Robust Large Scale Prone-Supine Polyp Matching Using Local Features: A Metric Learning Approach

Meizhu Liu, Le Lu, Jinbo Bi, Vikas Raykar, Matthias Wolf, Marcos Salganicoff

Siemens Medical Solutions USA, Malvern, PA 19355 (le-lu@siemens.com)
University of Florida, Gainesville, FL 32611 (mliu@cise.ufl.edu)

Abstract. Computer aided detection (CAD) systems have emerged as noninvasive and effective tools, using 3D CT Colonography (CTC) for early detection of colonic polyps. In this paper, we propose a robust and automatic polyp prone-supine view matching method, to facilitate the regular CTC workflow where radiologists need to manually match the CAD findings in prone and supine CT scans for validation. Apart from previous colon registration approaches based on global geometric information [1–4], this paper presents a *feature selection* and *metric distance learning* approach to build a pairwise matching function (where true pairs of polyp detections have smaller distances than false pairs), learned using *local polyp classification features* [5–7]. Thus our process can seamlessly handle collapsed colon segments or other severe structural artifacts which often exist in CTC, since only local features are used, whereas other global geometry dependent methods may become invalid for collapsed segmentation cases. Our automatic approach is extensively evaluated using a large multi-site dataset of 195 patient cases in training and 223 cases for testing. No external examination on the correctness of colon segmentation topology [2] is needed. The results show that we achieve significantly superior matching accuracy than previous methods [1–4], on at least one order-of-magnitude larger CTC datasets.

1 Introduction

Colon cancer is the second leading cause of cancer death in western countries, but it is one of the most preventable of cancers because doctors can identify and remove its precursor known as a polyp. Besides the well established fiber-optic colonoscopy, 3D Computed Tomography Colonography (CTC), or virtual colonoscopy has emerged as a powerful screening tool for polyp detection and diagnosis. The research field of computer aided detection (CAD) of colonic polyps in CTC is highly exploited [5–7]. To enhance polyp findings in collapsed or fluid-tagged colon segments, and better distinguish polyps from pseudo polyps (e.g. tagged stools), the current CTC clinical practice is to obtain two scans of a patient in prone and supine positions, respectively. However colon can move and deform significantly differently between the prone and supine scans, which makes

the manual registration of polyp findings or colon segments difficult, inaccurate and time-consuming.

In this paper, we present a novel computerized technique to achieve high performance polyp matching, by supervisedly optimizing a distance metric in the feature space of polyp classification where true pairs of polyp matches statistically have smaller distance than false pairs. A polyp instance may be represented by a variety of *local appearance features* for classification [5–7], including local geometric features, and morphological, shape/intensity and context features. Since the total union of these features may lead to redundancy, greater computational and spatial complexity, we first use *feature selection* method to choose the features that are most relevant to polyp matching (e.g., the feature difference variation is minimal between true polyp matches), but least redundant. After selecting a subset of task-specific features, from the polyp classification feature pool, we propose an efficient *metric learning* method to learn a *covariance-matrix boosted Mahalanobis distance* to measure the instance differences across views.

Extensive evaluation is executed using a representative, multi-site clinical database with 195 patient cases in training and 223 cases for testing, containing 106, 118 polyps respectively. We demonstrate superior performance results on polyp prone-supine view matching, compared with existing work mostly based on colon centerline/surface registration [1–4]. Note that previous polyp matching techniques are tested and reported on datasets which are at least one order-of-magnitude smaller than ours, as 20 [1] (with 20 polyps), 12 [2] (with 12 polyps), and 39 [3] (with 23 training and 16 testing polyps) patients/cases. This is partially because the pair of completely distended prone-supine colon scans (from rectum to cecum) is a prerequisite, and preparing topologically correctly segmented colon cases often requires manual editing or interactions [2] and can be labor-intensive, given a large number of 3D volumes.

2   Materials and Methods

Our approach consists of the following two steps. We first select a subset of features from the whole CAD classification feature pool, which is polyp matching-informative, using Minimum Redundancy Maximum Relevance (MRMR) algorithm [8]. Next, we learn an effective polyp matching distance metric on selected features (i.e., Mahalanobis distance by a positive semidefinite matrix that weights channels of features differently), in an additive, boosting based optimization manner. Fig. 1 summarizes the process diagram.

2.1   Matching-Sensitive Feature Selection

Data & Features: We collected 195 CTC cases (or 390 volumes) with 106 polyps appearing in both views for training; and 223 cases containing 118 polyps with double appearance for testing, from 8 hospitals in US, Europe and Asia. Images are acquired using Siemens, GE and Philips scanners with fecal tagging preparation. Only actionable polyps with diameters $\geq 6mm$ are considered. Our
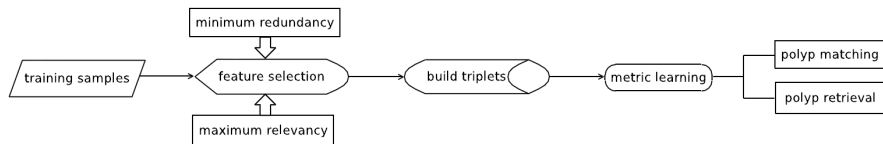
Fig. 1. The flow chart of using metric learning to do polyp matching and retrieval.

CAD system builds a tree-structured probabilistic classifier using 96 (morphological, intensity, texture-based or geometrical) features $\mathcal{F} = \{f_i\}$ on 61257 candidates in training dataset. Thus we can first perform a thresholding process to rule out false positive (FP) candidates with low probability values (ρ of being polyp). After this, we have ~ 8 candidates per patient with true positive (TP) detection sensitivities at 94.2% and 92.9% for training and testing perspectively. Note that our polyp matching approach is applicable on other CAD systems [5–7] which usually have a large variety of polyp descriptive features.

Let $\boldsymbol{x}_i^1$ be a true polyp instance in one view for a patient and let $\{\boldsymbol{x}_j^2\}$ be the set of corresponding instances in the other view. Note that the size of $\{\boldsymbol{x}_j^2\}$ can be larger than one since polyps can appear as two or more instances in each scan, especially for large polyps. This is called multiple instance problem. Here 1 or 2 indicates prone or supine view, without loss of generality. We define the positive (+) instance pairs of instances in two views rooted from the same unique true polyp, and other pairs as negative (-) (e.g., TP-TP pairs according to different polyps, TP-FP pairs, and FP-FP pairs). For each original feature f, a new "difference-of-feature" variable can be derived as $\triangle f = (f_i^1 - f_j^2)$, which is expected to be zero or a constant for positive pair population (i.e., tightly distributed in a statistical sense), or random for negatives.

Based on above motivation, we use feature selection algorithms to find a subset of $S \subset \mathcal{F}$ which are more informative on distinguishing true or false polyp pairs. For its numerical stability, accuracy and formulation simplicity, we choose Minimum Redundancy Maximum Relevance (MRMR) method [8] though other feature selection methods are also applicable. For self-contained content, we review MRMR algorithm as follows. $\triangle f$ is abbreviated as f, with a slight abuse of notation. For feature set $\mathcal{F} = \{f_i\}$, its MRMR feature subset S means that the average mutual information between the feature set S and the class labels is large while the mutual information between the features in S is small. The mutual information between feature set S and the class label set $\mathbf{y}$ is defined

$$I(S, \mathbf{y}) = \frac{1}{m} \sum_{f_i \in S} I(f_i, \mathbf{y}) \tag{1}$$

where m is the cardinality of S, I is the mutual information and $\mathbf{y}$ represents the positive or negative $\{+1; -1\}$ matched pair as defined above. The mutual

information between features in the set S is

$$I(S) = \frac{1}{m^2} \sum_{f_i, f_j \in S} I(f_i, f_j). \tag{2}$$

Then the objective of MRMR can be described as the following function

$$\gamma(S, \mathbf{y}) = \max_{S \subset \mathcal{F}} I(S, \mathbf{y}) - I(S) \tag{3}$$

The MRMR feature selection is a sequential process. The ith feature $f_i^\star$ can be chosen, given selected S_{i-1}, according to

$$f_i^\star = \arg \max_{f \in \mathcal{F} - S_{i-1}} I(f, \mathbf{y}) - \frac{1}{i-1} \sum_{f_j \in S_{i-1}} I(f, f_j)$$

Then the ith feature will be added to S_{i-1} to form $S_i = S_{i-1} \cup f_i$. If $\gamma(S_i, \mathbf{y}) < \gamma(S_{i-1}, \mathbf{y})$, then S_{i-1} is the optimal feature subset, and MRMR feature selection stops. Using the MRMR feature selection method, we obtain 18 features out of 96 features, describing polyp's shape morphology, segmented size, surface area, intensity profiles, classification score ("polypness") and their joint statistics. In the dimensions of these selected (individual instance appearance-specific) features, we will propose a metric learning algorithm for deriving a Mahalanobis distance to match polyps across prone-supine views.

2.2 Matching by Metric Distance Learning

In this section, we propose a new metric learning method called "MatrixBoost" to match polyps in prone-supine views, using the 18 matching sensitive polyp features selected by MRMR. The basic idea is that a good distance metric can be learned to assign different weights on features, so that low distances are given for pairs of instances to be matched and high distances for others, in the feature space. There are a number of ways to design distance metrics [9–11]. Metric learning can be derived from the optimal combination of weak learners to form a strong learner, based on training data. One type of weak learner is equivalence constrained, where equivalence constraints are provided for pairs $(\boldsymbol{x_i}, \boldsymbol{x_j})$, each associated with a binary label of "similar" or "dissimilar" [9]. Another weak learner representation often used in information retrieval [11] is the proximity relationship over triplet set $\mathcal{T} = \{(i, j, k)\}$, meaning that $\boldsymbol{x_i}$ is closer to $\boldsymbol{x_j}$ than to $\boldsymbol{x_k}$. Here $\boldsymbol{x_i}$ is the feature vector representation for the polyp instance i. The goal of metric learning is to learn a distance function d such that $d(\boldsymbol{x_i}, \boldsymbol{x_j}) < d(\boldsymbol{x_i}, \boldsymbol{x_k})$, and

$$d(\boldsymbol{x_i}, \boldsymbol{x_j}) = (\boldsymbol{x_i} - \boldsymbol{x_j})' M (\boldsymbol{x_i} - \boldsymbol{x_j}) \tag{4}$$

where $'$ is the vector/matrix transpose transformation and M is a positive semidefinite (PSD) matrix that leads to the Mahalanobis distance metric [9,10]. We follow the Mahalanobis distance metric formulation, but propose to construct the "covariance" matrix M by additively combining weak learners which are low rank PSD matrices. AdaBoost [12] method is utilized to learn the linear combination of low rank positive semidefinite (PSD) matrices, as a PSD matrix

Algorithm 1 Algorithm: Metric Learning by MatrixBoost

Input: A set of feature vectors $X = [\boldsymbol{x_1} \ \boldsymbol{x_2} \ \cdots \ \boldsymbol{x_n}]$,
The triplet set $\mathcal{T} = \{(i,j,k) \mid \boldsymbol{x_i}$ is closer to $\boldsymbol{x_j}$ than to $\boldsymbol{x_k}\}$.
Initialize $D_1((i,j,k))$ (usually set to $1/m$ where m is the total number of triplets).
repeat
   Learn the optimal weak hypothesis $h_t(\boldsymbol{x}, \boldsymbol{y}) : \mathcal{X} \times \mathcal{X} \to \mathcal{R}$ according to distribution D_t.
   Choose $\alpha_t \in \mathcal{R}$.
   Update distribution

$$D_{t+1}((i,j,k)) = \frac{D_t((i,j,k)) \exp(\alpha_t(h_t(\boldsymbol{x_i}, \boldsymbol{x_j}) - h_t(\boldsymbol{x_i}, \boldsymbol{x_k})))}{Z_t} \tag{5}$$

   where Z_t is a normalization factor (chosen so that D_{t+1} is a distribution).
until Some convergence criterion is met.
Output the final model $H(\cdot, \cdot) = \sum_t \alpha_t h_t(\cdot, \cdot)$.

M preserving the proximity relationships among triplet set $\mathcal{T} = \{(i,j,k)\}$. The input to our metric boosting algorithm for training are triplets of instances, with inequality constraints on distances as defined above.

Build triplets: In training, we select all the instances with classifier score greater than $\rho \geq \beta_c$ to build triplets. The classifier score threshold β_c is chosen to make a practically feasible trade-off between detection sensitivity and FP rate (sensitivity is 0.946, and FP rate per patient is 7.59, pruning obvious negatives). The retained instances will form the triplets in the following way. For each true positive (polyp) instance $\boldsymbol{x_i}$ in the prone view of a patient, we find all the positive instances $\{\boldsymbol{x_j}\}_{j=1}^n$ corresponding to the same polyp and all other instances (including positives corresponding to different polyps and negatives, or false positives) $\{x_k\}_{k=1}^m$ in the supine view. Then (i,j,k) will form a triplet, requiring $d(\boldsymbol{x_i}, \boldsymbol{x_j}) < d(\boldsymbol{x_i}, \boldsymbol{x_k})$. We repeat the same process on each true positive instance in the supine view to build more triplets, in a similar way. All the triplets form a triplet set $\mathcal{T}$ and we obtained 8646 triplets in total, which will be used as inputs for our metric learning algorithm to optimize the PSD matrix M.

Learn a PSD matrix using MatrixBoost: Since a PSD matrix can be Eigen-decomposed as a combination of lower rank matrix, e.g., $M = \alpha_i U_i$, where $U_i = \boldsymbol{u_i u_i}'$. The distance between two instances $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ is $d(\boldsymbol{x_i}, \boldsymbol{x_j}) = (\boldsymbol{x_i} - \boldsymbol{x_j})' M (\boldsymbol{x_i} - \boldsymbol{x_j})$. The algorithm is to learn a strong learner $H(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})' M (\boldsymbol{x} - \boldsymbol{y})$, which is a combination of weak learners $h_t(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})' U_t (\boldsymbol{x} - \boldsymbol{y})$, i.e. $H(\boldsymbol{x}, \boldsymbol{y}) = \sum_t \alpha_t h_t(\boldsymbol{x}, \boldsymbol{y})$, by minimizing the error rate of triplets violating the distance inequality as below.

$$\epsilon = \sum_{(i,j,k) \in \mathcal{T}} D((i,j,k)) \mathbf{1}_{(H(\boldsymbol{x_i}, \boldsymbol{x_j}) - H(\boldsymbol{x_i}, \boldsymbol{x_k}))}, \tag{6}$$

where D is a probability distribution over $\mathcal{T}$, and $\mathbf{1}$ is the Heaviside step function ($\mathbf{1}_{(a)} = 0$ if $a < 0$, and 1 otherwise). Our proposed MatrixBoost algorithm is

shown in Algorithm 1. This algorithm adapts the merits of the AdaBoost [12] and the decomposable nature of PSD distance matrix. In the algorithm, the weak model is $h_t(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})'U_t(\boldsymbol{x} - \boldsymbol{y})$ where $U_t = \boldsymbol{u_t}\boldsymbol{u_t}'$, and the final hypothesis is $H(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})'M(\boldsymbol{x} - \boldsymbol{y})$ where $M = \sum_t \alpha_t U_t$. Note that if M forms a metric that satisfies the triplet conditions, so does its multiplier. It can be proven that the training error of the final hypothesis H as defined in (6) is upper bounded by $\prod_{t=1}^{T} Z_t$, i.e.,

$$\sum D((i,j,k))\mathbf{1}_{(H(\boldsymbol{x_i},\boldsymbol{x_j})-H(\boldsymbol{x_i},\boldsymbol{x_k}))}$$
$$\leq \sum D((i,j,k))\exp(H(\boldsymbol{x_i},\boldsymbol{x_j}) - H(\boldsymbol{x_i},\boldsymbol{x_k})),\ \mathbf{1}_x \leq \exp(x)$$
$$= \sum D_{T+1}((i,j,k))\prod_{t=1}^{T} Z_t = \prod_{t=1}^{T} Z_t.$$

α_t and h_t will be chosen such that the error upper bound $\prod_{t=1}^{T} Z_t$ will be minimized. Let $h_t \in [0, 1]$, Z_t has the upper bound

$$Z_t \leq e^{\alpha_t}\frac{1-r}{2} + e^{-\alpha_t}\frac{1+r}{2}, \tag{7}$$

where

$$r = \sum_{(i,j,k)\in\mathcal{T}} D_t((i,j,k))\left(h_t(\boldsymbol{x_i}, \boldsymbol{x_k}) - h_t(\boldsymbol{x_i}, \boldsymbol{x_j})\right). \tag{8}$$

The right side of (7) can be minimized when $\alpha_t = \ln((1+r)/(1-r))/2$ which corresponds to $Z_t \leq \sqrt{1-r^2}$. Obviously, $Z_t \leq 1$ and if $r > 0$, we have $\alpha_t > 0$. Furthermore, the inequality implies that we can achieve smaller Z_t by minimizing its upper bound $\sqrt{1-r^2}$. Hence, a weak learner can be designed to maximize $|r|$ for a sensible model h_t.

$$\begin{aligned}
\max_{\substack{\mathbf{U}_t = \boldsymbol{u_t}\boldsymbol{u_t}' \\ ||\boldsymbol{u_t}|| = 1}} & \ |\sum_{(i,j,k)\in\mathcal{T}} D_t((i,j,k))\left(h_t(\boldsymbol{x_i}, \boldsymbol{x_k}) - h_t(\boldsymbol{x_i}, \boldsymbol{x_j})\right)| \\
\text{subject to} & \ \ h_t(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})'\mathbf{U}_t(\boldsymbol{x} - \boldsymbol{y})
\end{aligned} \tag{9}$$

Using simple matrix algebraic operations, Eq. (9) can be rewritten as

$$|\boldsymbol{u_t}'\left[\sum_{(i,j,k)\in\mathcal{T}} D_t((i,j,k))\left((\boldsymbol{x_i} - \boldsymbol{x_k})(\boldsymbol{x_i} - \boldsymbol{x_k})' - (\boldsymbol{x_i} - \boldsymbol{x_j})(\boldsymbol{x_i} - \boldsymbol{x_j})'\right)\right]\boldsymbol{u_t}| \tag{10}$$

The problem of maximizing the objective (10) subject to a normalization constraint $||\boldsymbol{u_t}|| = 1$ has a closed-form solution: the optimal $\boldsymbol{u_t}$ is the eigenvector corresponding to the eigenvalue λ, with the largest absolute value, of the matrix

$$\sum_{(i,j,k)\in\mathcal{T}} D_t((i,j,k))\left((\boldsymbol{x_i} - \boldsymbol{x_k})(\boldsymbol{x_i} - \boldsymbol{x_k})' - (\boldsymbol{x_i} - \boldsymbol{x_j})(\boldsymbol{x_i} - \boldsymbol{x_j})'\right). \tag{11}$$

Let $a = \max\{||\boldsymbol{x} - \boldsymbol{y}|| \mid \boldsymbol{x} \neq \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{y} \in X\}$ which is a constant for a given set of data.

$$0 \leq h_t(\boldsymbol{x}, \boldsymbol{y}) = \frac{(\boldsymbol{x} - \boldsymbol{y})'\boldsymbol{u_t}\boldsymbol{u_t}'(\boldsymbol{x} - \boldsymbol{y})}{a^2} \leq \frac{||\boldsymbol{x} - \boldsymbol{y}||^2||\boldsymbol{u_t}||^2}{a^2} \leq 1 \tag{12}$$

In testing, if $d(\boldsymbol{x_i}, \boldsymbol{x_k}) < \delta$ and δ is a distance threshold which can be statistically calibrated as shown later, we will claim $\boldsymbol{x_k}$ is a match of $\boldsymbol{x_i}$ and the confidence p_i for $\boldsymbol{x_i}$ to be detected as positive (assuming only true positives have matches) is inverse to the minimum distance, i.e., $p_i = 1/d(\boldsymbol{x_i}, \boldsymbol{x_k})$. An matched polyp pair example is shown in Fig. 2.
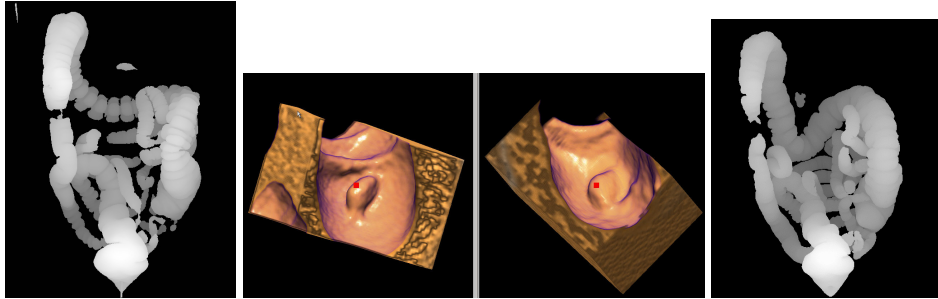
Fig. 2. An illustrative example of matched polyp pair, in a collapsed colon with background topological noises (e.g., small intestine) and large deformations in transverse and ascending colon sections.

3 Experimental Results

For each polyp x_i in a given view (prone or supine), the goal of polyp retrieval is to find its corresponding counterpart in the other view. Using $d(x_i, x_j)$ in Eq. (4) with learned M, we sort its k nearest neighbors and check whether there is a true match within k to trigger a hit. The retrieval rate is defined as the number of polyps retrieved divided by the total number of query polyps. In case of multiple instances, any true instance appearing in the top k neighbors will count the polyp as retrieved at k. The evaluation of retrieval rate versus the number of nearest neighbors, i.e., k is demonstrated in Fig. 3, for both training and testing datasets. Fig. 3 shows the superior performance of metric learning methods (via fusing local polyp appearance features), compared with the centerline geodesic distance based retrieval, similar to [3,4]. Not all polyps are retrievable because a small portion of polyps ($\leq 7\%$) only appeared in one view. For centerline based schemes, $> 40\%$ polyps are non-retrievable or can not be directly handled by [1–4], mainly due to collapsed colon segmentation in at least one prone or supine volume of 31% training, or 36% testing cases. [13]reports that $\sim 40\%$ volumes have collapsed colon segments in a clinical dataset. Note that, by normalizing against the polyp retrieval upper bounds ($55 \sim 59\%$ for geometric and $93 \sim 94\%$ for metric learning) respectively, in Fig. 3, i.e., assuming all polyps are matchable, our *local features + metric distance learning* approach still clearly has more appealing performances, as 85% versus 62% in training; 80% versus 57% in testing when $k = 1$.

We also evaluate our MatrixBoost algorithm against other metric learning methods (Mahalanobis [10], PSDBoost [11], ITML [14], BoostMetric [11] and COP [9]), using the same selected feature set by MRMR. MatrixBoost dominates the retrieval rate at the full range of k, with a larger margin in testing dataset. For example, when $k = 2$, the testing retrieval rate of our method is 80.51%, while the best result of all other techniques is 73.73%. High polyp match/retrieval rates under smaller numbers of k, can greatly facilitate the workflow for radiologists to effectively and efficiently match the polyp findings in prone-supine CTC
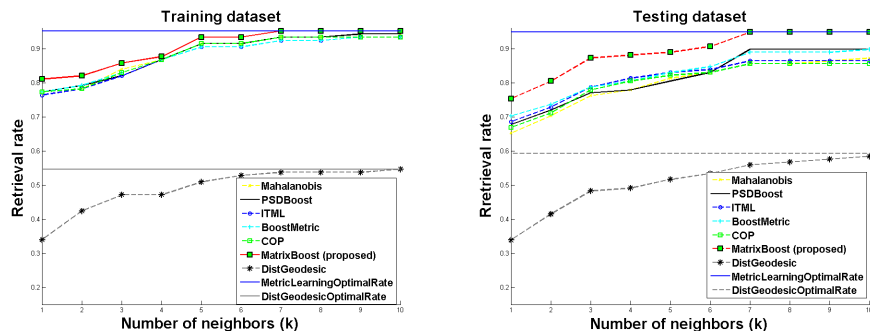
Fig. 3. Retrieval rates versus the number of nearest neighbors on the training (**Left**) and testing (**Right**) datasets.

views. Moreover, MatrixBoost permits faster convergence to the upper bounds of polyp retrieval rate at $k = 7$ in both training and testing. Lastly, the polyp retrieval performance can be presented using Precision-Recall curves that show the balance of retrieval accuracy (i.e., the percetage of false pairs retrieved) versus recall, in Fig. 4. For future work, we plan to generalize our method for temporal matching of lesion findings captured at different times.

4   Discussion

We proposed an effective and high performance polyp prone-supine view matching method, based on local polyp classification feature learning (via feature selection and metric learning). Our approach is evaluated on at least one order-of-magnitude larger, multiple hospitals dataset than previous work [1–4]. It can automatically and robustly handle highly varying colon segmentations from hundreds of patient cases, without any manual editing or preprocessing overhead. In summary, our method greatly advances the state-of-the-arts for polyp matching, and makes it more technically feasible for clinical practice.

References

1. Lai, Z., Hu, J., C.Liu, Taimour, V., Pai, D., Zhu, J., Xu, J., Hua, J.: Intra-patient Supine-Prone Colon Registration in CT Colonography Using Shape Spectrum. MICCAI (2010) 332–9
2. Roth, H., McClelland, J., Modat, M., Boone, D., Hu, M., Ourselin, S., Slabaugh, G., Halligan, S., Hawkes, D.: Establishing Spatial Correspondence between the Inner Colon Surfaces from Prone and Supine CT Colonography. MICCAI (2010)
3. Wang, S., Yao, J., Liu, J., Petrick, N., Periaswamy, S., Summers, R.: Registration of Prone and Supine CT Colonography Scans Using Correlation Optimized Warping and Canonical Correlation Analysis. Medical Physics **36** (2009) 5595–5603
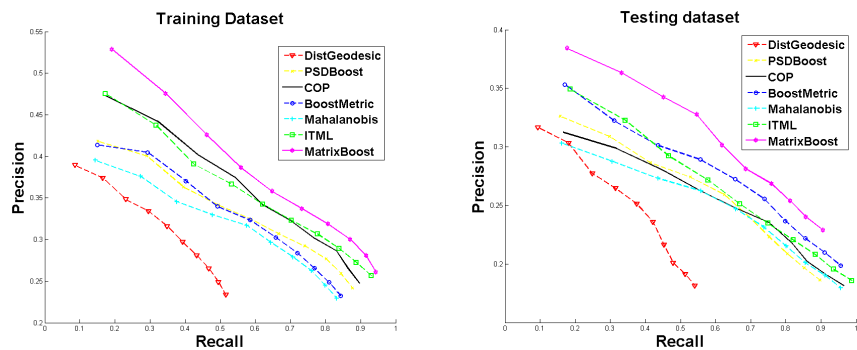
Fig. 4. The comparison of polyp retrieval Precision-Recall curves, using different metric learning methods and centerline Geodesic distance approach, on training (**Left**) and testing (**Right**) datasets. MatrixBoost method shows superior performance than others.

4. Li, P., Napel, S., Acar, B., Paik, D., Jeffery, R., Beaulieu, C.: Registration of Central Paths and Colonic Polyps Between Supine and Prone Scans in Computed Tomography Tolonography: Pilot Study. Medical Physics (2004) 31:2912–2923

5. Mang, T., Bogoni, L., Salganicoff, M., Wolf, M.: CT Colonography: Retrospective Evaluation of the Performance of Computer-aided Detection of Colonic Polyps in Tagged and Untagged Preparation. European Congress of Radiology (2010)

6. Slabaugh, G., Yang, X., Ye, X., Boyes, R., Beddoe., G.: A Robust and Fast System for CTC Computer-Aided Detection of Colorectal Lesions. Algorithms **3**(1) (2010)

7. van Ravesteijn, V., van Wijk, C., Vos, F., Truyen, R., Peters, J., Stoker, J., van Vliet, L.: Computer Aided Detection of Polyps in CT Colonography using Logistic Regression. IEEE Trans. on Med. Imag. (2010) 29:120–31

8. Peng, H., Long, F., Ding, C.: Feature Selection Based on Mutual Information: Criteria of Max-dependency, Max-relevance, and Min-redundancy. IEEE TPAMI (2005) 1226–1238

9. Xing, E., Ng, A., Jordan, M., Russell, S.: Distance Metric Learning, with Applicationto Clustering with Side-information. NIPS **15** (2002) 505–512

10. Bar-Hillel, A., Hertz, T., Shental, N., Weinshall, D.: Learning a Mahalanobis Metric from Equivalence Constraints. J. Mach. Learn. Res. **6** (2005) 937–965

11. Shen, C., Kim, J., Wang, L., van den Hengel, A.: Positive Semidefinite Metric Learning with Boosting. NIPS (2009) 1651–1659

12. Freund, Y., Schapire, R.: A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting. Comput. Learn. Theory (1995)

13. Lu, L., Wolf, M., Liang, J., Bi, J., Sagalnicoff, M.: A Two-Level Approach Towards Semantic Colon Segmentation: Removing Extra-Colonic Findings. MICCAI **5762** (2009) 1009–1016

14. Davis, J., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-Theoretic Metric Learning. ICML (2007) 209–216