A Meta Registration Framework for Lesion Matching^{*}

Sharmishtaa Seshamani¹, Purnima Rajan¹, Rajesh Kumar¹, Hani Girgis¹, Themos Dassopoulos³, Gerard Mullin², and Gregory Hager¹

¹ Department of Computer Science, Johns Hopkins University, Baltimore, MD sharmi@jhu.edu

² Johns Hopkins Hospital, Baltimore, MD
³ Washington University, School of Medicine St. Louis, MO

Abstract. A variety of pixel and feature based methods have been proposed for registering multiple views of anatomy visible in studies obtained using diagnostic, minimally invasive imaging. A given registration method may outperform another depending on anatomical variations, imaging conditions, and imaging sensor performance, and it is often difficult a priori to determine the best registration method for a particular application. To address this problem, we propose a registration framework that pools the results of multiple registration methods using a decision function for validating registrations. We refer to this as *meta registration*. We demonstrate that our framework outperforms several individual registration methods on the task of registering multiple views of Crohn's disease lesions sampled from a Capsule Endoscopy (CE) study database. We also report on preliminary work on assessing the quality of registrations obtained, and the possibility of using such assessment in the registration framework.

1 Introduction

Minimally invasive diagnostic imaging methods such as flexible endoscopy, and wireless capsule endoscopy (CE) often present multiple views of the same anatomy. Duplication issues are particularly severe in the case of CE, where peristaltic propulsion may lead to duplicate information over several consecutives frames, and also several minutes apart. This may be difficult to detect, since each individual image captures only a small portion of anatomical surface due to limited working distance of these devices, providing relatively little spatial context. Given the relatively large anatomical surfaces (e.g. the Gastrointestinal tract (GI)) to be inspected, it is important to identify duplicate information as well as to present all available views of anatomical and disease views to the clinician for improving consistency, efficiency and accuracy of diagnosis and assessment.

^{*} Supported in part by National Institutes of Health with Grant 5R21EB008227-02 and Johns Hopkins University internal funds.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 582–589, 2009.

In particular, we are interested in automatically assessing the severity of Crohn's disease, an inflammatory condition of the GI tract using CE images (Figure 1). The aggressiveness of the course of mucosal healing treatment for Crohn's disease depends upon the severity assessment of unique lesions in the CE images. We aim to identify and assess multiple views of a selected lesion, to reduce redundant information and to improve the consistency and accuracy of disease assessment.

In the literature, lesion finding has been commonly formulated as a detection problem [1] where a classifier is trained to learn the visual properties of the chosen object category (i.e. lesions). This process typically requires



Fig. 1. A mild Crohn's lesion in CE. CE color Images are typically 576x576 pixels.

feature extraction to generate a low dimensional representation of image content, followed by classifier training to distinguish the desired object model(s) [2]. For CE, appearance modelling has been used for blood detection [3,4], topographic segmentation[5] and lesion classification [6].

Generic detection is different from *matching* an instance of an object to another. Therefore, we consider detection of repetitive lesions as a registration and matching problem. A registration method evaluates an objective function or similarity metric to determine a location in the target image (a second view) where a reference view (i.e. a lesion) occurs. Once a potential registration is computed, a decision function must be applied to determine the validity of the match. Our goal is to develop a trained statistical classifier that makes a decision based on the quality of a match between two regions of interest (ROIs) or views of the same lesion, rather than the appearance of the features representing an individual ROI.

Decision functions for image matching have traditionally been designed as thresholded classifiers based on one or more registration metrics. Although such a method ensures low false positive rates, the overall retrieval rate is bounded by the recall rate of the most sensitive metric. The work of Szeliski et al. [7] and Stewart et al. [8] are examples of such problem formulations. In many cases, a single, unique global threshold may not exist; therefore the determination of an adaptive threshold is a challenging problem. An integrated classifier that distinguishes registrations based on a feature representation populated by a wide range of metrics is likely to outperform such threshold based methods. Chen et al. [9] introduce a new feature vector that represents images using an extracted feature set. However, this approach still requires the same similarity metric across the entire feature set. By contrast, we present a framework that incorporates multiple registration algorithms, a generalizable feature representation for classification and a regression based ranking system to choose the highest quality match.

2 Meta Registration Framework

Given an ROI R_i in an image I_i , and a target image I_i , we define the registration function $T(R_i, I_j) \mapsto$ R_i , which generates an ROI match pair (R_i, R_j) . The objective function for this registration is based upon the invariance properties of image data to be matched. For example, histogram representations are invariant to rotation, whereas pixel based methods are generally not. Feature based methods are typically less affected by changes in illumination and scale. Due to



Fig. 2. Multiple independent registration methods and a robust classifier are used to determine the location of ROI R in image I

large variation in these invariance properties within endoscopic studies, a single registration method may not be adequate for retreiving matches in endoscopic data. We use multiple independent global registration methods (each more accurate in a different subset of the data) and a decision function to estimate a valid match. Let us denote the similarity metric relating visual properties of the two ROIs R_i and R_j as $d(R_i, R_j)$. Using a set of global registration functions $\mathcal{T} = T_i(R, I) : i = 1 \dots n$ and estimated matches $R'_1 \dots R'_n$ from each of these methods, the decision function D which determines which estimates are correct matches can be written as in equation 1. In the case where there are multiple correct match estimates, we use a ranking function to determine the best result. Figure 2 shows the information flow in the proposed method.

$$D(R_i, R'_j) = \begin{cases} 1, & \text{if } d(R_i, R'_j) < \gamma \\ -1 & \text{otherwise} \end{cases}$$
(1)

Decision Function Design. For an ROI R, we use the following notation for representing appearance features. We start with pixel based features. The intensity band of the image is denoted as R_I . The Jacobian of the image is denoted $R_J = [R_x, R_y]$ where R_x and R_y are the vectors of spatial derivatives at all image pixels. Condition numbers and the smallest Eigenvalues of the Jacobian are denoted as R_{JC} and R_{JE} respectively. The Laplacian of the image is denoted as R_{LAP} . We denote histogram based features as : R_{RGBH} , R_{WH} and R_{WCH} for RGB histograms, Gaussian weighted intensity histograms and Gaussian weighted color histograms respectively. We also consider MPEG-7 descriptors [10]: R_{EHD} (Edge Histogram), R_{Har} (Haralick Texture), R_{DCD} (Dominant Color) and R_{HTD} (Homogeneous Texture).

Given two images I_a and I_b where A is an ROI in I_A with center x and B is an ROI in I_B , we generate a feature vector for a pair of regions A and B

Metric Name	Formula
RMS (rms)	$\sqrt{\left(\frac{1}{n}\sum_{k}(A_{I}-B_{I})^{2}\right)}$
RMS Shuffle	$\sqrt{\left(\frac{1}{n}\sum_{k} shuffle(A_{I}, B_{I})\right)}$
Ratio of Condition Numbers	$min(A_{JC}, B_{JC})/max(A_{JC}, B_{JC})$
Ratio of Smallest Eigenvalues	$min(A_{JE}, B_{JE})/max(A_{JE}, B_{JE})$
Laplacian Shuffle Distance	$shuffle(A_{LAP}, B_{LAP})$
Weighted Histogram Bhattacharya Distance	$sqrt(A_{WH}.B_{WH})$
RGB Histogram Bhattacharya Distance	$sqrt(A_{RGB}.B_{RGB})$
Edge Histogram Manhattan Distance	$\sum (A_{EHD} - B_{EHD})$
Haralick Descriptor Canberra Distance	$\sum \frac{ A_{Har} - B_{Har} }{ A_{Har} + B_{Har} }$
HTD Shuffle Distance	$shuffle(A_{HTD}, B_{HTD})$
Forward Backward check	$ x - T(I_b, I_a, T(I_a, I_b, x)) $

Table 1. Metrics Table

populated with the metrics shown in Table 1. The decision function can then be trained to distinguish between correct and incorrect matches using any standard classification method. We use Support Vector Machines (SVM) [11] in our experiments.

The Ranking Function: We treat the registration selection as an ordinal regression [12] problem. Given a feature set corresponding to correctly classified registrations, $F = \{f_1, ..., f_N\}$ and a set of N distances from the true registrations, we can form a set of preference relationships between the elements of F. Let us define the set of preference pairs P as, $P = \{(x, y) | f_x \prec f_y\}$. Our goal is to compute a continuous real-valued ranking function K such that, $f_x \prec f_y \in P \Longrightarrow K(f_x) < K(f_y)$. A preference pair $(x, y) \in P$ can be thought of as a pair of training examples for a standard binary classifier. We train a binary classifier C such that,

$$C(F_x, F_y) = \begin{cases} 0, & if(x, y) \in P \\ 1 & otherwise \end{cases}$$

and $C(F_y, F_x) = 1 - C(F_x, F_y)$. Given such a classifier, the rank can be computed as, $K(F) = \sum_{i=1}^{n} C(F, F_i)/n$ where K is the fraction of the training set that are less preferred to F based on the classifier. Thus K orders F relative to the training set. Let f_x represent the metrics or features of registration and $f_{i,j}$ represent the vector concatenation of f_i and f_j . The training set, $Train = \{ < f_{i,j}, 0 >, < f_{j,i}, 1 > | (i,j) \in P \}$ is used to train an SVM. For classification, we pair each vector in the test set with all the vectors in the training set and use the empirical order statistics K(F) described above for enumerating the rank.

3 Training Data

Given an ROI R and a set of CE images $\mathcal{I} = \{I_i : i = 1...N\}$, the task is to build a dataset of pairs of images representing correct and incorrect matches

of a global registration. We first compute the correct location of the center of the same ROI in \mathcal{I} through manual selection followed by a local optimization. We denote this set of locations as $\mathcal{X} = \{X_i : i = 1...N\}$. Next, we select any global registration method T and apply it between R and each image in the set \mathcal{I} to generate a set of estimated ROI center locations $\mathcal{X}' = \{X'_i : i = 1...N\}$ and pairs $\mathcal{R} = \{R, R_i : i = 1...N\}$. The pairs are designated a class label y (correct or incorrect matches) by thresholding on the Euclidean distance (in pixels) between X_i and X'_i . We refer to this as the ground truth distance. The training set \mathcal{T} contains all registered pairs and their associated class labels.

4 Experiments

Our CE study repository contains studies acquired with an approved Institutional Review Board (IRB) protocol. A CE study may contain up to 8 hours of data acquired at 2 images per second. The CE study database contains selected annotated images containing Crohn's Disease (CD) lesions manually selected by our clinical collaborators. These provide the ROIs we use for our experiments. A lesion may occur in several neighboring images, and these selected frames form a lesion set. Figure 3 shows an example of a lesion set. In these experiments, 150x150 pixel ROIs were selected. Various lesion sets contain between 2 and 25 image frames. Registration pairs were then generated for every ROI in the lesion set, totaling 266 registration pairs.

We used the following five standard techniques for 2D registration. These include SIFT feature matching, mutual information optimization, weighted histograms (grayscale and color) and template matching. For each of these methods, we performed a registration to estimate a registered location, resulting in a total of 1330 estimates (5 registration methods per ROI-image pair). The ground truth for these estimates was determined by using a threshold of 25 pixels on the Euclidean distance described in Section 3. The dataset contained 581 correct (positive examples) and 749 incorrect (negative examples) registrations.

For every registration estimate, we computed the registered ROI for the training pair. The feature vector respresenting this registration estimate was then computed as described in Section 2. We then trained the decision function using all registration pairs in the dataset. The performance of this integrated classifier was evaluated using a 10-fold cross-validation. Figure 5(a) shows the performance of our classifier, including comparison with the ROC curves of individual



Fig. 3. A Lesion Set: a set of neighboring CE images with the same lesion

	Iter1	Iter2	Iter3	Iter4	Iter5	Iter6	Iter7	Iter8
No: of pairs	300	600	900	1200	1500	1800	2100	2400
Train mis-classification rate	0.001	0.014	0.016	0.015	0.018	0.017	0.017	0.017
Train pair mismatch rate	0.16	0.18	0.17	0.16	0.16	0.16	0.16	0.15
Test pair mismatch rate	0.32	0.38	0.32	0.26	0.38	0.32	0.35	0.27
Test rank mean	0.53	0.69	0.55	00.35	0.69	0.55	0.61	0.44
Test rank std dev	0.14	0.15	0.20	0.28	0.19	0.23	0.21	0.29

Table 2. 8 Iterations of Ranking Experiment



Fig. 4. Estimated Ranks for training (left) and test(right) datasets

metrics used for feature generation. The true positive rate was 96 percent and the false negative rate was 8 percent.

For n registrations, a total of n choose 2 preference pairs can be generated. We used a subset of this data as the input to the ranking model. Features used to generate a training pair include the difference between EHDs and the difference between the DCDs. Training was initiated with a random selection of n = 300. This estimate was then improved by iteration and addition of preference pairs at every step. At each iteration, a 10-fold cross validation was performed where the dataset was divided into appropriately sized training and test sets. Training was then conducted using an SVM model with a radial basis kernel.

Once this classifier was trained, preference relationships were predicted by classifying vectors paired with all training vectors. Relative ranks within each set were determined and pair mismatch rates were calculated. A mismatch is any pair of registrations where $K(F_x) > K(F_y)$ and $F_x \prec F_y$ or $K(F_x) < K(F_y)$ and $F_x \succ F_y$. The training mis-classification rate is the percentage of contradictions between the true and predicted preference relationships in the training set. Table 2 shows the rank metrics for each iteration. Figure 4 shows the sorted ranking function of training and test set registrations at the end of 8 iterations.

The registration framework was applied to all 266 image pairs. For each pair, all five registration methods were applied to estimate matching ROIs. The first row of Table 3 shows the number of correct registrations evaluated using the ground truth distance. Features were extracted for all registrations and the integrated classifier trained above was applied. A leave one out cross-validation



Fig. 5. Meta Registration Results. LEFT: ROC Curves of all metrics used individually overlaid with the integrated classifier (Green X), RIGHT: Percentage true positives for cross validation for template matching(red), SIFT (green), Mutual Information (blue), Intensity Weighted Histogram matching (black) Color Weighted Histogram matching (magenta) and meta registration (cyan).

Table 3. Meta Registration Results. The first row shows the number of correct matches qualified by a manual validation. The second row shows the number of correct matches qualified by the automated integrated classifier. The third row shows the number of automatically qualified matches that match the manual (ground truth) validation.

Type	Template	Sift	Mutual	Intensity	HSV	Meta
	Matching		Info	Weighted	Weighted	Registration
				Histogram	Histogram	
Ground	165	122	54	111	129	266
Truth						
Classifier	129	62	25	75	77	188
True	106	59	10	46	47	188
Positives						

was performed for each ROI-image pair. The second row of Table 3 shows the number of matches that the classifier validated as correct. Finally, the last row shows the number of true positives which is the number of correctly classified matches that are consistent with the ground truth classification. The last column in the table shows the performance of the meta registration. The new framework retreived a larger number of registrations than any single registration method. We performed a range of n-fold validations on the same dataset for n ranging from 2-266 (where n = 2 divides the set into two halves and n = 266 is the leave one out validation). Figure 5(b) shows the percentage of true positives retreived (which is the ratio of true positives of the meta registration method and the integrated classifier (top curve in cyan). The meta registration was able to retrieve 70 percent of matches where the best registration method only retreived 39 percent.

5 Discussion

We have proposed a meta registration framework for matching of lesions in capsule endoscopic video. This general approach incorporates multiple independent optimizers and an integrated classifier combined with a trained ranker to select the best correct match from all registration results. Our results show the integrated method outperforms the use of any single method for ROI matching and retreival in CE imagery. In future work, we plan to extend this to matching ROIs without explicit application of global optimizers.

References

- Taylor, C.J., Cooper, D.H., Graham, J.: Training models of shape from sets of examples. In: Proc. of the British Machine Vision Conference, pp. 9–18. Springer, Heidelberg (1992)
- 2. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 511–518 (2001)
- Jung, Y.S., Kim, Y.H., Lee, D.H., Kim, J.H.: Active blood detection in a high resolution capsule endoscopy using color spectrum transformation. In: Proc. of the International Conference on BioMedical Engineering and Informatics, Washington, DC, USA, pp. 859–862. IEEE Computer Society, Los Alamitos (2008)
- 4. Li, B., Meng, M.Q.H.: Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments. Computers in Biology and Medicine 39(2), 141–147 (2009)
- Cunha, J.P., Coimbra, M.T., Campos, P., Soares, J.M.: Automated topographic segmentation and transit time estimation in endoscopic capsule exams. IEEE Transactions on Medical Imaging 27(1), 19–27 (2008)
- Bejakovic, S., Kumar, R., Dassopoulos, T., Mullin, G., Hager, G.: Analysis of crohns disease lesions in capsule endoscopy images. In: Proc. of the International Conference on Robotics and Automation, pp. 2793–2798 (2009)
- Szeliski, R.: Prediction error as a quality metric for motion and stereo. In: Proc. of the International Conference on Computer Vision, Washington, DC, USA, p. 781. IEEE Computer Society, Los Alamitos (1999)
- Yang, G., Stewart, C., Sofka, M., Tsai, C.L.: Registration of challenging image pairs: Initialization, estimation, and decision. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(11), 1973–1989 (2007)
- Chen, X., Cham, T.J.: Learning feature distance measures for image correspondences. In: Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, pp. 560–567. IEEE Computer Society, Los Alamitos (2005)
- Coimbra, M., Cunha, J.: Mpeg-7 visual descriptors contributions for automated feature extraction in capsule endoscopy. IEEE Transactions on Circuits and Systems for Video Technology 16(5), 628–637 (2006)
- Vapnik, V.N.: The nature of statistical learning theory. Springer-Verlag New York, Inc., New York (1995)
- Herbrich, R., Graepel, T., Obermayer, K.: Support vector learning for ordinal regression (1999)