INTERACTIVELY LEARNING A PATIENT SPECIFIC \(K\)-NEAREST NEIGHBOR CLASSIFIER BASED ON CONFIDENCE WEIGHTED SAMPLES

Eva M. van Rikxoort\(^1\), Jonathan G. Goldin\(^1\), Bram van Ginneken\(^2\)\(^3\), Maya Galperin-Aizenberg\(^1\), Chiayi Ni\(^1\), Matthew S. Brown\(^1\)

\(^1\) Department of Radiological Sciences, University of California-Los Angeles, Los Angeles, USA
\(^2\) Image Sciences Institute, University Medical Center Utrecht, Utrecht, the Netherlands
\(^3\) Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre, the Netherlands

ABSTRACT

An automatic segmentation method that fails for one scan of a patient is likely to fail in all follow up scans as well. We propose to construct a patient specific \(k\)-nearest neighbor classifier that learns from the test data while the user is interactively correcting the segmentation in the baseline scan. We apply the system to lung segmentation in chest CT scans. The system is set up in such a way that interaction is limited to single clicks in misclassified areas. Voxels indicated by a user as erroneously labeled are added to the training data. In classification, patient specific confidence weights are applied relative to the similarity between the test and training samples. The method is quantitatively validated on baseline and follow up scans from 16 patients. The results improve substantially in both baseline and follow up scans with only five clicks from the user in the baseline scan on average.

1. INTRODUCTION

Anatomical segmentation is an important prerequisite for quantitative analysis of medical images. Segmentation methods are usually developed a priori on a particular patient population. As a result, when the method is applied to a different population it often results in failures. The wide variety of abnormal patterns, scanning protocols, and other variables makes it practically impossible to design fully automatic segmentation methods that perform well in every case. Especially for applications where severe abnormalities are the standard not the exception, for example in clinical trials, this poses a major problem. For these applications effective interactive segmentation tools are crucial. In addition, in clinical trials, subjects are generally imaged at several time points to monitor the progression of the disease. When an automatic segmentation method fails in the baseline scan of the patient, it is likely to fail in all follow up scans in the same manner. As a result, radiologists are editing the same segmentation errors repeatedly. There are several possible strategies to overcome this. An often used strategy is to use image registration, in which the baseline scan of the patient is registered to the follow up scan and the labels propagated. This approach leads to good results as long as the registration method is able match the scans with high precision. However, a segmentation for the baseline scan needs to be available, and it does not provide a method to interact with the system. We propose to use techniques from pattern recognition to train a patient specific \(k\)-nearest neighbor \((knn)\) classifier that learns from the test data while the user is interactively correcting the segmentation result in the baseline scan.

Voxel classification is a powerful, generic segmentation method that has been applied to several applications with good results, e.g. [1]. Since texture patterns in abnormalities are variable and often show unique patterns for specific patients, it is impossible to make a training dataset that contains all textures and generalizes well. An interesting field of research is how to effectively interact with the voxel classification and incorporate input from a user into the training data to improve the results on scans from the same patient. We propose an interaction scheme in which the user indicates mislabeled areas with a single click in the area. Next a small sphere is defined with voxels inside the area. Voxels from areas that are indicated by a user as mislabeled by the original classification have two distinct properties. First, in the original training dataset the samples are among the samples of a different class. And second, when we add new samples from this area to the training dataset, we have a high confidence that those samples are correctly labeled for the specific patient. We propose to use this knowledge to introduce patient specific confidence weights into the \(knn\) classifier. During classification, samples among the \(k\) nearest neighbors in the training dataset that are from the same patient as the test scan are assigned a confidence weight relative to their similarity to the test sample; the closer the test and training sample are, the higher the confidence weight.

The contribution of this paper is twofold: we present an effective way of interactively updating a voxel classification result, and we present a new patient specific confidence weights into the \(knn\) classifier. During classification, samples among the \(k\) nearest neighbors in the training dataset that are from the same patient as the test scan are assigned a confidence weight relative to their similarity to the test sample; the closer the test and training sample are, the higher the confidence weight.

\(^{\text{This research was funded by the Netherlands Organization for Scientific Research.}}\)
weighted knn classification method to improve segmentation results in follow up scans from the same patient. The method is applied to segmentation of the lungs in chest CT scans of patients with severe abnormalities from three different diseases: asthma, scleroderma, and chronic obstructive pulmonary disease (COPD). Lung segmentation in scans with severe abnormalities is challenging due to the change in appearance of the lung parenchyma as opposed to healthy lungs, which are often used for training [2]. This is illustrated in Figure 1 in which slices from scans containing healthy lungs, asthma, and COPD are shown. In addition, chest CT scans on average consist of over 400 slices, making manual editing time consuming and tedious work. For the evaluation of the method, 32 scans from 16 patients were used; a baseline and follow up scan for each patient. Ground truths were obtained for all 32 scans.

2. METHODS

This section describes the different parts of the system. First, conventional voxel classification is described that is used as a starting point for the interactive segmentation. Next, the interactive system and confidence weights are specified. Finally, the application to follow up scans is explained.

Before applying the method, all scans were subsampled with a factor 2 in each direction using block averaging (the mean of 8 voxels becomes the new voxel value) to reduce required computation times. All computations were performed on those subsampled data. To obtain the result at the original resolution the result was supersampled.

2.1. Conventional voxel classification

In voxel classification two stages can be distinguished, a training stage in which a classifier is trained and a test stage, in which the trained classifier is applied to the test data. In the training stage, a number of voxels are sampled from training images, a set of features is calculated for each voxel, and a classifier is trained. In order to train the classifier a ground truth is required which gives the class label for each voxel.

As a classifier, we used a $k$-nearest neighbor classifier, with $k$ set to 15. For training the system, from a set of training scans with lung segmentations, every fourth voxel was sampled and a set features were calculated for each voxel. As features we used the output of Gaussian filters up to and including second order derivatives ($L_x$, $L_{x^2}$, $L_z$, $L_{xz}$, $L_{yz}$, $L_{zy}$), and the gradient ($L_i$) at four scales ($\sigma = 1, 2, 4, 8$), as well as the original Hounsfield unit ($L_0$). This makes a total of 45 features. The most important features were selected using sequential forward selection. Feature selection lead to a selection of the following 10 features for the final data set: $L_0$, $L$ and $L_y$ at $\sigma = 1$, $L_{zz}$ at $\sigma = 2$, $L$, $L_{yy}$, and $L_z$ at $\sigma = 4$, and $L$, $L_y$, and $L_{xx}$ at $\sigma = 8$.

The lungs in all scans are segmented by applying the trained 15nn classifier to each voxel in the scan. To obtain a binary segmentation, the probabilities are blurred using a Gaussian of unit standard deviation, the result is thresholded at 0.5 and only the largest two components are retained.

2.2. Interactive confidence weighted classification

The result of the conventional voxel classification is presented to the user as the initial lung segmentation. If the user detects an area in the segmentation that is mislabeled, a single click in this area indicates a misclassified voxel. Next, the user sets the radius of a sphere that falls inside this area. The user does not need to fit the sphere precisely inside the area, a small sphere is sufficient as long as all voxels inside the sphere have the same label. An example of a clicked point and area can be seen in Figure 2.

The voxels inside the sphere are added to the training data with their correct label and the scan is reclassified. To speed up the process the user can indicate an area that should be reclassified instead of the whole scan. During reclassification we introduce confidence weights based on three observations: First, the newly added samples are misclassified by the original training dataset. Second, the labels of the newly added samples are very reliable for scans of the same patient. And third, the number of added samples is very low compared to the size of the complete dataset. Giving the new samples a higher weight makes sure that they have an impact despite their low number, and is justified by their higher reliability. However, at the same time we need to make sure that voxels that were correctly classified by the original dataset will not be misclassified due to this weighting. Therefore, we define the confidence weights relative to the similarity of the test and training sample. The rationale behind this approach is that even though the confidence in those samples is higher, they are among samples of the opposite class in the training dataset. By only giving a high confidence weight to training samples when the test sample is truly similar to the confident sample, we assure that we do not mislabel voxels from the opposite class that were previously classified correctly.

For conventional, unweighted knn classification, every sample gets a weight of $\frac{1}{k}$. For "confidence weighted knn" classification, the (unnormalized) weight $w'$ of each training
sample $i$ that is among the $k$ nearest neighbors is defined as:

$$w'_i = \frac{1 + c \times \exp\left(\frac{\text{median}_{j \in N}(\text{sqrldst}(j))}{\text{sqrldst}(i)}\right)}{k}$$

where $k$ is the number of neighbors, $N$ is the set of all $k$ neighbors, $c$ is a constant, $\text{median}(\text{sqrldst}(j))$ is the median of the squared Euclidean distances of the normalized features of all $k$ neighbors to the test sample, and $\text{sqrldst}(i)$ is the squared Euclidean distance of the neighbor under consideration to the test sample. $c$ is set for each sample $i$ separately depending on whether it is a sample from the same patient as the test sample or not. When $c$ is set to 0, the conventional weight $\frac{1}{k}$ is given to the sample. By multiplying $c$ with the exponential of the relative distance of the training sample to the test sample, we ensure that training samples only get a high confidence weight when the test sample is truly among the samples from the same patient. Since the confidence weights do not necessarily sum up to one, we define the final weight $w_i$ as $w_i = \frac{w'_i}{\sum_{j \in N} w'_j}$.

The area indicated by the user (or the whole scan) is reclassified. The soft result is merged with the current segmentation by thresholding at 0.5 and taking the largest component. This process is repeated until the user is satisfied with the segmentation.

2.3. Patient specific confidence weighted classification

For a follow up scan of a patient for which the interactive segmentation has been performed on a previous scan, a patient-specific confidence training set is available. Confidence weighted $k$nn classification is applied using equation 1. For the follow up scan, the confidence $c$ for samples from the same patient is generally set to a lower value than for the baseline scan since they are not taken from the same scan and scan characteristics might be different.

3. EXPERIMENTS AND RESULTS

The voxel classification described in Section 2.1 was trained using 20 scans from asymptomatic subjects for which ground truth lung segmentations were available. An example slice of a training scan can be seen in the first frame of Figure 1. For the evaluation of this study, 32 scans from 16 different patients from three cohorts showing abnormalities due to COPD, scleroderma, and asthma were used. For each patient, a baseline and follow up scan were available. The details of the data are provided in Table 1. By using patients from different cohorts we make sure that different disease patterns are available in the test data. For all 32 test scans, reference segmentations were obtained semi-automatically using a previously proposed method [3] followed by manual editing by a radiologist.

The interactive segmentation as described in Section 2.2 was applied to all 16 baseline scans. The confidence value $c$ from equation 1 was set to 4 for samples from the same patient, and to 0 for all other samples. The number of clicks from the user as well as the intermediate results were stored. To quantify the performance, volumetric overlap fraction was used. For two volumes the volumetric overlap fraction $O$ is defined as the volume of their intersection divided by the volume of their union. The more equal two volumes are, the closer $O$ is to one. Figure 3 shows for each scan, the overlap after each number of clicks from the user. It can be seen that for all scans the segmentation result improves with every click. It should be noted that since the user often selects a small area to be reclassified, the improvements in overlap for each click are often minor compared to the whole lung volume. On average, five clicks were performed per scan. The average overlap after interactive classification was 0.93, compared to 0.87 before interaction. Figure 2 shows for one scan the different steps of the interactive segmentation.

<table>
<thead>
<tr>
<th>cohort (#)</th>
<th>supine/prone</th>
<th>slice spacing(mm)</th>
<th>exposure (mAs)</th>
<th>RV/TLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asthma (5)</td>
<td>supine</td>
<td>0.7</td>
<td>20-60</td>
<td>RV</td>
</tr>
<tr>
<td>Scleroderma (3)</td>
<td>prone</td>
<td>1-1.25</td>
<td>100</td>
<td>TLC</td>
</tr>
<tr>
<td>COPD (8)</td>
<td>supine</td>
<td>0.75-2.5</td>
<td>70-150</td>
<td>TLC&amp;RVC</td>
</tr>
</tbody>
</table>
For the 16 follow up scans, patient specific confidence weighted knn classification was applied as described in Section 2.3. For the follow up scans, \( c \) from equation 1 was set to 3 for samples from the same patient and to 0 for all other samples. In the follow up scans, \( O \) was 0.90 for the confidence weighted knn, as compared to 0.83 for the conventional method. In Figure 4 the results of two scans and the corresponding follow up scan are shown for both the confidence weighted and conventional classification. It can be seen that there is a substantial improvement in performance.

With the interactive system designed for this paper, the user clicked on average five times per scan to get a satisfying lung segmentation. The average size of the area indicated to be reclassified was \( 40 \times 40 \times 40 \) voxels. With this setting, reclassification takes about two minutes on a single core, so the total interaction took 10 minutes per scan on average. The method was not optimized for speed, and could be speeded up by using a smaller dataset during reclassification. Using the patient specific knn classification constructed during the interaction, the performance in the follow up scans improved without additional interaction or computation time.

5. REFERENCES


4. CONCLUSION AND DISCUSSION

In this paper, an interactive voxel classification method is presented that learns a patient specific model to be used for follow up scans while the user is correcting the segmentation in the baseline scan. The main novelty of this paper is the confidence weighted knn classification, which is able to improve the segmentation result without mislabeling previously correctly labeled samples. The method was applied to segmentation of the lungs in 32 chest CT scans of 16 patients. The scans were taken from three different cohorts, ensuring that different texture types were present in the test data. With only five clicks by a user on average, the segmentation result in the baseline scan improved substantially. For the follow up scans, the overlap for the confidence weighted knn improved from 0.83 to 0.90 as compared to conventional knn classification, without additional user interaction. Since the lungs are large convex objects, volumetric overlap measures are generally high and might not reflect the improvements completely. For example, for one of the scans from the asthma cohort, the overlap improved from 0.78 to 0.89, which resulted in 203 ml out of the total 1710 ml of lungs at residual volume being included in the segmentation of the confidence weighted knn that was missed by the conventional method.

Inspecting the results shows a different pattern for the asthma and scleroderma cases than for the COPD cases. For the asthma and scleroderma patients, user interaction was more efficient; less clicks were needed per scan on average to improve the results. However, the initial voxel classification also performed better in these cases. In the COPD cases more clicks were needed per scan on average to improve the result, but the improvement in terms of overlap were better both in the baseline and the follow up scans.

Fig. 3. This figure shows for each test scan, after each number of clicks from the user, the overlap.

Fig. 4. In this figure the results for two follow up scans are shown. The first row shows the result of a scan from the asthma cohort, in the second row a scan from the COPD cohort is shown. In both rows, the first frame shows the original slice. The second frame shows the result of the conventional voxel classification. In the third frame the output of the confidence weighted classification is shown. The last row shows the ground truth.

4. CONCLUSION AND DISCUSSION

In this paper, an interactive voxel classification method is presented that learns a patient specific model to be used for follow up scans while the user is correcting the segmentation in the baseline scan. The main novelty of this paper is the confidence weighted knn classification, which is able to improve the segmentation result without mislabeling previously correctly labeled samples. The method was applied to segmentation of the lungs in 32 chest CT scans of 16 patients. The scans were taken from three different cohorts, ensuring that different texture types were present in the test data. With only five clicks by a user on average, the segmentation result in the baseline scan was improved substantially. For the follow up scans, the overlap for the confidence weighted knn improved from 0.83 to 0.90 as compared to conventional knn classification, without additional user interaction. Since the lungs are large convex objects, volumetric overlap measures are generally high and might not reflect the improvements completely. For example, for one of the scans from the asthma cohort, the overlap improved from 0.78 to 0.89, which resulted in 203 ml out of the total 1710 ml of lungs at residual volume being included in the segmentation of the confidence weighted knn that was missed by the conventional method.

Inspecting the results shows a different pattern for the asthma and scleroderma cases than for the COPD cases. For the asthma and scleroderma patients, user interaction was more efficient; less clicks were needed per scan on average to improve the results. However, the initial voxel classification also performed better in these cases. In the COPD cases more clicks were needed per scan on average to improve the result, but the improvement in terms of overlap were better both in the baseline and the follow up scans.

With the interactive system designed for this paper, the user clicked on average five times per scan to get a satisfying lung segmentation. The average size of the area indicated to be reclassified was \( 40 \times 40 \times 40 \) voxels. With this setting, reclassification takes about two minutes on a single core, so the total interaction took 10 minutes per scan on average. The method was not optimized for speed, and could be speeded up by using a smaller dataset during reclassification. Using the patient specific knn classification constructed during the interaction, the performance in the follow up scans improved without additional interaction or computation time.

5. REFERENCES

