Robust Registration of Statistical Shape Models for Unsupervised Pathology Annotation

Dana Rahbani(✉), Andreas Morel-Forster, Dennis Madsen, Marcel Lüthi, and Thomas Vetter

Department of Mathematics and Computer Science, University of Basel, Basel, Switzerland
dana.rahbani@unibas.ch

Abstract. We present a method to automatically label pathologies in volumetric medical data. Our solution makes use of a healthy statistical shape model to label pathologies in novel targets during model fitting. We achieve this using an EM algorithm: the E-step classifies surface points into pathological or healthy classes based on outliers in predicted correspondences, while the M-step performs probabilistic fitting of the statistical shape model to the healthy region. Our method is independent of pathology type or target anatomy, and can therefore be used for labeling different types of data. The method is able to detect pathologies with higher accuracy than standard robust detection algorithms, which we show using true positive rate and F1 scores. Furthermore, the method provides an estimate of the uncertainty of the synthesized label. The detection also directly improves surface reconstruction results, as shown by a decrease in the average and Hausdorff distances to ground truth. The method can be used for automated diagnosis or as a pre-processing step to accurately label large amounts of images.

Keywords: Statistical shape model · Label synthesis · Outlier detection · Robust non-rigid registration · EM algorithm

1 Introduction

Automatic labeling of biomedical data remains a necessity, whether for diagnosis in health-care or image annotation in datasets. This is especially the case for pathology labeling in volumetric data such as CT or MR images, where there are time, cost, and error constraints on getting expert labels. One main challenge for automatic labeling is the extreme variation which can be seen across pathologies. This limits the ability to generalize labeling algorithms across imaging domains or even within the same pathology type.

Algorithms that rely on generative methods assume there is an underlying model which can be used to analyze an image. One example is a statistical shape model (SSM), which is a linear parametric model of shape variation.

L. Zhou et al. (Eds.): LABELS 2019/HAL-MICCAI 2019/CuRIOUS 2019, LNCS 11851, pp. 13–21, 2019.
https://doi.org/10.1007/978-3-030-33642-4_2

SSMs can generalize well to represent novel instances within the same shape family. For example, an SSM built from healthy mandibles can be used to extract information about a novel mandible extracted from CT, such as location of teeth. However, the extreme variation problem also prevents the direct application of SSMs to pathological data, mainly because of unavailable correspondences needed for model building and fitting. Solutions usually involve disease-specific models [19] or handcrafted pathology features [12], but this is not always possible given limited data and intra-disease pathology variations.

We show how SSMs built from healthy anatomies can be exploited to perform pathology labeling in novel images. We treat pathology labeling as an outlier detection step in our proposed robust non-rigid registration algorithm. Outliers are all SSM points without a corresponding point in the target and vice versa. Our work extends combined fitting and segmentation with the EM-algorithm [2,5] to outlier detection on surfaces. We avoid pathology-specific modeling of features by introducing a probabilistic metric which does not depend on imaging modality or pathology type. The metric evaluates the target reconstruction and learns to perform unsupervised classification of individual data points into shape or pathology. The metric we implement is a probabilistic extension of a double-projection distance used in the iterative closest points (ICP) algorithm [1,16], explained in detail in Sect. 3.1. Our main contributions are:

1. an unsupervised-learning and probabilistic approach to label surfaces extracted from biomedical images as healthy or pathological
2. a robust registration algorithm for fitting SSMs to pathological data

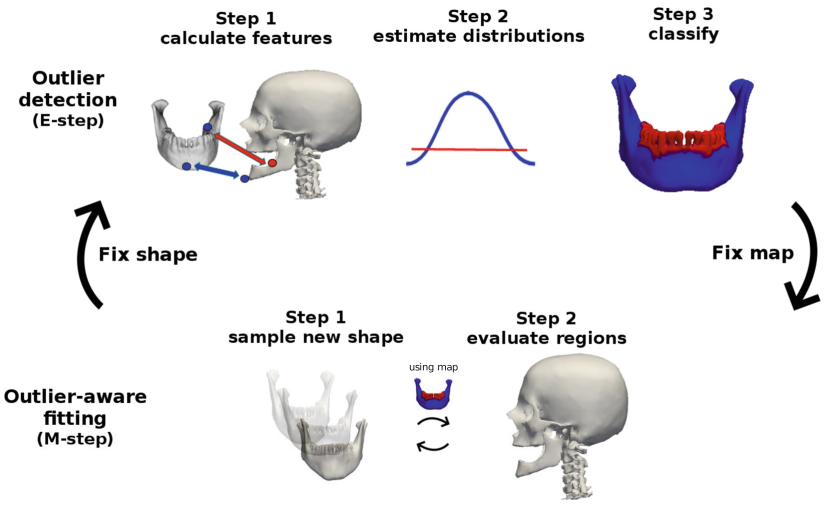

Fig. 1. Proposed pipeline for label synthesis and reconstruction. The input is an unlabeled pathological target surface. The outputs are the reconstruction, the label map and the estimated distributions. They are obtained by iterating between outlier detection (E-step) and outlier-aware fitting (M-step). The label map splits the reference topology into: healthy-region to be used in SSM fitting (blue) and outlier-region to be ignored by the SSM (red). (Color figure online)

2 Related Work

Given a dataset, the goal of outlier detection is to distinguish extreme values that are statistically relevant from those due to measurement errors [7]. A method can do so robustly if it has a high "breakdown point", a value which describes the number of outliers that can be present in a dataset before an algorithm fails [13]. For SSMs, outliers are model points which do not match their counterparts in a novel target. Approaches have been implemented to handle outliers:

Registration. In the trimmed ICP method (TrICP) [1], the robust Least Trimmed Squares algorithm is used in surface alignment. Other approaches for rigid registration of 3D surfaces with outliers make use of surface feature descriptors to match regions in correspondence [6]. However, we aim at non-rigid registration with missing or added data, which remains a difficult challenge. Non-rigid registration of point sets with outliers has been addressed in an extension to robust point matching [3] and in coherent point drift [15]. Instead of enforcing regularization on the allowed deformation fields as they do, we obtain deformation likelihoods directly from the SSM shape prior. In addition, outliers in our case include highly unlikely points under the shape prior instead of only missing or additional points in the set.

SSM-Based Approaches. Outlier detection for SSMs is a pre-processing step for building or fitting. Semantic patches are often introduced to narrow down the PCA space or reference topology [9,21]. SSMs have also been used for pathology segmentation from fitting errors [4]. Our method does not rely on a manual segmentation of the reference topology. We use fitting failures as in the second approach but go further by accounting for uncertainty in correspondences before pathology detection and improving reconstruction results.

Other Generative Approaches. Part-based models (PBMs) split SSMs into parts with a binary occurrence parameter [20], while Gaussian process morphable models (GPMMs) of shape and intensity [11,17] account for pathologies with local deformation kernels [10]. Recently, reconstruction errors from generative adversarial networks (GANs) have been used for pathology detection [18]. Our approach does not require local definitions of pathologies as PBMs or GPMMs do, nor does it depend on classification thresholds as GANs do.

3 Method

We extend the standard SSM fitting formulation with an additional segmentation of outliers. The segmentation takes the form of binary labels: one label for every point on the reference SSM topology. The labels separate two regions: an outlier-region and a healthy-region. SSM fitting is then restricted to the healthy-region. We formulate the detection and reconstruction steps together as a maximum a posteriori (MAP) estimation problem. The goal is to find the SSM shape and pose parameters $\boldsymbol{\theta}$ and the point-level label map $\boldsymbol{z}$ that maximize the posterior distribution function given a target surface Γ:

$$P(\boldsymbol{\theta}, \boldsymbol{z} \mid \Gamma) \propto L(\Gamma \mid \boldsymbol{\theta}, \boldsymbol{z}) P(\boldsymbol{\theta}, \boldsymbol{z}) \tag{1}$$

The likelihood evaluates the similarity of the SSM reconstruction $M(\boldsymbol{\theta})$ to the target Γ given a specific combination of $\boldsymbol{\theta}$ and $\boldsymbol{z}$, formulated as follows:

$$L(\Gamma \mid \boldsymbol{\theta}, \boldsymbol{z}) = \prod_{i \epsilon n} l_h(M(\boldsymbol{\theta})_i, \Gamma_i)^{z_i} l_o(M(\boldsymbol{\theta})_i, \Gamma_i)^{1-z_i} \tag{2}$$

The likelihood factors over the n reference topology points, since they are assumed to be independent. Every point i is evaluated by either the healthy-region distribution l_h or the outlier-region distribution l_o. The point label z_i indicates which of the two distributions should be used for the point i: if $z_i = 1$, then l_h is used, else $z_i = 0$ and l_o is used. The Euclidean distance is used to compare point i on the model surface $M(\boldsymbol{\theta})_i$ and its corresponding point on the target Γ_i.

Starting with a surface, the shape parameters $\boldsymbol{\theta}$, label map $\boldsymbol{z}$, and distributions l_h and l_o are unknown, making optimization intractable. We use an EM algorithm to solve this problem. In the E-step, we fix $\boldsymbol{\theta}$, learn the distributions l_h and l_o, then infer $\boldsymbol{z}$. In the M-step, we fix $\boldsymbol{z}$ and infer $\boldsymbol{\theta}$. The novel segmentation algorithm and the reconstruction strategy are presented in this section. Details on how to build SSMs can be found in Sect. 3 of [8] and Sect. 2 of [14].

3.1 Outlier Detection: Inferring the Label Map $\boldsymbol{z}$

We want to infer the binary label map $\boldsymbol{z}$ defined on the domain of the SSM reference topology. Each of the n points has a label for one of two classes: healthy-region or outlier-region. In our examples, we consider the mandible with teeth as the healthy shape. Outliers could be holes from missing teeth, shape deformations from injuries or surgery, or artifacts.

Starting from Eq. 1, we fix the values of $\boldsymbol{\theta}$. We can then infer the labels which give the MAP solution. The unknown variables in the likelihood function are the distributions and the label map, both of which depend on the accuracy of the corresponding pairs. We propose a probabilistic interpretation of the distances between corresponding points to account for correspondence uncertainties, accomplished in the three steps below.

Determine Correspondences. A simple double-projection method proposed in an ICP-based alignment [16] is used. For every point of the SSM reference topology, we first find its closest point on the target, then project from this point back to the SSM. The output is a set of bidirectional correspondence pairs. Incorrect pairs are expected, not only because of this rough correspondence estimation approach, but also because in the beginning the fixed parameters $\boldsymbol{\theta}$ are far from the MAP solution.

Estimate Distributions. If the ground truth $\boldsymbol{\theta}$ and perfect correspondences were used, then we could expect l_h to be a univariate Gaussian with zero mean and standard deviation from reconstruction noise. However, since neither are

available for pathological targets, l_h is a Gaussian distribution learned from the current double-projection distances. We assume a uniform distribution for l_o since the method is independent of pathology type. The likelihood is fixed at the value three standard deviations away from the mean of a healthy distribution, which we learn by fitting to 100 healthy shapes sampled from the SSM.

Infer Label Map. A point is considered an outlier if its double-projection distance has a higher likelihood of belonging to the outlier-region distribution than to the healthy-region distribution. We infer every z_i by choosing the label corresponding to the larger likelihood value. This is equivalent to maximizing the likelihood function in Eq. 2 with respect to $\boldsymbol{z}$.

3.2 Outlier-Aware SSM Fitting: Inferring the SSM Parameters $\boldsymbol{\theta}$

To fit the SSM to the target, we need to maximize Eq. 1 with respect to the SSM parameters $\boldsymbol{\theta}$. To do so, we first fix the values of $\boldsymbol{z}$ obtained from the E-step. This leaves $\boldsymbol{\theta}$ as the only remaining unknown in the likelihood Eq. 2. The prior on the shape parameters $P(\boldsymbol{\theta})$ is provided by the SSM. With this information, we can find $\boldsymbol{\theta}$ by maximizing Eq. 1 using the approach in [14], which approximates the posterior distribution then takes the MAP solution as the best reconstruction.

4 Evaluation

We evaluate our method on a mandible SSM built from eight surfaces extracted from healthy CT scans. The meshes are registered with Gaussian process morphable models [11]. The healthy mandible SSM is built from the registered meshes using PCA. We generate pathological targets with known ground truth label maps and SSM parameters. For this, we sample shapes from the healthy model and deform or clip away parts of their surfaces. This ensures that the model is able to represent the target shape without pathologies and allows us to evaluate the effect of the labeling algorithm, instead of the SSM generalization ability, on the registration. This results in 25 test cases, two of which are shown in Fig. 2. We then apply the method on real data, with target surfaces extracted from CT images. Visualization is performed using *Scalismo*[1].

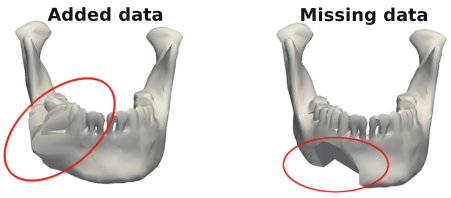

Fig. 2. Mandible shape with example pathologies circled in red. (Color figure online)

[1] https://scalismo.org.

Model fitting is performed for shape and pose parameters after rigid landmark alignment. Three approaches are compared: (1) standard SSM, which only performs reconstruction without segmentation, (2) robust SSM, which diminishes the effect of outliers based on the Huber Loss function with a 5 mm threshold, and (3) outlier-aware SSM. The threshold for the robust SSM is manually assigned because the learned threshold proved to be too large. For every outlier detection step, 1,000 iterations are taken in the fitting step. The entire process is repeated ten times as this empirically showed convergence by a constant healthy-region distribution. For comparison, 10,000 iterations are performed for the standard and robust SSMs.

Labeling Evaluation. Label maps are evaluated by the true positive rate (TPR) and the F1 scores compared to the ground truth labels. The TPR is the ratio of the number of true detected outlier points to the number of ground truth outlier points. It is used to give an idea of how much of the outlier region is detected. The F1 score is the harmonic mean of TPR and precision, where precision is the ratio of true detected outlier points to all detected outlier points. The F1 score is used to evaluate binary classification tasks with class imbalance. Figure 3 shows examples of the label maps for a missing data case. TPR and F1 scores in Table 1 reveal that the outlier-aware SSM outperforms the robust SSM.

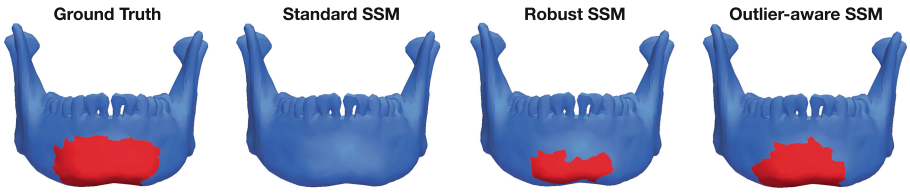

Fig. 3. Visual comparison of label maps projected onto the reference SSM topology: healthy-region (blue) and outlier-region (red). Pathology detection is not a feature of the standard SSM, which is why the entire topology is labeled as healthy. (Color figure online)

Reconstruction Evaluation. Reconstructions are evaluated by their Hausdorff and average distances to the ground truth healthy surface, seen in Table 1. There is a strong decrease in both distances when the outlier-aware SSM is used instead of the standard and robust SSMs. This is accredited to a closer reconstruction of the ground truth healthy surface in both the healthy and outlier regions.

Breakdown Point Evaluation. The breakdown point is defined as the fraction of points that can be outliers before the algorithm fails. We generate pathologies that cover an increasing fraction of the reference surface. The fitting degrades strongly if more than a third of the observed surface is pathological. Just before the breakdown point, we still reach an F1 score of 0.74 and an average distance of 1.08 mm.

Table 1. Labeling and fitting evaluations: mean values followed by standard deviations in parentheses. The mean values of the true positive rates (TPR) and F1 scores are best at 1, while those of the Hausdorff and Average distances (HD and AD) at 0 mm.

	Standard SSM (no labeling)	Robust SSM (thresholded labeling)	Outlier-aware SSM (probabilistic labeling)
TPR	–	0.39 (0.22)	**0.56 (0.22)**
F1	–	0.51 (0.27)	**0.68 (0.22)**
HD	4.48 (5.46)	6.07 (5.37)	**1.98 (0.89)**
AD	1.14 (0.51)	1.52 (0.49)	**0.88 (0.23)**

Label Uncertainty. We apply the algorithm to radius and mandible surfaces extracted from pathological CT images. Pathologies are regions of overgrowth for the radius and regions of missing teeth for the mandible, as seen in the targets in Fig. 4. We use a radius SSM built from 37 surfaces and mandible SSM from 8 surfaces as the models, all built from healthy CT images.

We use our proposed outlier-aware SSM to register and label the target surfaces and their reconstructions. Unlike the synthetic data case, surfaces extracted from real CT images by simple thresholding are not as clean. This can be seen for the mandible example in Fig. 4, where the cranium and the spine are parts of the input target surface. The proposed method can be applied on the surface with irrelevant data, without requiring user input other than the initial rigid alignment. We accomplish this by including both the model and target surfaces into the likelihood function, in Eq. 2. Using the distributions learned from step 2 of the E-step (Fig. 1) and the distances between the target and reconstruction, we can compute the uncertainty for the synthesized labels. The target and fitting

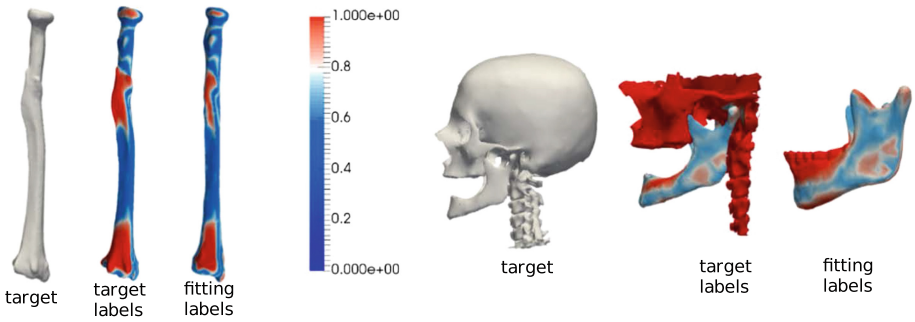

Fig. 4. Uncertainty in generated labels for target radius and skull surfaces, both extracted from pathological CT images. Pathology labels with high certainty are in red. Note that for the mandible example, we crop the target for computational purposes with a box from six landmarks. Given the remaining surface, the method is able to correctly locate irrelevant regions, as indicated by red labels on the cranium and spine. (Color figure online)

labels in Fig. 4 show the certainty levels for the generated pathology labels. Red indicates high certainty for the pathological label and blue for the healthy label.

5 Conclusion

Supervised pathology detection algorithms depend on expert labels or pathology thresholds. However, our proposed outlier-aware SSM is able to perform the detection given only a target surface without any further assumptions or user annotations. Pathology detection with our approach accomplishes higher true positive rates and F1 scores than classical robust statistics methods do. This results in a closer approximation of the ground truth healthy target, seen with reduction in the average distances, and also an uncertainty estimate on the synthesized labels. Our pipeline is non-specific to pathology type or imaging domain. This implies it can be used to point out regions of interest to clinicians or as a pre-processing step for training end-to-end classifiers. Future work will investigate other probabilistic metrics that can work alongside the distance-based one, as well as further testing of the method on current biomedical image segmentation challenges.

References

1. Chetverikov, D., Stepanov, D., Krsek, P.: Robust euclidean alignment of 3D point sets: the trimmed iterative closest point algorithm. Image Vis. Comput. **23**(3), 299–309 (2005)
2. Chitphakdithai, N., Duncan, J.S.: Non-rigid registration with missing correspondences in preoperative and postresection brain images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 367–374. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15705-9_45
3. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. Comput. Vis. Image Underst. **89**(2–3), 114–141 (2003)
4. Dufour, P.A., Abdillahi, H., Ceklic, L., Wolf-Schnurrbusch, U., Kowal, J.: Pathology hinting as the combination of automatic segmentation with a statistical shape model. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7512, pp. 599–606. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33454-2_74
5. Egger, B., et al.: Occlusion-aware 3D morphable models and an illumination prior for face image analysis. Int. J. Comput. Vis. **126**, 1269–1287 (2018)
6. Gelfand, N., Mitra, N.J., Guibas, L.J., Pottmann, H.: Robust global registration. In: Symposium on Geometry Processing, vol. 2, p. 5 (2005)
7. Grubbs, F.E.: Procedures for detecting outlying observations in samples. Technometrics **11**(1), 1–21 (1969)
8. Heimann, T., Meinzer, H.P.: Statistical shape models for 3D medical image segmentation: a review. Med. Image Anal. **13**(4), 543–563 (2009)
9. Lüthi, M., Albrecht, T., Vetter, T.: Building shape models from lousy data. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 1–8. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04271-3_1

10. Lüthi, M., Forster, A., Gerig, T., Vetter, T.: Shape modeling using Gaussian process morphable models. In: Statistical Shape and Deformation Analysis, pp. 165–191. Elsevier (2017)
11. Lüthi, M., Gerig, T., Jud, C., Vetter, T.: Gaussian process morphable models. IEEE Trans. Pattern Anal. Mach. Intell. **40**(8), 1860–1873 (2018)
12. Madabhushi, A., Lee, G.: Image analysis and machine learning in digital pathology: challenges and opportunities (2016)
13. Meer, P., Mintz, D., Rosenfeld, A., Kim, D.Y.: Robust regression methods for computer vision: a review. Int. J. Comput. Vis. **6**(1), 59–70 (1991)
14. Morel-Forster, A., Gerig, T., Lüthi, M., Vetter, T.: Probabilistic fitting of active shape models. In: Reuter, M., Wachinger, C., Lombaert, H., Paniagua, B., Lüthi, M., Egger, B. (eds.) ShapeMI 2018. LNCS, vol. 11167, pp. 137–146. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-04747-4_13
15. Myronenko, A., Song, X.: Point set registration: coherent point drift. IEEE Trans. Pattern Anal. Mach. Intell. **32**(12), 2262–2275 (2010)
16. Pajdla, T., Van Gool, L.: Matching of 3-D curves using semi-differential invariants. In: Proceedings of IEEE International Conference on Computer Vision, pp. 390–395. IEEE (1995)
17. Reyneke, C., Thusini, X., Douglas, T., Vetter, T., Mutsvangwa, T.: Construction and validation of image-based statistical shape and intensity models of bone. In: 2018 3rd Biennial South African Biomedical Engineering Conference (SAIBMEC), pp. 1–4. IEEE (2018)
18. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: Niethammer, M., Styner, M., Aylward, S., Zhu, H., Oguz, I., Yap, P.-T., Shen, D. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 146–157. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_12
19. Thompson, P.M., Woods, R.P., Mega, M.S., Toga, A.W.: Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain. Hum. Brain Mapp. **9**(2), 81–92 (2000)
20. Toews, M., Arbel, T.: A statistical parts-based model of anatomical variability. IEEE Trans. Med. Imaging **26**(4), 497–508 (2007)
21. Yokota, F., Okada, T., Takao, M., Sugano, N., Tada, Y., Tomiyama, N., Sato, Y.: Automated CT segmentation of diseased hip using hierarchical and conditional statistical shape models. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8150, pp. 190–197. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40763-5_24