It will be demonstrated that the Mahalanobis distance to the target class mean is appropriate for the current problem. Imposing the ordering may be achieved for any other problem provided that the distance to the most typical representation of the target class can be defined.

II. DATA DESCRIPTION AND FEATURE DESIGN

A CAD system for CTC starts with the acquisition of CT colonography data. In these data, candidate objects are detected and segmented. The segmented candidates are typically characterized by features describing, for instance, the candidate's shape and its internal intensity distribution. Such data serve as input for the classification system. All preprocessing steps will be addressed in this section.

A. CT Colonography Data

Data sets from four different medical centers were used to evaluate the performance of our system. Data sets from different sources differ in polyp prevalence, the patient preparation, the scanning protocol, the protocol for determining the ground truth, and the type of rectal tube used for colon distension during CT examination. An arbitrary number of patients were randomly selected from each source, irrespective of the number of polyps and their shape. The most important characteristics of the data sets are shown in Table I. More details may be retrieved from the references included in the table. All patients adhered to an extensive laxative regime. The reference standard (ground truth) for data sets "A," "B," and "C" was optical colonoscopy. An expert radiologist served as the reference for data set "D." Radiologists retrospectively indicated the location of polyps by annotating a point in the 3-D data set based on the reference standard. The candidate segmentations (see below) were labeled by comparison to these annotations. Data sets "A," "B," and "C" consisted of scans in both prone and supine positions. A polyp was counted as a true positive CAD detection if it was found in at least one of the two scanned positions. Only dataset "A" has been used during development of the system.

B. Candidate Detection

Polyps are often described as objects that protrude from the colon wall. For that reason, the candidate detection method is

TABLE I
PROPERTIES OF THE DATA SETS

Data set	Medical Center	Slice Thickness (mm)	Fecal Tagging	Scans per Patient	Number of Patients	Number of Polyps ≥ 6 mm	Ref.
'A'	AMC / Amsterdam	3.2	No	2	86	59	[15]
'B'	WRAMC / Washington, DC	1.2	Yes	2	48	28	[16]
'C'	UW / Madison, WI	1.2	Yes	2	141	176	[17] ¹
'D'	Charité / Berlin	1.0	Yes	1	32	8	[18] ¹

¹Information about the patient preparation can be retrieved from the reference. However, the specific data set we used is not described.

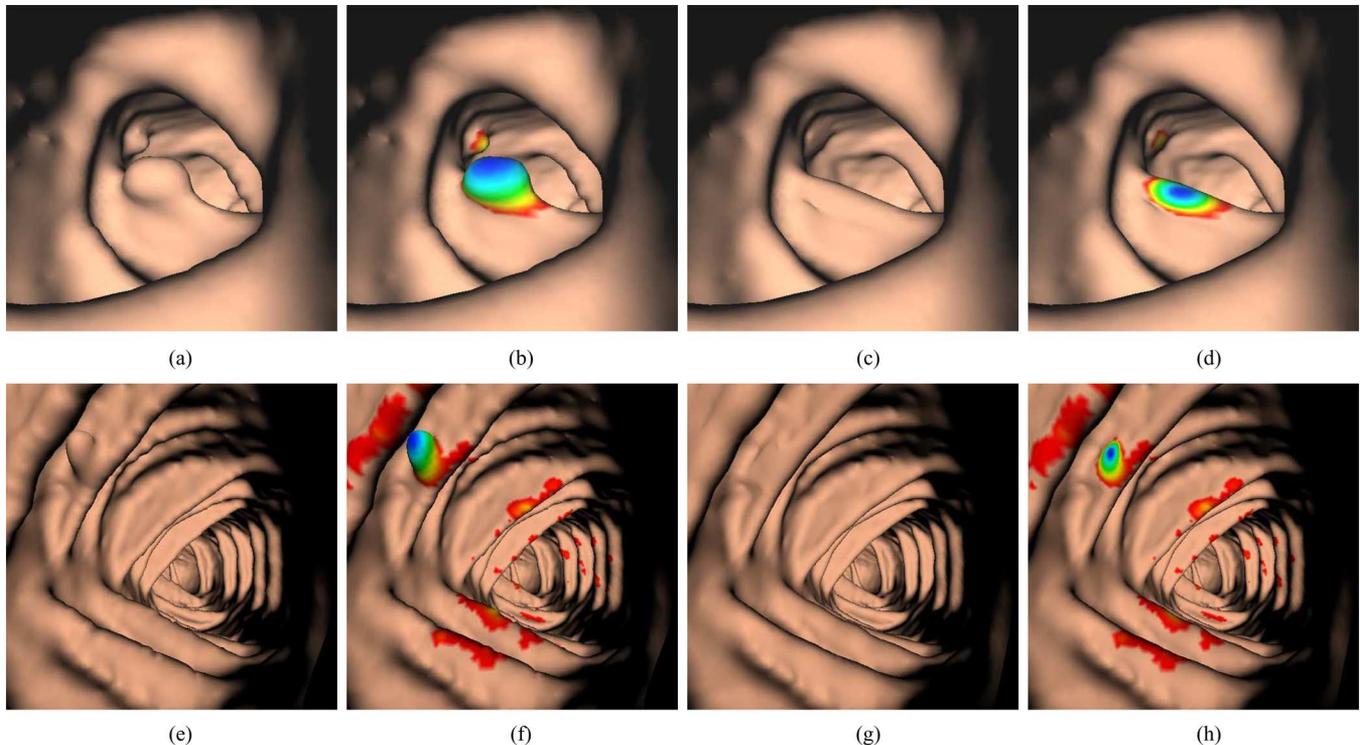


Fig. 1. Candidate detection method applies a nonlinear “flattening” operation to the colon wall. The protrusion field is defined as the difference in position of the colon wall before (a)–(b), (e)–(f) and after (c)–(d), (g)–(h) application of the operation. The coloring (b,d,f,h) indicates the protrusion of the mesh vertices of detected candidates (blue denotes a large protrusion and red denotes a protrusion of 0.2 mm, i.e., the low hysteresis threshold). Notice that the folds are hardly affected by the operation (a) before deformation, (b) before deformation, (c) after deformation, (d) after deformation, (e) before deformation, (f) before deformation, (g) after deformation, and (h) after deformation.

designed to detect all objects that protrude from the colon wall, irrespective of their shape. Suppose that the points on the convex parts of a protruding object are iteratively moved inwards. Effectively, this will “remove” the object. After a certain amount of deformation, the protrusion is completely removed and the colon wall appears “normal.” The amount of deformation as a result of the operation is a measure of “protrudedness.” Fig. 1 illustrates this process by showing images before and after application of the nonlinear “flattening” operation.

Practically, the colon wall was represented by a triangle mesh, which was obtained by thresholding the CT colonography data at -750 Hounsfield units (HU). A nonlinear PDE [35] was solved to remove all protruding structures from the mesh that displayed a positive second principal curvature. A similar approach that acts directly on the grey valued image is presented in [39]. In this procedure, the global shape of the colon including the folds was retained, since these structures

display a second principal curvature that is smaller than or equal to zero. The protrusion field was computed by the position difference of the mesh vertices before and after processing. Subsequently, hysteresis thresholding was applied to this field to detect and segment the candidates. The high threshold on the protrusion was 0.4 mm and determines the sensitivity. The value of 0.4 mm was selected since it yields 100% sensitivity per polyp annotation in our training set. All retained regions of the colon surface were augmented by adding the adjacent mesh points with a protrusion of at least 0.2 mm (the low threshold). The regions thus obtained form the segmented candidates.

C. Features

Radiologists that evaluate CTC data primarily use two properties of a candidate for classification: the shape and the voxel intensities inside the candidate. There is still debate about the optimal way to analyze CTC data. Radiologists using the 3-D

rendering of the colon (virtual colonoscopy) detect polyps based on shape, but they will often fall back to the 2-D representation (grey values) before a final decision is made. Using the 2-D representation, both the internal intensities and the shape are assessed, although shape is often hard to extract from the grey-value images. The features used in the presented CAD system are based on the same two properties that are primarily used by radiologists.

Shape was previously described by the shape index and curvedness [22], mean curvature, average principal curvatures and sphericity ratio [19], [20], and spherical harmonics [24]. An alternative method to measure shape, which is based on the protrusion field, will be introduced (see Section II-C1).

The internal intensity of the candidates has been found before to be a discriminative feature to discard a large number of false detections [25]–[27], [34]. It may be expected that due to the partial volume effect false detections arise that have low internal intensity. False detections that are stool often have air inside, which also lowers the intensity. Such information about the candidates will be included through statistics on the object's internal voxel intensities (see Section II-C2).

At last, it was experimentally found that many false positives turned out to be detections on the rectal enema tube (RET) (previously also reported in [33] and [40]). Therefore, a third feature will be proposed to discard such false detections (see Section II-C3).

1) *Shape Feature From Protrusion Field*: Polyps are conventionally characterized by the single largest diameter, excluding the stalk [11], [41]. However, Fig. 2(a) shows that this measure does not distinguish polyps from false detections well. It appears that especially among the less protruding candidates (≤ 2 mm), the candidates with the larger diameters are predominantly false detections. Alternatively, it might be natural to select the maximum protrusion of a candidate as a feature, but it appears that a lot of polyps have only modest protrusion. As an illustration, Fig. 2(c) and (d) shows two candidates that have approximately the same maximum protrusion but a completely different appearance. The first candidate (candidate “c”) has a large diameter, but does not resemble a polyp at all, whereas the second candidate (candidate “d”) with a small diameter does so. To conclude, a large diameter relative to the maximum protrusion indicates a nonpolypoidal shape (candidate “c”) and a small diameter or a relative low protrusion points to a small clinically irrelevant candidate. A feature that is derived from the thresholded protrusion field should therefore include the size of a candidate as well as the ratio between the largest diameter and the maximum protrusion. Moreover, the feature should characterize the whole segmented area instead of the extrema (like the largest diameter or the maximum protrusion).

We designed a feature that takes into account both the protrusion as well as the lateral size of the object. Effectively, it measures the percentage of the area of the candidate that has a protrusion larger than a certain threshold T . This feature is further denoted as Φ_T . A large circumference as well as shallow edges lead to relatively large areas with protrusion below T and result in a low response. Thus, this feature favors compact objects with steep edges. Fig. 2(b) shows that according to Φ_T ($T = 0.6$ mm) candidate “d” is indeed favored over candidate

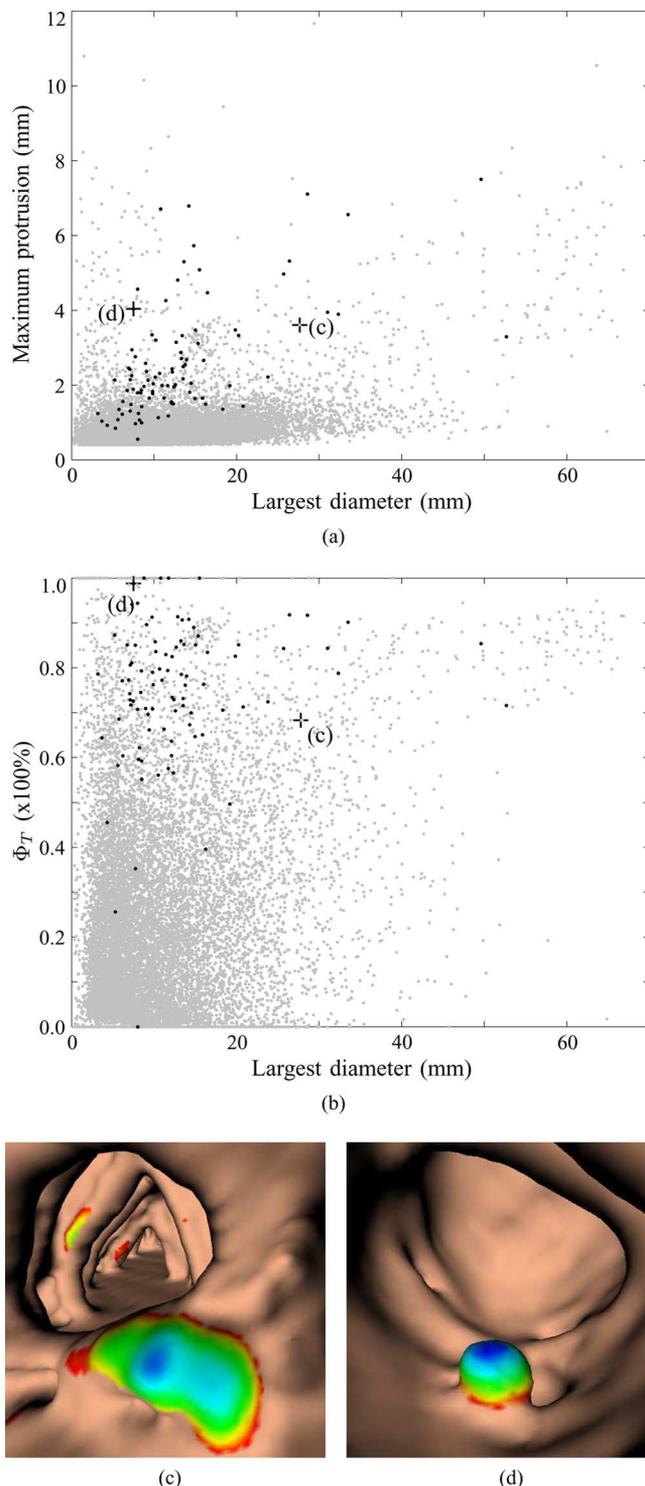


Fig. 2. (a)–(b) Scatter plots of features calculated for data set “A.” Grey dots denote false detections and black dots indicate polyps ≥ 6 mm. Note that each polyp may appear as two separate dots in the scatter plot, since each patient is scanned twice. (a) The maximum protrusion versus the single largest diameter of a candidate. The threshold of the candidate detection can be seen at a maximum protrusion of 0.4 mm. (b) Φ_T ($T = 0.6$ mm) versus the largest diameter. (c)–(d) Two candidates with the same maximum protrusion that are ordered differently according to Φ_T .

“c.” Ordering the candidates based on Φ_T is thus expected to improve the performance of the CAD system over simply using the maximum diameter alone.

2) *Intensity Features*: Consider all mesh vertices that are part of the segmentation mask of a candidate object (see Section II-B). For each vertex, a weighted average of colon wall intensities was calculated along the line segment from the vertex under consideration to the center of mass of the candidate's vertices. The weight of the intensity of each voxel depends on the Gaussian scaled squared-distance between the intensity and the maximum intensity along the line segment. The tonal scale σ_t used for weighting was set to 140 HU. This value is substantially larger than two times the image noise (previously measured to be 43.4 HU for data acquired with 50 mAs [42]). Consequently, σ_t facilitated that the edges of the candidate contributed less to the weighted average than the internal voxels of the candidate. In other words, the candidate's true internal intensity was emphasized. The center of mass falling inside the polyp is supported by the smooth apex of polyps.

Subsequently, the mean ($f_{I,\text{mean}}$), median ($f_{I,\text{median}}$), maximum ($f_{I,\text{max}}$), minimum ($f_{I,\text{min}}$), and standard deviation ($f_{I,\text{std}}$) were determined from the weighted averages of all vertices. The latter four were only used in the classifier selection stage (see Section V-A).

3) *Feature for Suppressing Candidates on the Rectal Enema Tube*: The rectal enema tube is a prominent source of false positive classifications [33], [40]. This is because the tube's attenuation in CT is similar to that of tissue. Moreover, the size and shape (25 mm in diameter) resembles a large polyp. Cross-sectional examples of a rectal enema tube are shown in Fig. 3(a). To suppress the false detections on the rectal tubes, a feature has been developed to distinguish these false detections from the other candidates. For each candidate it was measured how much field-of-view (FOV) the candidate "blocks" as seen from the rectal enema tube [Fig. 3(b)]

$$f_{\text{FOV}} = \frac{1}{4\pi} \sum_{\text{points} \in \text{candidate}} A_{1\text{-ring}} \frac{(\vec{q}_i \cdot \vec{n}_i)}{\|\vec{q}_i\|^3} \quad (1)$$

in which \vec{q}_i is the vector from a mesh point i of the candidate to an arbitrary point on the rectal tube, \vec{n}_i is the vertex normal, and $A_{1\text{-ring}}$ is the surface area of the one-ring neighborhood defined as the average area of the cells adjacent to the point of interest. A positive value means that the candidate is bent away from the tube and a negative value indicates that the candidate is bent toward the tube.

Fig. 3(c) shows a scatter plot of false detections (grey) and true polyps (black) with f_{FOV} on the horizontal axis and with the mean radius of the candidates on the vertical axis. The mean radius is calculated as a weighted sum of the distances of all mesh points i to the center of gravity of the candidate, $\|\vec{r}_i\|$, weighted by the area of the one-ring neighborhood $A_{1\text{-ring},i}$. Apparently, four clusters are identifiable in this feature space: candidates at the end of the tube have negative values for f_{FOV} and a rather small mean radius (dotted line); candidates on the balloon also yield negative f_{FOV} , but come with a large mean radius (dashed line); candidates inside the tube have positive response for f_{FOV} (dash-dotted); and candidates that are not related to the tube have negligible blocking and form an elongated

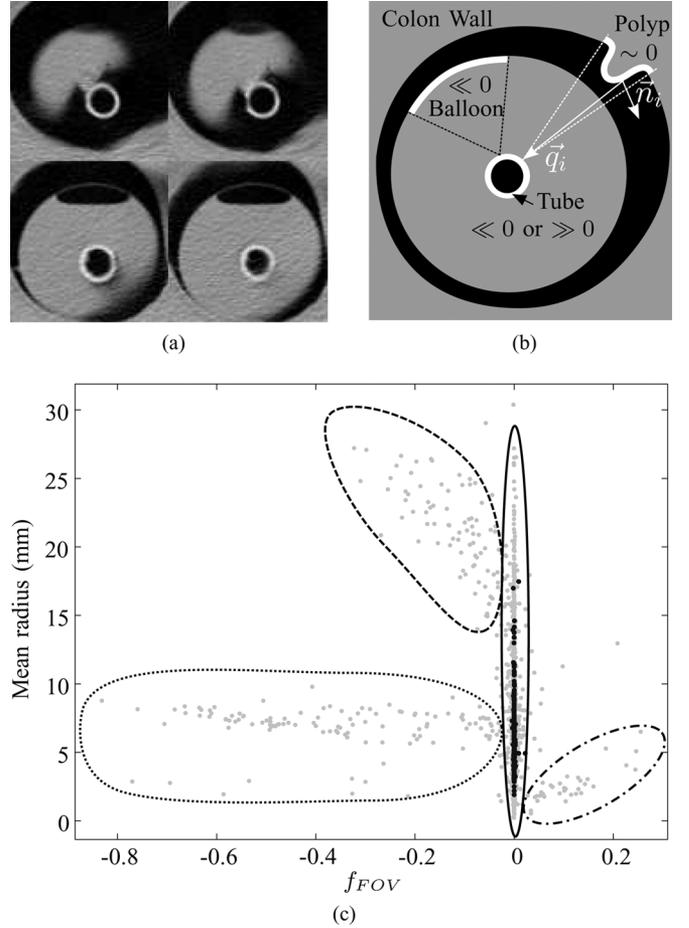


Fig. 3. (a) Example of a rectal enema tube in data set "A" as seen in different slices of a CT image. (b) Schematical explanation of the responses of f_{FOV} . (c) Scatter plot of the mean radius versus f_{FOV} . The grey dots are false detections and the black dots are polyps. In the text, we identify the four clusters.

cluster centered at $f_{\text{FOV}} = 0$ (solid line). To conclude, non-zero values of this feature tend to indicate detections on the rectal enema tube.

III. CHARACTERISTICS OF THE FEATURE SPACE

A first prerequisite for clinical application is that the system has high sensitivity for the detection of polyps. To limit the risk of missing a polyp in the candidate detection step, this step unavoidably yields a large number of detections. Consequently, the number of objects from the two classes is severely unbalanced. For instance, only 0.3% of the candidates detected in data set "A" were polyps ≥ 6 mm. Any classifier relies heavily on the few polyp examples. Complex classifiers may not be expected to generalize well to other data sets, because they are typically sensitive to small changes in training data. Furthermore, the misclassification costs for objects from the two classes are unbalanced and unknown: a missed polyp is far more troublesome than a false positive classification. Finally, it has to be realized that the size of a polyp indicates the risk of it becoming malignant.

A part of the feature space is presented in Fig. 4(a) and (b) by two scatter plots. It can be seen that the distribution of the polyps is rather uniform with respect to Φ_T , though it appears

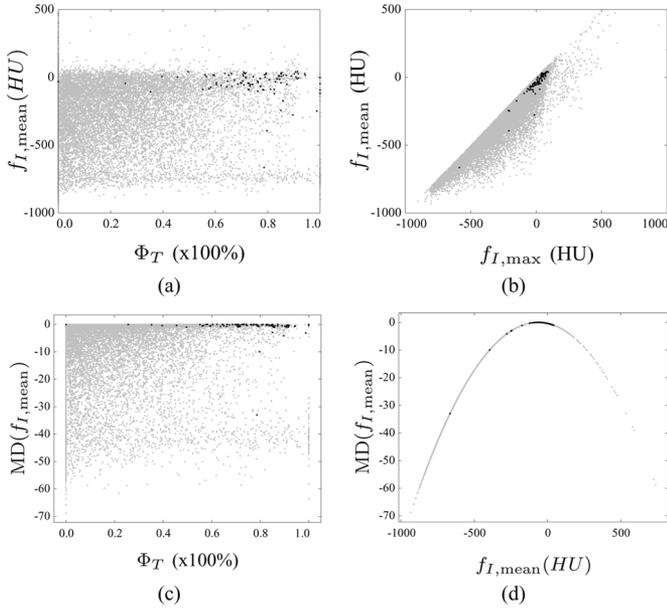


Fig. 4. Scatter plots demonstrating the distribution of the candidates for data set “A.” The grey dots are false detections and the black dots are polyps. (a) Mean intensity versus Φ_T . (b) Mean intensity versus maximum intensity. (c) Same feature space as (a) with the output of the negated Mahalanobis distance mapping on the vertical axis. This mapping is introduced in Section IV.A. (d) Influence of the mapping on $f_{I,\text{mean}}$. Note that candidates with a high and low mean intensity have a lower mapped feature than the polyps.

truncated at a certain level ($\Phi_T \approx 55\%$). This occurs because polyps < 6 mm are not clinically relevant and were therefore excluded *a priori* (i.e., not annotated in the data). The false detections display a different behavior. As our focus is on irregularities on the colon surface (protruding objects), it may be expected that far more candidates with small protrusion are detected than candidates with large protrusion, e.g., due to natural fluctuations of the colon wall and noise. This can also be seen in the distribution of the candidates with respect to the maximum protrusion in Fig. 5(a) and with respect to Φ_T in Fig. 5(b) (dotted curves). An exponential decaying function fitted to the distribution is also shown (solid curves). Thus, one must not only reckon with many false detections, the false detections are also unevenly distributed in the feature space. Finally, it can be observed that the classes largely overlap and that the way the candidates were generated imposes abrupt cluster boundaries, which may hamper density based classifiers. The abrupt cluster boundaries can be seen at $\Phi_T = 0\%$ and $\Phi_T = 100\%$ in Fig. 4(a).

We approach the classification problem not just as a two-class classification task, but rather as a regression problem. In other words, the classification system should be designed to facilitate a clinically relevant ordering of the candidates. Ideally, this means that the polyps should be ranked above the false detections and that the larger polyps are ranked above the smaller polyps. The classifier that is used in the regression analysis should be robust to the large class imbalance, the uneven distribution of candidates in the feature space, and the abrupt boundaries in the feature space. Moreover, the classification system as a whole must be low-complex in order to be robust to variations in the data sets from different sources.

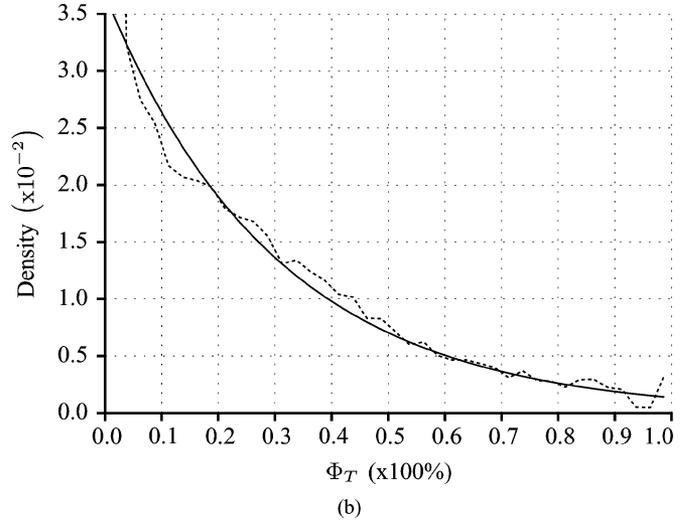
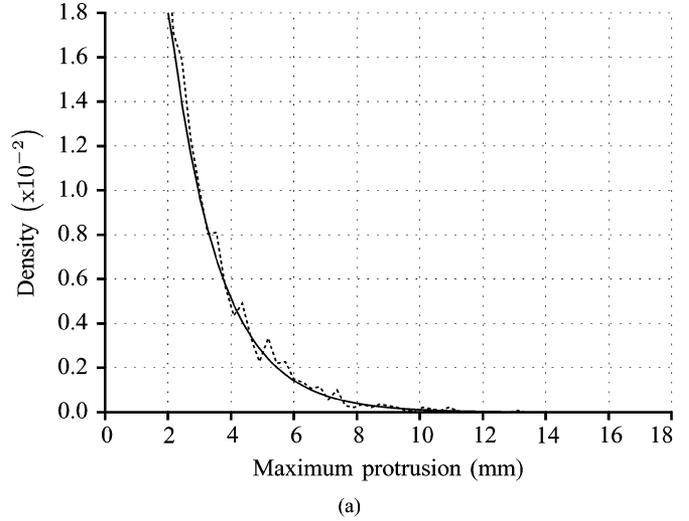


Fig. 5. Distribution of (a) the maximum protrusion and (b) Φ_T of the false detections in data set “A” (dotted curves). Exponential decaying functions were fitted to the distributions (solid curves).

IV. CLASSIFICATION SYSTEM

This section describes a classification system that fulfills the demands derived in the previous section. It is schematically depicted in Fig. 6. The input feature vector consists of two types of features, namely those suitable for ordering the candidates (f_O) and those allowing for density estimation and outlier rejection (f_D). The features of the first type are directly used in the regression analysis, whereas the other features are mapped first by a Mahalanobis distance mapping. Subsequently, regression analysis leads to an ordering. The ordering can then be used to compute FROC curves to estimate the performance. Three discriminant classifiers will be applied in the regression problem (see Section V): the normal-based linear discriminant classifier (LDC) [43], the normal-based quadratic discriminant classifier (QDC) [43], and the logistic discriminant classifier [43].

We did not opt for support vector machine (SVM) classifiers due to the large class overlap. Due to this large overlap, it is not expected that a unique classification boundary can be found confidently. Moreover, we did not opt for neural networks too

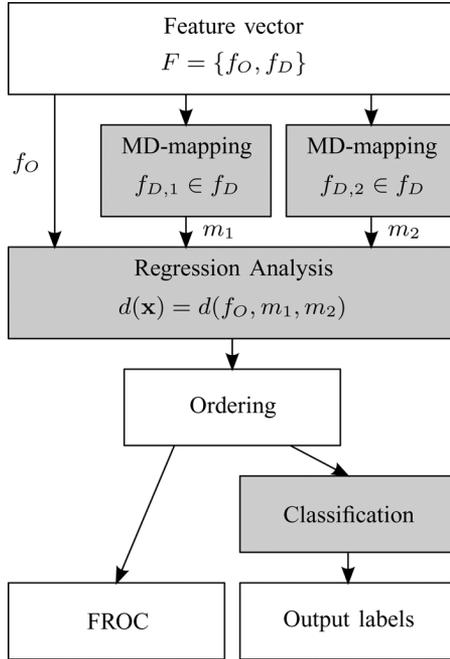


Fig. 6. Schematic representation of the classification system. The classification starts with a feature vector consisting of features suitable for ordering (f_O) and features suitable for density estimation (f_D). The feature sets $f_{D,1}$ and $f_{D,2}$ are processed through two mappings. An ordering of the candidates is determined by regression that incorporates both the features f_O and the outputs of the mappings, m_1 and m_2 . The ordering may be thresholded for classification in order to construct FROC curves.

because, obviously, multi-layer neural networks based solutions may increase complexity. On the other hand, one can think of low-complex neural networks, like single layer networks with sigmoidal transfer functions (as used in [8], [30], and [32]). However, these are known to be closely related to the logistic classifier.

A. Mahalanobis Distance Mapping

Let us assume that, for a certain subset of features, a Gaussian properly describes the distribution of the objects from the target class, i.e., the polyps. One might say that the mean of this distribution corresponds to a typical representation of a polyp (“the most polyp-like polyp”). Moreover, the Mahalanobis distance to the mean of the polyp class may act as an efficient feature to reject outliers, i.e., objects not belonging to the target class. This procedure compares to the operation of a Gaussian one-class classifier [38].

Instead of comparing this distance to a preset threshold, the (negated) Mahalanobis distance is used as a feature. The mean of the polyp class was derived from the train data set. Consequently, this acts as a mapping transforming one or more features into a single feature. The output feature is suitable for ordering the candidates, since zero Mahalanobis distance (the mean of the Gaussian) is considered most polyp-like. The feature can thus be used in the regression analysis. In practice, the mapping was applied to f_{FOV} and $f_{I,mean}$. Effectively, candidates on the rectal tubes as well as candidates with an abnormal intensity are rejected. Fig. 4 illustrates the influence of the mapping on $f_{I,mean}$.

In comparison to Wang *et al.* [27], our mapping replaces the normalization procedure of their two-level classifier. This allows us to use a standard technique from statistical pattern recognition to determine the parameters of the mapping.

B. Normal-Based Discriminant Classifiers

Let us consider the linear normal-based discriminant classifier (LDC) to represent a common, low-complexity type of classifier. Such an LDC includes a weighted sum of the covariance matrices of both classes, in which the weights are the prior probabilities. In the case of a large class imbalance, however, as in the polyp detection problem, the prior of the minority class is extremely small. As a consequence, the weighted sum is almost identical to the covariance matrix of the majority class and the covariance matrix of the minority class is neglected. In other words, contrary to common preference, the detection of objects from the minority (target) class is largely based on information of the objects from the majority (outlier) class. One might conceive this as the opposite of a one-class classifier, which typically uses information about the target class only.

One might consider a quadratic normal-based discriminant classifier (QDC) instead, since it does not weight the covariance matrices by the prior probabilities. One underlying problem here is that the classes have non-Gaussian distributions. In order to capture a polyp inside the tip of the quadratic decision boundary, simultaneously an exponentially increasing number of false positives are included (see Fig. 5). The more conservative linear decision boundary will make a different error to detect such a polyp, but this error is less pronounced. What is more, the quadratic classifier depends strongly on the covariance matrix of the polyp class. This covariance matrix might be somewhat unstable, however, due to the limited number of polyps.

C. Logistic Discriminant Classifier

It was previously demonstrated that the false detections are distributed in an exponential fashion with respect to size and Φ_T (see Fig. 5). Fig. 4 illustrated that the polyps are somewhat uniformly distributed. This implies that the ratio of the posterior probabilities must also follow an exponential function, which is represented in the next relation

$$\log \left(\frac{p(\mathbf{x} | \omega_p)}{p(\mathbf{x} | \omega_f)} \right) = d(\mathbf{x}) \quad (2)$$

in which $d(\mathbf{x})$ is the linear discriminant function of the feature vector and ω_p and ω_f denote the polyp class and the false detection class, respectively. One can recognize in (2) the assumption made by a logistic classifier, which corresponds to sigmoidal posterior probability density functions

$$\begin{aligned} p(\omega_p, \mathbf{x}) &= \frac{1}{1 + \exp(d(\mathbf{x}))} \\ p(\omega_f, \mathbf{x}) &= 1 - p(\omega_p, \mathbf{x}). \end{aligned} \quad (3)$$

The linear logistic classifier estimates the posterior probabilities $p(\omega_i | x)$ instead of the class-dependent distributions $p(x | \omega_i)$ [43]. These posterior distributions are assumed to be the sigmoidal functions. This is a valid assumption when e.g., the classes are distributed Gaussian, or, as in this case, one of the distributions is exponentially decreasing while the

other is more or less uniform. Then, a maximum likelihood (ML) estimation is made to find the linear direction in the data that best fits these assumed sigmoidal posterior functions. This ML estimator will give the weights of the discriminant function $d(\mathbf{x})$. Using the posterior probabilities instead of the class-dependent distribution functions makes this classifier less sensitive to the large class imbalance.

V. RESULTS

Classifier selection aims at choosing the best method for the regression analysis in our classification system (see Fig. 6). Three classifiers will be analyzed: the LDC, the QDC, and the logistic classifier (see Section IV). The specific choice will be based on two types of analysis: FROC analysis using a variety of sets of features in order to select the best classifier for the problem (instead of the best classifier for a specific feature set), and stability analysis by bootstrapping the training set.

The feature vector F in Fig. 6 consists of three features: Φ_T , $f_{I,\text{mean}}$, and f_{FOV} . Φ_T is related to the size of the candidates and is therefore directly used in the regression analysis, thus $f_O = \{\Phi_T\}$. The Mahalanobis distance mapping is applied to the other two features prior to the regression analysis. It is applied to $f_{D,1} = \{f_{I,\text{mean}}\}$ to sort all candidates based on the mean intensity in order of increasing distance to the normal tissue values of polyps; and to $f_{D,2} = \{f_{\text{FOV}}\}$ to aid discarding the candidates on the rectal tube. The added value of these features and the influence of the mappings will be analyzed in Section V-B.

In practice, the usefulness of a CAD system depends on whether it will generalize to data sets from different sources. The robustness of the complete system will be tested in Section V-C by means of an evaluation using data sets from four different medical centers (see Section II-A).

A. Classifier Selection: Performance and Stability

The performance of the classifiers was analyzed by means of FROC analysis. The FROC curves were calculated for a large pool of different feature sets to secure that the classifier selection step is not dependent on a certain choice of features. The FROC curves were calculated from a repeated ten-fold cross-validation. Only data set ‘‘A’’ was used in this learning phase to remain completely independent of the other data sets.

The aggregate of the different sets of features employed in the experiment will be called the feature pool. This pool was not created in order to select the best features, but merely to study the performance of the classifiers without choosing a specific feature set first. If some feature set were chosen first (before the classifier selection step), one might select the best classifier for the specific set of features and not necessarily the classifier which is best for the problem at hand. The feature pool consisted of 29 sets of features chosen from a total of nine different features: three protrusion-based features Φ_T with various thresholds T : 0.5, 0.6 and 0.7 mm; the features related to the intensity (i.e., the mean, maximum, minimum, and median intensity and the standard deviation of the intensity) and f_{FOV} to discard candidates on the rectal tubes. Each set contained at most five features of which one was chosen from the set of protrusion-based features.

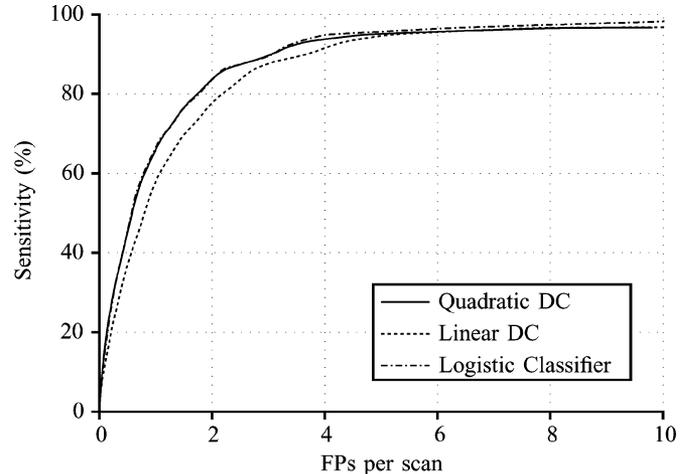


Fig. 7. FROC curves averaged over all feature sets for the LDC, QDC, and logistic classifiers.

TABLE II
INSTABILITY OF VARIOUS CLASSIFIERS

Classifier	Instability	Percentage (%)
Logistic	33.7	0.11
QDC	220.0	0.76
LDC	15.6	0.05

An FROC curve was computed for each classifier and for each set of features from the pool. The average FROC curve for a classifier is shown in Fig. 7. The standard deviation that was derived from the variation between the FROC curves for different feature sets was less than 0.03 FPs per scan for sensitivities below 95%. The FROC curves reveal that the logistic classifier and the QDC do not differ in their performance as their FROC curves almost completely overlap. The performance of LDC was significantly worse by approximately 15 times the standard deviation.

The second criterion used for classifier selection was the stability of the classifiers. This stability was assessed by means of bootstrapping the training set. This results in a perturbed orientation of the classifiers, which consequently leads to a number of differently classified candidates. The average number of different decisions is then used as a measure of instability [44]. Table II lists the instability measures. The table clearly shows that the logistic classifier and the LDC are the most stable classifiers. The instability has been measured for a sensitivity of 85%, but the results generalize well to other sensitivity levels, i.e., different locations of the decision boundary.

More specifically, it is noticeable that the LDC is much more stable than the QDC. This is explained by the covariance matrix estimated by the LDC being nearly identical to the covariance matrix of the majority class, which barely changes due to bootstrapping. On the other hand, the QDC also estimates a covariance matrix for the polyp class. Because of the low number of polyps, bootstrapping leads to a different covariance matrix for the polyp class. This is reflected by the poor instability of the QDC. The logistic classifier is expected to be more stable since it poses an assumption onto the relative posterior probabilities of the two classes rather than estimating both (class-dependent) probability distribution functions.

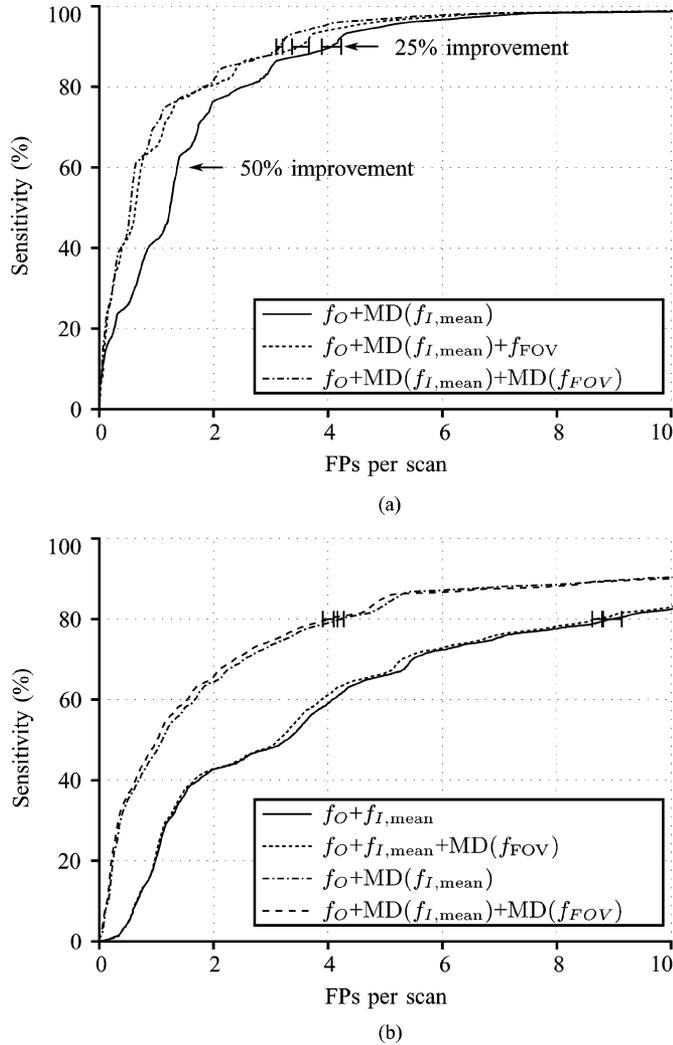


Fig. 8. FROC curves that indicate the added value of the feature f_{FOV} and the use of the Mahalanobis distance mapping. (a) Data set “A” with and without f_{FOV} . Using the Mahalanobis distance mapping leads to a small increase in performance. (b) Data set “C” with and without f_{FOV} and with the unmapped and mapped mean intensity feature. The graph reveals that it is an absolute necessity to apply the mapping in the case of fecal-tagged data.

To conclude, it is shown that the logistic classifier combines a good performance in terms of FROC analysis with a good stability value. Therefore, the logistic classifier will be used as the regressor in the classification system.

B. Outlier Rejection by Mahalanobis Distance Mapping

Let us now look into the performance of outlier rejection by the Mahalanobis distance mapping. The starting point of our analysis is the FROC curve generated by the logistic classifier using Φ_T with a threshold T of 0.6 mm, and $f_{I,mean}$ (prior to mapping). FROC curves are computed for data sets “A” and “C.” Among other differences, these data sets differ in the type of rectal tubes used and the administration of a fecal tagging agent (see also Table I).

Fig. 8(a) shows the FROC curves for data set “A.” In this data set, no fecal tagging agent was administered to the patients. As a consequence, only false detections with low mean intensities were present. This means that this feature is already

suitable for ordering the candidates. Mapping $f_{I,mean}$ did not result in a significantly different FROC curve; for this reason and for the purpose of clarity the curves with the “unmapped” $f_{I,mean}$ are not shown. The solid curve is the FROC curve of a system with only the $MD(f_{I,mean})$ and Φ_T . The dotted line is obtained when the feature f_{FOV} is added directly, without prior Mahalanobis distance mapping; the dash-dotted FROC curve is the outcome when a mapped version of this feature is used instead. The improvement by adding this feature may be a reduction up to 25%–50% of the number of false positives depending on the required sensitivity (see arrows). The error bars denote two times the standard deviation of the number of false positives over all scans.

The results for data set “C” are shown in Fig. 8(b). In contrast to data set “A,” patients from this data set were administered a fecal tagging agent. As a consequence, it may be expected that the Mahalanobis distance mapping of $f_{I,mean}$ has a larger influence due to the presence of both candidates with a low mean intensity as candidates with a high mean intensity. Here again, the solid curve corresponds to classification using Φ_T and $f_{I,mean}$. Similar to the analysis of data set “A,” the feature f_{FOV} is added and the MD-mapping is applied to this feature and to $f_{I,mean}$. In contrast to the rectal tubes in data set “A,” the tubes in this data set did not have a balloon attached, but included a marker of high attenuation material. Because of this, less candidates on the rectal tubes were found and those which were found could often be easily discarded by means of their intensity. As a consequence, adding the feature f_{FOV} may be expected not to improve the performance. This is confirmed by the dotted line, indicating no significant improvement. Again, for the purpose of clarity, the FROC curves with the “unmapped” f_{FOV} are not shown in this figure, as they do not differ significantly. Observe that adding f_{FOV} does not lead to worse results.

The second step was to compute the same FROC curves with the mapped mean intensity feature. A striking improvement can be seen. This result can be explained by the fact that in this case there are both false detections with lower mean intensity as there are false detections with higher mean intensity. According to these results, only the mapped features will be used in further FROC analyses.

C. Multicenter Evaluation

An important aspect of a CAD system for CT colonography is its ability to generalize to data sets differing in a variety of aspects. The generalization power of the presented system will be investigated by FROC analysis and a cross-center evaluation.

The patients from data sets “A,” “B,” and “C” were scanned in both prone and supine positions. At the basis of this (conventional) approach is that a polyp is not always visible in both CT scans, e.g., due to suboptimal distension or remaining fluid rests. Consequently, a polyp may not be annotated in both scans. Let us initially focus on the annotated polyp “findings” to assess the performance of the candidate detection step.

The candidate detection returned 88.8% (436/491) of the annotated findings ≥ 6 mm in total (see Table III). The preparation of the patients is at the basis of the differences in the number of missed findings. The patients of data set “A” had undergone an extensive preparation. This might explain the fact that the system detected almost all annotations in this data set

TABLE III
RESULTS OF THE CANDIDATE DETECTION SYSTEM

Data set	Polyp findings (≥ 6 mm)			Polyps (≥ 6 mm)			Number of false detections
	Number of annotations	Number of detections	Detection rate (%)	Number of annotations	Number of detections	Detection rate (%)	
'A'	94	93	99	59	59	100	28 678
'B'	49	38	78	28	28	100	12 334
'C'	340	297	87	176	174	99	53 698
'D'	8	8	100	8	8	100	8026
Total	491	436	89	271	269	99	102 736

(93/94). On the other hand, data set "B" appeared to contain a large amount of residual fluid (confirmed by [45]). Consequently, many polyps were obscured by fecal remains, reducing the detection rate to 77.6% (38/49). Data set "C" had less contrast-enhanced fluid in the colon, which resulted in a higher detection rate of 87.4% (297/340). The percentage of polyps detected in either scan was 99.0% (269/271) (sensitivity is conventionally measured in this way [46]).

Fig. 9 shows the results of the cross-center evaluation. It is generally known that a large amount of features decreases the generalization power of a classifier, especially when the data sets differ as much as the four data sets of our study. Therefore, we consciously limited the number of features in this evaluation to the three features described before: Φ_T with a threshold 0.6 mm, $MD(f_{I,\text{mean}})$, and $MD(f_{\text{FOV}})$. Each graph in Fig. 9 corresponds to one test set; the line styles in the figures indicate the specific data set on which the classifier was trained. In the case of testing and training on the data from the same medical center, a ten-fold, repeated cross-validation was performed. The standard deviation indicated in the graphs is estimated as the standard deviation of a binomial distribution [47] and depends on the number of polyps and the sensitivity. This standard deviation characterizes the variation in the FROC curves when a new subset is drawn from the same distribution.

It can be seen that in all graphs, the FROC curves for classifiers trained on the different data sets are generally within one standard deviation from each other. In other words, the same performance is attained no matter on which data set the classifier is trained. Concurrently, there are small differences in the performance of the CAD system for the four data sets. Despite this, all yield a sensitivity larger than 85% at the cost of five false positive detections per scan. Four polyps in data set "B" remained undetected at 86% (25/29) sensitivity. The missed polyps were all reviewed by a fellow researcher with a background in CAD of polyps in CTC. All missed polyps were covered by contrast-enhanced material in at least one of the two scans and were annotated in only one position. Consequently (no electronic cleansing was used), the CAD system did not get a second chance of finding these polyps. In data set "C," 14 polyps remained undetected by the CAD system at 90% sensitivity. The false negatives consisted of tumors with lobulated shapes, polyps covered by fecal remains, "nonprotruding" polyps annotated as a flat polyp by the radiologists and polyps that were located between haustral folds. Even though data set "D" contained only one scan per patient, the FROC curves for this data set compete with the FROC curves for the other data sets.

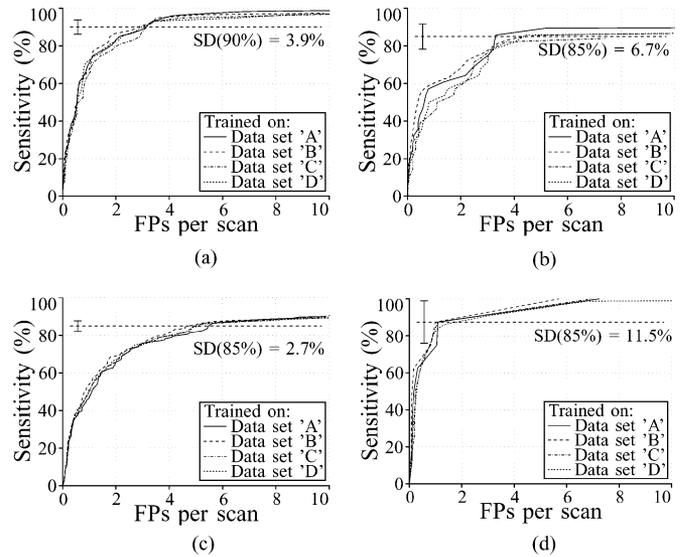


Fig. 9. Each graph shows the results of classifying a certain data set, using four different classifiers that are each trained on one of the four data sets. The line style indicates the data set on which is trained. When the same data set is used for training and classifying, a ten-fold, repeated cross-validation was used. (a) Test set "A," (b) Test set "B," (c) Test set "C," and (d) Test set "D."

In conclusion, the FROC curves for the different data sets show that the CAD system is independent on the specific data set used for training. The differences between the curves are a result of the administration of a fecal tagging agent, the preparation of the patients and natural fluctuations in the appearance of the polyps in the data sets.

VI. DISCUSSION/CONCLUSION

We developed a classification system based on logistic regression for CAD of polyps in CT colonography data. Typically, there are unbalanced and unknown misclassification costs and a huge class imbalance. The latter occurs because there are only a few examples of the abnormality class in a sheer endless sea of normal "healthy" samples. Our classification system can cope with the aforementioned characteristics by carrying out a regression analysis instead of classifying the candidates into one of the two classes. The ordering correlates with the clinical relevance of the candidates. The exponential distribution of the candidates and the small number of polyps available for training led to the use of the logistic classifier for regression. The logistic classifier is low-complex and proved to be stable.

Candidates were detected based on their protrudedness from the colon wall. A feature derived from the protrusion field was sensitive for candidates that had steep edges and large protrusion. Other features used were the internal intensity distribution, and a feature to discard detections on the rectal tubes.

The features were divided into two types of features, namely features that allowed directly an ordering of the candidates and features that were well described by a Gaussian density distribution. The features of the second type were mapped by a Mahalanobis distance mapping to impose an ordering. This mapping was chosen because it emulates a Gaussian one-class classifier. In this way, outlier rejection was incorporated into the classification system.

After discarding the candidates on the rectal tubes, polyps and non-polyps could be distinguished using only information about the protrusion and the internal intensity of the candidates. The observed sensitivity was comparable to the sensitivity of radiologists using CTC [7], [15], [16] and competed with other CAD systems [7]–[9], [26]. It was also shown that the CAD system generalizes well to data sets from different medical centers.

To conclude, we introduced a low-complex CAD system that took into account all the characteristics of the classification problem. These characteristics will frequently occur in medical image processing problems. The Mahalanobis distance mapping in conjunction with logistic regression is generally applicable to obtain a clinically relevant ordering of the candidates. For automatic polyp detection, the generalization to data sets from different medical centers and with different patient preparations is essential to application in large-scale screening.

ACKNOWLEDGMENT

The authors would like to thank Dr. R. Choi, Virtual Colonoscopy Center, Walter Reed Army Medical Center, Washington, DC; Dr. P. Rogalla, Charité Hospital, Humboldt University, Berlin, Germany; and Dr. P. J. Pickhardt, University of Wisconsin Medical School, Madison, WI, for providing the data sets.

REFERENCES

- Colorectal Cancer Facts & Figures Amer. Cancer Soc., Atlanta, GA, Tech. Rep. 8617.00, 2005.
- B. C. Morson, "Evolution of cancer of the colon and rectum," *Cancer*, vol. 34, pp. 345–349, 1974.
- J. H. Bond, "Clinical evidence for the adenoma-carcinoma sequence, and the management of patients with colorectal adenomas," *Semin Gastrointest. Dis.*, vol. 11, pp. 176–184, 2000.
- P. J. Pickhardt, "CT colonography (virtual colonoscopy) for primary colorectal screening: Challenges facing clinical implementation," *Abdom. Imag.*, vol. 30, pp. 1–4, 2005.
- J. T. Ferrucci, "Colon cancer screening with virtual colonoscopy: Promise, polyps, politics," *Am. J. Roentgenol.*, vol. 177, pp. 975–988, 2001.
- S. Winawer, R. Fletcher, D. Rex, J. Bond, R. Burt, J. Ferrucci, T. Ganiats, T. Levin, S. Woolf, D. Johnson, L. Kirk, S. Litin, and C. Simmann, "Colorectal cancer screening and surveillance: Clinical guidelines and rationale—Update based on new evidence," *Gastroenterology*, vol. 124, pp. 544–560, 2003.
- R. M. Summers, J. Yao, P. J. Pickhardt, M. Franaszek, I. Bitter, D. Brickman, V. Krishna, and J. R. Choi, "Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population," *Gastroenterology*, vol. 129, pp. 1832–1844, 2005.
- J. Näppi and H. Yoshida, "Fully automated three-dimensional detection of polyps in fecal-tagging CT colonography," *Acad. Radiol.*, vol. 14, pp. 287–300, 2007.
- R. M. Summers, L. R. Handwerker, P. J. Pickhardt, R. L. van Uitert, K. K. Deshpande, S. Yeshwant, J. Yao, and M. Franaszek, "Performance of a previously validated CT colonography computer-aided detection system in a new patient population," *Am. J. Roentgenol.*, vol. 191, pp. 169–174, 2008.
- H. Yoshida and J. Näppi, "CAD in CT colonography without and with oral contrast agents: Progress and challenges," *Comput. Med. Imag. Graph.*, vol. 31, pp. 267–284, 2007.
- P. J. Pickhardt, A. D. Lee, E. G. McFarland, and A. J. Taylor, "Linear polyp measurement at CT colonography: In vitro and in vivo comparison of two-dimensional and three-dimensional displays," *Radiology*, vol. 236, pp. 872–878, 2005.
- M. E. Zalis, M. A. Barish, J. R. Choi, A. H. Dachman, H. M. Fenlon, J. T. Ferrucci, S. N. Glick, A. Laghi, M. Macari, E. G. McFarland, M. M. Morrin, P. J. Pickhardt, J. Soto, and J. Yee, "CT colonography reporting and data system: A consensus proposal," *Radiology*, vol. 236, pp. 3–9, 2005.
- P. J. Pickhardt, C. Hasssan, A. Laghi, A. Zullo, D. H. Kim, and S. Morini, "Cost-effectiveness of colorectal cancer screening with computed tomography colonography," *Cancer*, vol. 109, pp. 2213–2221, 2007.
- S. A. Taylor, A. Laghi, P. Lefere, S. Halligan, and J. Stoker, "European society of gastrointestinal and abdominal radiology (ESGAR): Consensus statement on CT colonography," *Eur. Radiol.*, vol. 17, pp. 575–579, 2007.
- R. E. van Gelder, C. Y. Nio, J. Florie, J. F. Bartelsman, P. Snel, S. W. de Jager, S. J. van Deventer, J. S. Laméris, P. M. Bossuyt, and J. Stoker, "Computed tomographic colonography compared with colonoscopy in patients at increased risk for colorectal cancer," *Gastroenterology*, vol. 127, no. 1, pp. 41–8, 2004.
- P. J. Pickhardt, J. R. Choi, I. Hwang, J. A. Butler, M. L. Puckett, H. A. Hildebrandt, R. K. Wong, P. A. Nugent, P. A. Mysliwiec, and W. R. Schindler, "Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults," *N. Eng. J. Med.*, vol. 349, pp. 2191–2200, 2003.
- D. H. Kim, P. J. Pickhardt, A. J. Taylor, W. K. Leung, T. C. Winter, J. L. Hinshaw, D. V. Gopal, M. Reichelderfer, R. H. Hsu, and P. R. Pfau, "CT colonography versus colonoscopy for the detection of advanced neoplasia," *N. Eng. J. Med.*, vol. 357, pp. 1403–1412, 2007.
- Virtuelle Koloskopie*, [Online]. Available: <http://radiologie.charite.de>
- R. M. Summers, W. S. Selbie, J. D. Malley, L. M. Pusanik, A. J. Dwyer, N. A. Courcousakis, D. J. Shaw, D. E. Kleiner, M. C. Sneller, C. A. Langford, S. M. Holland, and J. H. Shellhamer, "Polypoid lesions of airways: Early experience with computer-assisted detection by using virtual bronchoscopy and surface curvature," *Radiology*, vol. 208, pp. 331–337, 1998.
- R. M. Summers, C. D. Johnson, L. M. Pusanik, J. D. Malley, A. M. Youssef, and J. E. Reed, "Automated polyp detection at CT colonography: Feasibility assessment in a human population," *Radiology*, vol. 219, pp. 51–59, 2001.
- H. Yoshida and J. Näppi, "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps," *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1267–1274, Dec. 2001.
- H. Yoshida, J. Näppi, P. MacEneaney, D. T. Rubin, and A. H. Dachman, "Computer-aided diagnosis scheme for detection of polyps at CT colonography," *Radiograph.*, vol. 22, no. 4, pp. 963–979, 2002.
- G. Kiss, J. van Cleynenbreugel, S. Drisis, D. Bielen, G. Marchal, and P. Suetens, "Computer-aided detection for low-dose CT colonography," in *Proc. MICCAI'05*, 2005, vol. LNCS 3749, pp. 859–867.
- G. Kiss, S. Drisis, D. Bielen, F. Maes, J. van Cleynenbreugel, G. Marchal, and P. Suetens, "Computer-aided detection of colonic polyps using low-dose CT acquisitions," *Acad. Radiol.*, vol. 13, no. 9, pp. 1062–1071, 2006.
- J. Näppi and H. Yoshida, "Feature-guided analysis for reduction of false positives in CAD of polyps for computed tomographic colonography," *Med. Phys.*, vol. 30, no. 7, pp. 1592–1601, 2003.
- A. Jerebko, S. Lakare, P. Cathier, S. Periaswamy, and L. Bogoni, "Symmetric curvature patterns for colonic polyp detection," in *Proc. MICCAI'06*, 2006, vol. LNCS 4191, pp. 169–176.
- Z. Wang, Z. Liang, L. Li, X. Li, B. Li, J. Anderson, and D. Harrington, "Reduction of false positives by internal features for polyp detection in CT-based virtual colonography," *Med. Phys.*, vol. 32, no. 12, pp. 3620–3616, 2005.
- P. Sundaram, A. Zomorodian, C. Beaulieu, and S. Napel, "Colon polyp detection using smoothed shape operators: Preliminary results," *Med. Imag. Anal.*, vol. 12, no. 2, pp. 99–119, 2008.

- [29] S. B. Göktürk, C. Tomasi, B. Acar, C. F. Beaulieu, D. S. Paik, R. B. Jeffrey, Jr., J. Yee, and S. Napel, "A statistical 3-D pattern processing method for computer-aided detection of polyps in CT colonography," *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1251–1260, Dec. 2001.
- [30] A. K. Jerebko, J. D. Malley, M. Franaszek, and R. M. Summers, "Multiple neural network classification scheme for detection of colonic polyps in CT colonography data sets," *Acad. Radiol.*, vol. 10, pp. 154–160, 2003.
- [31] Y. Zheng, X. Yang, and G. Beddoe, "Reduction of false positives in polyp detection using weighted support vector machines," in *Proc. 29th IEEE EMBS*, Aug. 2007, pp. 4433–4436.
- [32] M. A. Kupinski, D. C. Edwards, M. L. Giger, and C. E. Metz, "Ideal observer approximation using Bayesian classification neural networks," *IEEE Trans. Med. Imag.*, vol. 20, no. 9, pp. 886–899, Sep. 2001.
- [33] K. Suzuki, H. Yoshida, J. Näppi, and A. H. Dachman, "Massive-training artificial neural network (MTANN) for reduction of false positives in computer-aided detection of polyps: Suppression of rectal tubes," *Med. Phys.*, vol. 33, no. 10, pp. 3814–3824, 2006.
- [34] L. Bogoni, P. Cathier, M. Dundar, A. Jerebko, S. Lakare, J. Liang, S. Periaswamy, M. E. Baker, and M. Macari, "Computer-aided detection (CAD) for CT colonography: A tool to address a growing need," *Br. J. Radiol.*, vol. 78, pp. 57–62, 2005.
- [35] C. van Wijk, V. F. van Ravesteijn, F. M. Vos, R. Truyen, A. H. de Vries, J. Stoker, and L. J. van Vliet, "Detection of protrusions in curved folded surfaces applied to automated polyp detection in CT colonography," in *Proc. MICCAI'06*, 2006, vol. LNCS 4191, pp. 471–478.
- [36] J. Dehmeshki, S. Halligan, S. A. Taylor, M. E. Roddie, J. McQuillan, L. Honeyfield, and H. Amin, "Computer assisted detection software for CT colonography: Effect of sphericity filter on performance characteristics for patients with and without fecal tagging," *Eur. Radiol.*, vol. 17, no. 3, pp. 662–668, 2007.
- [37] E. Konukoglu, B. Acar, D. S. Paik, C. F. Beaulieu, and S. Napel, "HDF: Heat diffusion fields for polyp detection in CT colonography," *Signal Process.*, vol. 87, no. 10, pp. 2407–2416, 2007.
- [38] D. M. J. Tax, "One-class classification," Ph.D. dissertation, Delft Univ. Technol., Delft, The Netherlands, Jun. 2001.
- [39] C. van Wijk, V. F. van Ravesteijn, F. M. Vos, and L. J. van Vliet, "Detection and segmentation of colonic polyps on implicit isosurfaces by second principal curvature flow," *IEEE Trans. Med. Imag.*, accepted for publication.
- [40] G. Iordanescu and R. M. Summers, "Reduction of false positives on the rectal tube in computer-aided detection for CT colonography," *Med. Phys.*, vol. 31, no. 10, pp. 2855–2862, 2004.
- [41] C. van Wijk, J. Florie, C. Y. Nio, E. Dekker, A. H. de Vries, H. W. Venema, L. J. van Vliet, J. Stoker, and F. M. Vos, "Protrusion method for automated estimation of polyp size on CT colonography," *Am. J. Roentgenol.*, vol. 190, no. 5, pp. 1279–1285, 2008.
- [42] I. W. O. Serlie, F. M. Vos, H. W. Venema, and L. J. van Vliet, CT imaging characteristics 2006 [Online]. Available: <http://www.ist.tudelft.nl/qi>
- [43] A. R. Webb, *Statistical Pattern Recognition*, 2nd ed. New York: Wiley, 2002.
- [44] M. Skurichina, "Stabilizing Weak Classifiers," Ph.D. dissertation, Delft Univ. Technol., Delft, The Netherlands, Oct. 2001.
- [45] I. W. O. Serlie, A. H. de Vries, F. M. Vos, Y. Nio, R. Truyen, J. Stoker, and L. J. van Vliet, "Lesion conspicuity and efficiency of CT colonography with electronic cleansing based on a three-material transition model," *Am. J. Roentgenol.*, vol. 191, no. 5, pp. 1493–502, 2008.
- [46] M. A. Barish, J. A. Soto, and J. T. Ferrucci, "Consensus on current clinical practice of virtual colonoscopy," *Am. J. Roentgenol.*, vol. 184, pp. 786–792, 2005.
- [47] C. Chatfield, *Statistics for Technology*, 3rd ed. London, U.K.: Chapman & Hall, 1983.