Eurographics Workshop on Visual Computing for Biology and Medicine (2015)
K. Bühler, L. Linsen, and N. W. John (Editors)

Visual Analysis of Medical Image Segmentation Feature Space for Interactive Supervised Classification

V. Molchanov[1], T. Chitiboi[1,2], and L. Linsen[1]

[1]Jacobs University, Bremen, Germany
[2]Fraunhofer MEVIS, Bremen, Germany

Abstract

Classification of image regions is a crucial step in many image segmentation algorithms. Assigning a segment to a certain class can be based on various numerical characteristics such as size, intensity statistics, or shape, which build a multi-dimensional feature space describing the segments. It is commonly unclear and not intuitive, however, how much influence or weight should be assigned to the individual features to obtain a best classification. We propose an interactive supervised approach to the classification step based on a feature-space visualization. Our visualization method helps the user to better understand the structure of the feature space and to interactively optimize feature selection and assigned weights. When investigating labeled training data, the user generates optimal descriptors for each target class. The obtained set of descriptors can then be transferred to classify unlabeled data. We show the effectiveness of our approach by embedding our interactive supervised classification method into a medical image segmentation pipeline for two application scenarios: detecting vertebral bodies in sagittal CT image slices, where we improve the overall accuracy, and detecting the pharynx in head MRI data.

1. Introduction

The task of medical image segmentation is to identify semantically meaningful structures from images acquired using medical imaging techniques. While the detection of low-level objects in form of small homogeneous regions can be easily automized, assembling them automatically to higher-level structures remains a challenge. Technically, a classification process is involved that decides which low-level objects contribute to the target structure. The classification is carried out by looking into properties (or features) of the low-level objects such as location, geometrical attributes, or local intensity statistics. Those properties form a multi-dimensional *feature space*. Classification can be performed by detecting clusters in the feature space, but it is typically unknown which of the features are most descriptive to discriminate objects of the target class from others or how to weight the features in order to find the best classifier.

We propose a visual analysis tool that allows for the interactive inspection of the feature space to investigate the influence of features, to identify patterns, and to detect descriptive features. Moreover, one can interactively define weights to the features and describe a classifier in the weighted multidimensional feature space. We obtain a supervised classification approach by, first, defining a classifier using our interactive visual analysis tool for a labeled training set and, second, applying the classifier to an unlabeled data set.

We embed our supervised classification approach into a medical image segmentation pipeline and apply it to two scenarios, namely, to detect vertebral bodies in sagittal CT image slices and to detect the pharynx in head MRI data. We show the effectiveness of our classification approach by comparing our results to commonly used automatic classifiers and evaluate our interactive visual tool by performing a controlled user study.

2. Related Work

Several approaches for image segmentation have as a core step the classification of image regions [Bla10]. Thus, part of the segmentation process is to solve a classification problem defined in the feature space. This classification problem can be solved in an unsupervised manner by clustering approaches specifically designed for segmentation, e.g. [CM02] or by supervised, semi-automatic approaches that require training data.

© The Eurographics Association 2015.

DOI: 10.2312/vcbm.20151204

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

Semi-automatic systems for image segmentation may exploit the user domain knowledge in different ways [ZX13]. They can ask the user to manually specify seed points and/or strokes in the image space to indicate groups of regions that belong to the same class, e.g. [NZZW10], or they allow for user intervention to guide and correct results, e.g. [SMH10]. The expert's work load is reduced with the use of trained classifiers, where his/her input is only required for a variable amount of training data, and automatically transferred to new data. However, interpreting the semantics behind a trained classifier model is a complex task, as shown in recent that attempts a statistical visual exploration of, in this case, a Random Forest model [Ehr15].

While the typical goal in image analysis is to improve the segmentation result, Schultz and Kindlmann [SK13] use interactive visual tools to analyze the performance of segmentation algorithms. Our approach follows a similar idea, namely, using interactive visual analysis techniques for an understanding of the relevant features. This knowledge can help an expert to identify regions where segmentation is problematic. For example, Von Landesberger et al. [VLBK*13] try to analyze the results of different automatic segmentation approaches to possibly improve the algorithm. We use an interactive analysis of the feature space to develop an effective supervised classification step in a transparent and intuitive way. The interactive classification step is directly embedded into specific segmentation approaches, as we demonstrate in sections 5 and 6.

3. Visual Analysis of the Feature Space

We present a visual analysis approach that allows for an interactive investigation of the multi-dimensional feature space of low-level homogeneous image regions. The first step is to generate an oversegmentation of the given image into low-level objects. To ensure a good classification of the image regions, an edge-preserving segmentation like the watershed transform on the gradient image [HP03] is applied to capture the individual structures present in the image. To reduce over-segmentation, images are typically preprocessed with a Gaussian or rank filter. Anisotropic diffusion filters [SNS*98] are additionally used to enhance the relative brightness of image borders before segmentation.

We characterize the detected image regions by a range of descriptive features, as done in the literature [SCHH13]. In particular, we make use of the spectral distribution of the enclosed pixels using standard statistical methods (e.g. mean, standard deviation, upper and lower quartiles), shape descriptors based on central image moments [BB09], and distance-related features. To make the paper self-contained, we briefly describe the features that will be used later on. Regarding the region's shape, there are many descriptors available in literature [MIJ08]. One of the standard methods to compute shape features is based on central image moments [BB09] of the binary mask of a region r, where $r(x,y) = 1$ if the pixel (x,y) belongs to the region and $r(x,y) = 0$ otherwise. The central image moments are defined by:

$$\mu_{pq}(r) = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q r(x,y),$$

where the size of the region is given by μ_{00}, while $(\bar{x},\bar{y})$ represents its center of gravity. Furthermore, one can perform principal component analysis (PCA) on the pixel distribution of the region. Features such as the principal eigenvectors and corresponding eigenvalues λ_1 and λ_2 can be computed for the covariance matrix of r, which is defined as

$$cov(r) = \begin{bmatrix} \mu_{20}(r) & \mu_{11}(r) \\ \mu_{11}(r) & \mu_{02}(r) \end{bmatrix}.$$

The principal eigenvectors define the region's orientation, while the ratio of their corresponding eigenvalues measures the region's eccentricity by:

$$eccentricity = 1 - \frac{4 \cdot \lambda_1 \cdot \lambda_2}{(\lambda_1 + \lambda_2)^2},$$

Using this result, a best-fitting ellipse having the same area as the given region can be determined, whose overlap with the original shape determines the region's compactness. When describing orthogonal shapes, further features measuring rectangularity and square fit can be extracted that quantify the similarity of region r with its minimum bounding rectangle (MBR):

$$rectangularity = \frac{r_{area}}{MBR_{area}},$$

$$squarelikeness = \frac{MBR_{area}}{\left(\frac{MBR_{height}+MBR_{width}}{2}\right)^2}.$$

Another class of shape features used in our experiments is based on the distance transform. In particular, we computed the Euclidean distance from each pixel inside the image region to the background such that the region border has minimum distance 0 and the center of gravity is the point with the maximum distance. Then, for each segment, we computed the max, mean, upper and lower quartiles as distance statistics of pixels giving the geometry of the segment.

All these potentially descriptive features describe a multi-dimensional space, where each low-level image region corresponds to a multi-dimensional point in that feature space. As the numerical values of the features may have different ranges, a normalization is applied. Now, the challenge is to explore the space of potentially useful existing features in order to determine which of them have a decisive contribution to the classification and how they interact.

We propose an intuitive interactive exploration of the multi-dimensional space using a star-coordinates widget. Dimensionality reduction methods are applied to map multi-dimensional data to a low-dimensional space. Mapping to a 2D or 3D space allows for a visual exploration of the

© The Eurographics Association 2015.

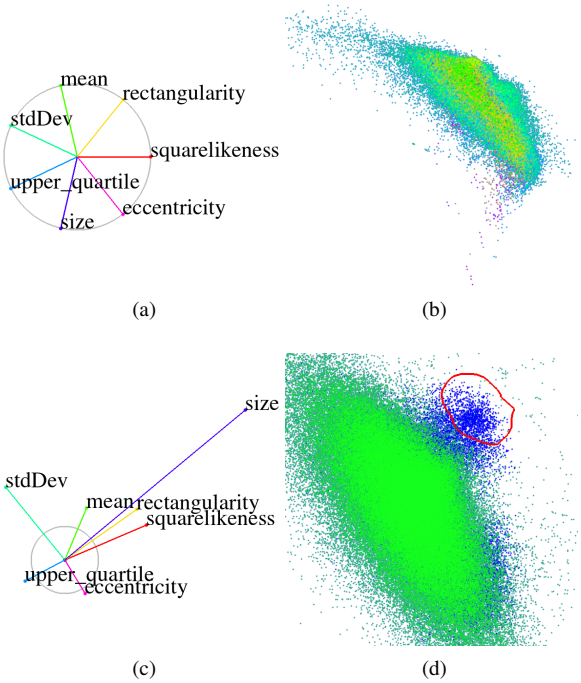

Figure 1: Default configuration of star-coordinates widget (a). The unit circle is shown in gray. Colors in the resulting projected view of a training spine CT dataset (b) define samples belonging to different classes. A result of partial decoupling of target samples (blue) from the rest (green) is shown in (d). Red polyline denotes a user made selection made. The corresponding projection matrix is shown in (c).

data distribution. PCA [Jol86] or multi-dimensional scaling [BG10] are the most commonly used classical dimensionality reduction approaches. The choice of an optimal method is usually done by compromising between preservation of distances and good clusters segregation. Other criteria include simplicity and efficiency of the approach. Since our aim was to develop an intuitive system that allows for interactive modification, linear projections are the most suitable choice, as they do not introduce much distortion and keep computation costs low.

Any linear projection of the m-dimensional feature space Q onto the 2-dimensional visual space U can be represented in terms of a $2 \times m$ matrix P. The columns of P consist of the coordinates of the images in U of basis vectors from Q. Since the image of the origin in Q is always the origin in U, we visualize columns of P as axes of a star-coordinates system [TM03]. Thus, there is a one-to-one correspondence between the space of linear projections and the configurations of basis vectors in the star-coordinates widget. The weights of the individual features are given by the length and relative position of the basis vectors. In the default configuration, all basis vectors are distributed uniformly over the unit circle as shown in Figure 1a, which corresponds to the projection matrix P:

$$P = \begin{pmatrix} 1 & \cos(2\pi/m) & \dots & \cos(2\pi(m-1)/m) \\ 0 & \sin(2\pi/m) & \dots & \sin(2\pi(m-1)/m) \end{pmatrix}.$$

Given an initial linear projection, our exploratory system allows the user to interactively manipulate the projection matrix by changing the positions of the end-points of the basis vectors in the star-coordinates widget. Elements of P are then recomputed and used for the new projection the multi-dimensional feature data. Through this process, the user's interactions lead to immediate update of the projected data thus providing real-time feedback. Placing the end of a basis vector to the origin leads to a vanishing corresponding column of P, such that values of the selected feature do not influence the resulting projection. Therefore, our approach naturally and intuitively includes the means of analyzing feature subspaces. In particular, traditional 2D scatterplots are obtained when only two orthonormal axes are active. How the interactions are performed is best understood by looking at the accompanying video.

4. Supervised Classification

When visualizing a training dataset, where the objects are labeled, it is possible to encode the classes in the projected view using one color per class, see Figure 1b. Commonly, one is interested in a single target class, such that the binary decision of belonging to that class is encoded using only two colors, which reduces the perceptional load. In general, classes may be overlapping, i.e., it is not possible to find a projection matrix that allows for a perfect separation. Our goal is to find a configuration of the star-coordinates widget, such that a representative subset of samples from the target class is visually decoupled from the rest. In Figures 1c and 1d, one can see that the chosen configuration of the star-coordinates widget creates a projected view that exhibits a region with mainly blue samples. After having identified such a region, the classification step is executed interactively by the user surrounding the region by a polygonal line. The number of blue samples inside (or outside) the region compared to the overall number of blue samples gives an immediate and intuitive understanding of how many true (or false) positives are to be expected. Respectively, the number of green samples inside (or outside) the region compared to the overall number of green samples gives an immediate and intuitive understanding of how many false (or true) negatives are to be expected. For our applications it was easy to automatically remove obvious false negatives in a post-processing step. With such knowledge, the user can select a larger region to increase the number of true positives.

The selected area in conjunction with the recorded projection matrix represents our classifier, which can be applied to new unlabeled data. In order to apply it, one maps the new feature data using the recorded projection matrix and selects

© The Eurographics Association 2015.

only those samples that fall into the selected region. The selected samples are classified as positives (belonging to the target class), while the others are classified as negatives.

For image segmentation purposes, our interactive feature space exploration and classification tool gets complemented by a linked view showing the classification result overlaid with the medical images. Hence, one can immediately validate (and possibly correct) the classification result.

5. Application to Spine Segmentation in CT Images

We illustrate the capacity of our approach by embedding it into a segmentation pipeline for detecting vertebral bodies in sagittal CT image slices. The task is to segment a class of objects (vertebrae) with regular appearance, located in a specific spatial relation in the image space to form the spine. As a segmentation task, this problem has been tackled by several approaches, for either CT or MRI, which roughly consist of two steps: First, the recurring structures are detected (vertebrae [HCLN09] or inter-vertebral disks [SSQW07]) to estimate the spine position. Then, the spine is reconstructed using various segmentation methods such as active shape [TP11] and active appearance [KOE*09], graph cuts [KWZ*13], etc. The initial object detection is most often performed by first locating the object in the image space and then segmenting. However, the reverse approach is also possible: The image is first segmented and the regions belonging to the spine are detected by classification. This reduces the search space and supplies more information for finding the vertebrae (such as the shape of segmented regions). Yao et al. [YOS06] perform region-based vertebra detection in axial slices but report significant leaks in the segmentation.

Our approach is to detect segments that represent vertebral bodies in the sagittal plane, where they have rectangular shape and a specific intensity profile. The segmentation approach first creates a watershed-based oversegmentation [HP03] of a region of interest that includes all the bone structures obtained by thresholding and morphological operations (Figures 2a,2e and 2b,2f). For the low-level image regions, we build the multi-dimensional feature space, which is analyzed by applying the proposed star-coordinates widget to training data labeled by experts (Figure 1). Within the interactive analysis, a classifier with very high specificity that distinguishes vertebrae regions in the multi-dimensional feature space is built (Figure 2c,2g). This classifier can, then, be applied to classify unlabeled test data. After classification, the spine is reconstructed using a model based on the vertical alignment of the vertebrae and consistency over consecutive sagittal image slices (Figures 2d,2h).

We applied the proposed technique to 20 CT datasets, each consisting of 21 saggital slices. The low-level image regions were labeled by a domain specialist to identify which belong to the spine for ground truth. On average, the total

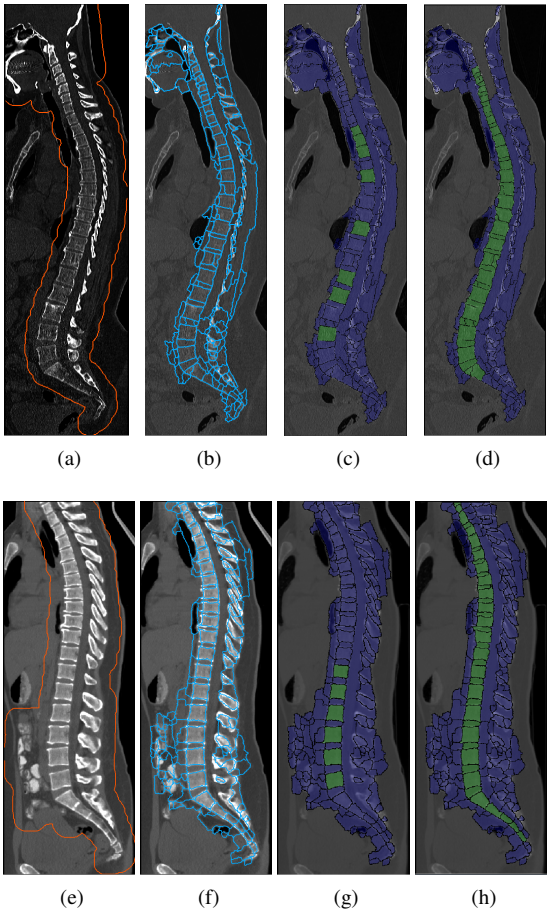

Figure 2: Segmentation steps: (a),(e) Original image and ROI. (b),(f) Segmented regions. (c),(g) Vertebrae detected by classification. (d),(h) Final segmentation.

number of segments is $6,664 \pm 3,275$ per dataset, of which 421 ± 70 were classified as spine. Detailed information can be found in Table 1.

For each segment we considered the following seven features: *squarelikeness*, *rectangularity*, *eccentricity*, mean intensity (*mean*), standard deviation (*stdDev*), upper quartile of voxel intensities (*upper_quartile*), and world *size*, described in detail in section 3. The classification by visual analysis was evaluated by cross-validation in a controlled user study. In a set of experiments, 19 of the 20 datasets were used for training and one for evaluation. The users were asked to find a layout of the star-coordinates widget such that two classes are best separated in the projection and mark a polygonal area containing a representative subset of positive samples. The goals of the user study were to investigate how robustly users would find the best projected view and how the interactively obtained classifier performs on the test data. For the user study, we recruited eight subjects with no

© The Eurographics Association 2015.

Table 1: Total number of segments and number of positive segments (that belong to spine) in each of the saggital CT datasets. The average (AVG) and standard deviation (SD) are provided.

dataset	1	2	3	4	5	6	7	8	9	10	**AVG**
all	4,251	4,385	8,362	13,475	5,168	9,176	5,117	3,249	11,999	3,368	**6,664**
positive	468	541	520	468	351	424	305	315	344	366	**420.9**
dataset	11	12	13	14	15	16	17	18	19	20	**SD**
all	3,432	5,181	6,473	8,154	4,786	5,947	5,667	14,153	6,639	4,301	**3,275**
positive	435	536	410	381	346	454	393	483	443	435	**70.46**

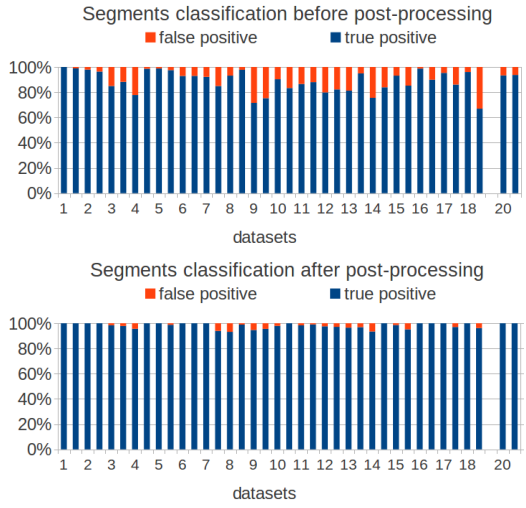

Figure 3: Relative number of false and true positive segments resulting from the vertebrae classifier trained by two different users for each dataset. Top - before, bottom - after post-processing.

a priori domain knowledge and unaware of the underlying segmentation task.

Each datasets was segmented using the classifiers built interactively by two different observers. The results of the vertebra classification (relative number or true/false positive segments) before and after post-processing are provided in Figure 3. The average precision for all data sets obtained based on the interactively determined projections was on average 89.05% and increased to 98.17% after post-processing.

One randomly selected test configuration (corresponding to the dataset No. 19) was performed by all users to assess variability. The average precision obtained in this case was 83.06% before and 94.83% after the post-processing step. The detailed results for each user are shown in Table 2.

The times spent by users to complete each experiment (Table 2 - top) varies from 90 to 365 seconds with an average of 159 ± 101 seconds. This shows that finding a projection matrix using the star-coordinates widget and selecting a region of interest takes only a few minutes with minimum training in case of seven features.

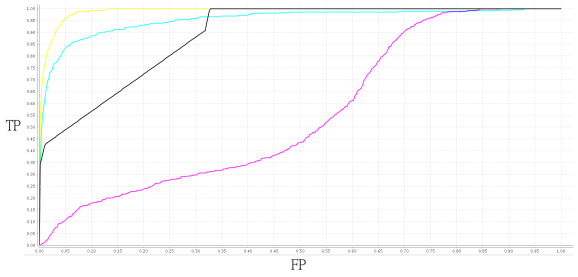

Figure 4: ROCs comparing the performance of different classifiers on the spine datasets: red - Support Vector Machine, black - Random Forest, blue - Naïve Bayes, yellow - Neural Network.

Figure 5 shows projection matrices in form of star-coordinates widgets, which were found by all eight participants for the dataset No. 19. Note that the axes rotation around the origin of the star-coordinates widget as well as a simultaneous scaling of all axes do not qualitatively change the resulting projection. Important are the relative size of the axes and angles between them. Thus, to simplify the comparison of results, we rotated all original images to get the dominating axis labeled size pointing upwards. We can observe that *size* was the most relevant feature to discriminate vertebral structures, followed by the shape features *rectangularity* and *squarelikeness* and the standard deviation *stdDev* of the intensity values. Overall, we noticed common patterns in all solutions regarding the relative size and angles between the feature axes. Considering this, it could be possible to derive a probabilistic projection matrix with better score based on outputs from a sufficiently large number of participants.

We compared our supervised classification results to the output of four state-of-the-art classifiers whose ROC curves are shown in Figure 4. For the vertebrae detection task, we require a classifier with high precision. Our subjects were able to detect on average 13.9% of the true samples while including only 0.16% of the set of negative samples, situating us extremely close to the Y axis on the same plot. Thus, our results are comparable to the performance of the Random Forest, Naïve Bayes, and Neural Network, outper-

© The Eurographics Association 2015.

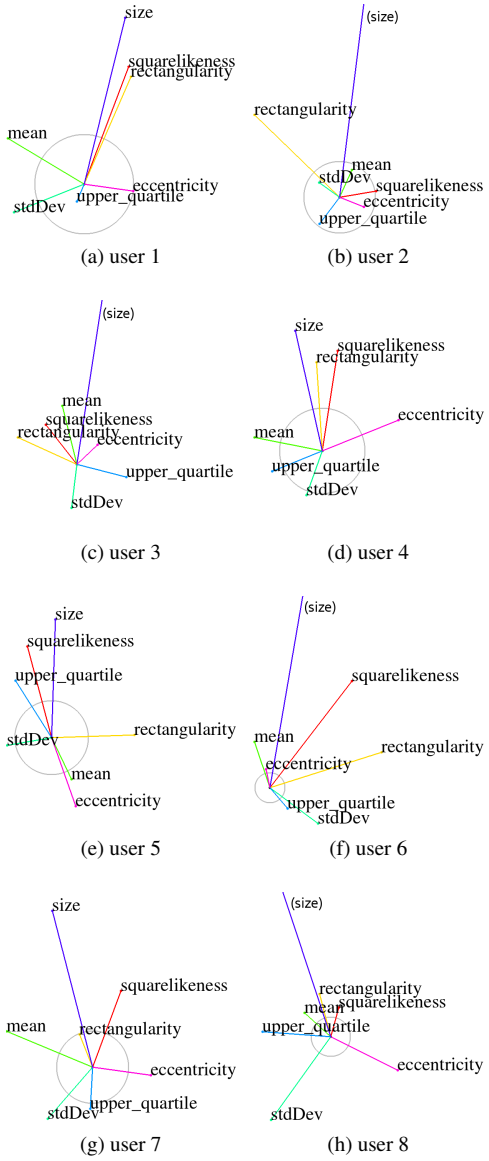

Figure 5: Final configurations of the projection matrix found by all users for dataset 18.

forming the Support Vector Machine. For the automatic classifiers, however, additional interactive fine-tuning would be required to adjust their acceptance threshold to obtain a high precision. The choice of these thresholds is rather unintuitive, while the user interactions in our visual tool provide an intuitive understanding of the choice made.

Our tool is not only able to support the generation of a classifier, but it is also helpful to analyze the classification process. For example, one can investigate the features of false-negative segments to understand how they differ from the selected area and why they were missed. The analysis allows for a better understanding of how certain features discriminate the target class. This information can be used to improve automatic classification by adding additional features or adjusting how current features are computed.

After the classification, an automatic post-processing step based on domain knowledge was applied to reject false positives. Samples with too large or too small size or lying in the upper 10% of the image which could represent other skeletal structures were eliminated. After post-processing the precision increased from 89.05% before to 98.17% . The remaining selected segments were used as seeds in a model-based region-growing method to reconstruct the spine, following the work flow presented in [SCHH13]. Our final results reached an average Dice score of $0.899(\pm 0.069)$. For the dataset processed by all participants we obtained an average Dice score $0.956(\pm 0.014)$. This result improves the Dice score of 0.948 obtained in [SCHH13], where the initial vertebrae seeds were obtain using a random forest classifier.

6. Application to Pharynx Segmentation in MR Images

Another application where we illustrate our supervised classification approach based on feature space visual analysis is the detection of the pharynx in axial MR image slices. In this setting, the pharynx regions have a less regular shape and a more diverse appearance (Figure 7a), which makes the problem of directly choosing the right combination of features to distinguish the targeted regions more difficult than in the case of the vertebrae. The large intensity variations in MRI make pharynx segmentation more challenging than in CT images. Existing semi-automatic segmentation approaches start from manually placed markers and perform a guided 3D image growing on a pixel level, such as Ivanovska et al. [IDL*13] and Liu et al. [LUO*03]. However, operating exclusively on a pixel level the spatial relation between the pharynx and other neighboring anatomical regions is more difficult to establish, which is why these approaches face difficulties in separating the pharynx from the complete respiratory track. Our approach is to detect segmented regions that represent pharynx cross-sections in over-segmented axial slices. Again, we extract low-level image regions using the watershed segmentation [HP03] and compute the following features: world *size*, *eccentricity*, direction of the principal component of the object mask (*prime_axis_X*, *prime_axis_Y*), mean intensity (*mean*), upper and lower quartiles of voxel intensities (*upper_quartile*, *lower_quartile*), and statistics over the distance transform of the region pixels to the background (*distance_median*, *distance_upper_quartile*, *distance_lower_quartile*), as described in section 3. The feature space is then analyzed using the star-coordinates widget to define a classifier for the pharynx samples from training data (Figure 6).

The classifier can, then, be applied to new data fully automatically. The classification result can further be improved

© The Eurographics Association 2015.

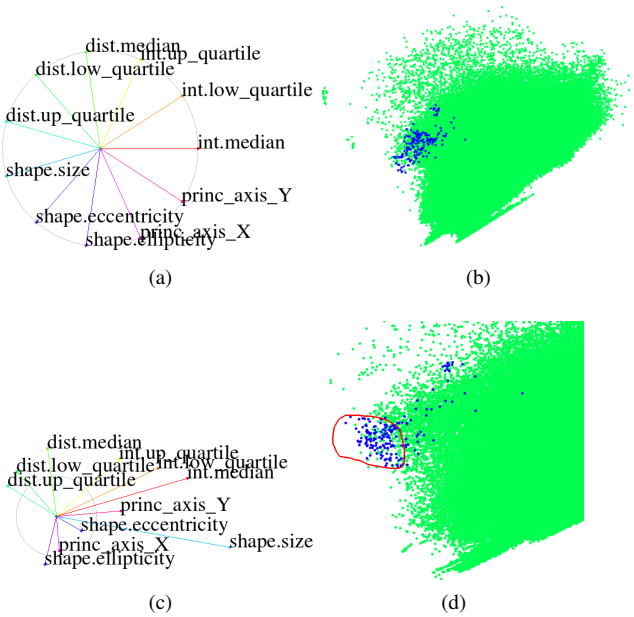

Figure 6: (a) Default configuration of star-coordinates widget (a). Colors in the resulted projected view of a training pharynx MRI dataset (b) define samples belonging to different classes. (c) Typical configuration of star-coordinates widget. (d) Resulting projection of training data, the blue and green samples belonging to the two classes. Red polyline denotes a selection made by the user.

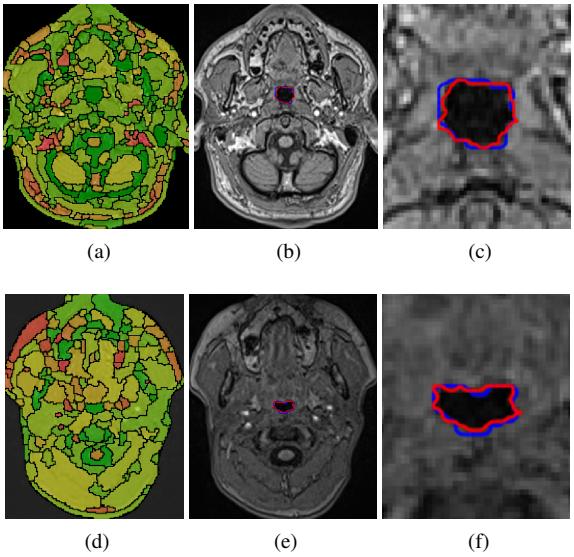

Figure 7: (a), (d) Oversegmented MR images. Segments are colored according to mean intensity from green = low intensity values to red = high values. (b), (e) Our result shown in red compared to ground truth (blue). (c), (f) Zoomed-in result.

in a post-processing step by exploiting alignment over consecutive axial slices.

We applied the proposed technique to ten MRI datasets, each consisting of 45 axial slices with $2 \times 2\ mm^2$ resolution. Images were segmented and labeled by a domain specialist to identify the pharynx. These data serve as the ground truth in our experiment. The total number of segments per dataset is on average $9,689 \pm 1132$, out of which on average 21 ± 6 segments belong to the pharynx (see Table 3).

We performed a similar evaluation study with eight participants who had no prior knowledge about the datasets, target class, or the significance of the features. For cross-validation, a series of experiments was set up where nine datasets were used for training and the tenth for testing. Overall every dataset was classified by two or three participants. The mean precision obtained was $59.29\%(\pm 15.27\%)$ and increased to $70.18\%(\pm 24.65\%)$ after post-processing. The relative numbers or true and false positives obtained by each user per dataset are shown in Figure 8. The average time for completion per experiment was 238 seconds and decreased to about 2 minutes when the users got more experience with the feature space.

A typical projection obtained after user's explorative interactions is shown in Figure 6c. All users discovered a high positive correlation of three intensity features (mean, lower and upper quartiles), see Figure 6c. Other important features are the *shape.size* and *shape.eccentricity*, whose axes usually point to a direction nearly orthogonal to the intensity-related axes.

Figure 7a illustrates the classification problem. The oversegmented image is colored according to the mean intensity of each region. By visual inspection, it is challenging for a human observer to distinguish the pharynx segments using only shape and intensity information. In contrast, the problem could be easily solved by visual exploration in feature space. Our sample results highlighted in red color in Figures 7b and 7c are compared with the ground truth (blue).

7. Conclusion and Future Work

We have presented an interactive visual analysis tool for exploring image segmentation feature spaces. Our approach allows for intuitive generation of classifiers represented as a projection matrix and a selected area in the projected view. The system serves to better understand the features' roles, weights, and their interplay, in order to characterize the structures to-be-detected. We have shown the potential of our supervised classification approach when embedded in real-world medical image segmentation problems, namely detecting vertebral bodies in CT slices and the pharynx in

© The Eurographics Association 2015.

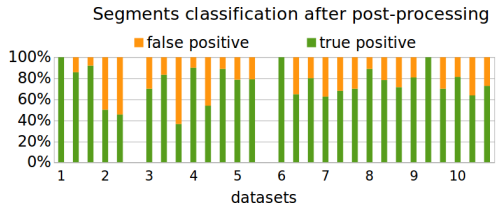

Figure 8: Relative number of false and true positive segments classified by two or three different users for each dataset in the pharynx segmentation task, after post-processing.

MR slices. The classification results were comparable with or better than those obtained with state-of-the-art classifiers, even after fine-tuning their parameters. In future work, we want to analyze results from a number of users to derive projection matrices with better stability. Moreoever, the process of finding the optimal projection matrix and a proper selection area can be supported in an automated manner by solving an optimization problem with an error functional reflecting the fraction of true-positive samples for the current setting.

References

[BB09] BURGER W., BURGE M. J.: *Principles of Digital Image Processing: Core Algorithms*. Springer, 2009. 2

[BG10] BORG I., GROENEN P. J. F.: *Modern Multidimensional Scaling Theory and Applications*, 2nd edition ed. Springer Series in Statistics. Springer, 2010. 3

[Bla10] BLASCHKE T.: Object based image analysis for remote sensing. *ISPRS 65*, 1 (Jan 2010), 2–16. URL: http://dx.doi.org/10.1016/j.isprsjprs.2009.06.004, doi:10.1016/j.isprsjprs.2009.06.004. 1

[CM02] COMANICIU D., MEER P.: Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 24*, 5 (2002), 603–619. 1

[Ehr15] EHRLINGER J.: ggrandomforests: Visually exploring a random forest for regression. *arXiv preprint arXiv:1501.07196* (2015). 2

[HCLN09] HUANG S.-H., CHU Y.-H., LAI S.-H., NOVAK C.: Learning-based vertebra detection and iterative normalized-cut segmentation for spinal MRI. *IEEE Trans. on Med. Imag. 28*, 10 (oct. 2009), 1595 –1605. 4

[HP03] HAHN H., PEITGEN H.-O.: IWT - Interactive Watershed Transform: A hierarchical method for efficient interactive and automated segmentation of multidimensional gray-scale images. In *Proc. SPIE Medical Imaging* (2003), vol. 5032, pp. 643–653. 2, 4, 6

[IDL*13] IVANOVSKA T., DOBER J., LAQUA R., HEGENSCHEID K., VÖLZKE H.: Pharynx segmentation from mri data for analysis of sleep related disoders. In *Advances in Visual Computing*, vol. 8033. Springer, 2013, pp. 20–29. URL: http://dx.doi.org/10.1007/978-3-642-41914-0_3, doi:10.1007/978-3-642-41914-0_3. 6

[Jol86] JOLLIFFE I. T.: *Pincipal Component Analysis*. Springer-Verlag, 1986. 3

[KOE*09] KLINDER T., OSTERMANN J., EHM M., FRANZ A., KNESER R., LORENZ C.: Automated model-based vertebra detection, identification, and segmentation in CT images. *Medical Image Analysis 13*, 3 (2009), 471 – 482. 4

[KWZ*13] KELM M. B., WELS M., ZHOU K. S., SEIFERT S., SUEHLING M., ZHENG Y., COMANICIU D.: Spine detection in CT and MR using iterated marginal space learning. *Medical Image Analysis 17*, 8 (Dec 2013), 1283–1292. doi:10.1016/j.media.2012.09.007. 4

[LUO*03] LIU J., UDUPA J. K., ODHNERA D., MCDONOUGH J. M., ARENS R.: System for upper airway segmentation and measurement with mr imaging and fuzzy connectedness. *Academic radiology 10*, 1 (2003), 13–24. 6

[MIJ08] MINGQIANG Y., IDIYO K. K., JOSEPH R.: A survey of shape feature extraction techniques. In *Pattern Recognition* (Nov. 2008), Yin P.-Y., (Ed.), IN-TECH, pp. 43–90. 2

[NZZW10] NING J., ZHANG L., ZHANG D., WU C.: Interactive image segmentation by maximal similarity based region merging. *Pattern Recognition 43*, 2 (Feb 2010), 445–456. doi:10.1016/j.patcog.2009.03.004. 2

[SCHH13] SCHWIER M., CHITIBOI T., HÜLNHAGEN T., HAHN H.: Automated spine and vertebrae detection in CT images using object-based image analysis. *Int J Numer Method Biomed Eng 29*, 9 (2013), 938–963. 2, 6

[SK13] SCHULTZ T., KINDLMANN G. L.: Open-box spectral clustering: Applications to medical image analysis. *IEEE Trans. Vis. Comput. Graph. 19*, 12 (2013), 2100–2108. 2

[SMH10] SAAD A., MÖLLER T., HAMARNEH G.: Probexplorer: Uncertainty-guided exploration and editing of probabilistic medical image segmentation. In *IEEE - VGTC Conference on Visualization* (2010), EuroVis'10, Eurographics Assoc., pp. 1113–1122. URL: http://dx.doi.org/10.1111/j.1467-8659.2009.01691.x, doi:10.1111/j.1467-8659.2009.01691.x. 2

[SNS*98] SATO Y., NAKAJIMA S., SHIRAGA N., ATSUMI H., YOSHIDA S., KOLLER T., GERIG G., KIKINIS R.: Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. *Medical Image Analysis 2*, 2 (1998), 143 – 168. 2

[SSQW07] SHI R., SUN D., QIU Z., WEISS K.: An efficient method for segmentation of MRI spine images. In *IEEE - ICME* (may 2007), pp. 713 –717. doi:10.1109/ICCME.2007.4381830. 4

[TM03] TEOH S. T., MA K.-L.: Starclass: Interactive visual classification using star coordinates. In *SDM* (2003), Barbará D., Kamath C., (Eds.), SIAM. 3

[TP11] TANG Z., PAULI J.: Fully automatic extraction of human spine curve from MR images using methods of efficient intervertebral disk extraction and vertebra registration. *International Journal of Computer Assisted Radiology and Surgery 6* (2011), 21–33. 4

[VLBK*13] VON LANDESBERGER T., BREMM S., KIRSCHNER M., WESARG S., KUIJPER A.: Visual analytics for model-based medical image segmentation: Opportunities and challenges. *Expert Syst. Appl. 40*, 12 (Sept. 2013), 4934–4943. URL: http://dx.doi.org/10.1016/j.eswa.2013.03.006, doi:10.1016/j.eswa.2013.03.006. 2

[YOS06] YAO J., O'CONNOR S., SUMMERS R.: Automated spinal column extraction and partitioning. In *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on* (april 2006), pp. 390 –393. 4

[ZX13] ZHAO F., XIE X.: An overview of interactive medical image segmentation. *Annals of the BMVA 7* (2013), 1–22. 2

© The Eurographics Association 2015.

Table 2: Statistics of experiments of vertebrae detection on saggital CT dataset No. 19. Shows for each user the time required for training, the number of true positive (TP) and false positive (FP) samples obtained, as well as the precision (prec). The average (AVG) and standard deviation (SD) over all users are provided.

	user	1	2	3	4	5	6	7	8	**AVG**	**SD**
	time[s]	90	245	140	106	142	365	92	95	**159**	**100.94**
before post-processing	TP	69	103	64	57	33	113	18	34	**61.38**	**36.17**
	FP	12	23	1	4	16	26	7	6	**11.88**	**9.82**
	prec.	85.19%	81.75%	98.46%	93.44%	67.35%	81.29%	72%	85%	**83.06**%	**10.97**%
after post-processing	TP	69	103	64	57	33	111	18	34	**61.13**	**35.69**
	FP	3	7	1	0	4	6	1	3	**3.13**	**2.67**
	precision	95.83%	93.64%	98.46%	100%	89.19%	94.87%	94.74%	91.89%	**94.83**%	**3.69**%

Table 3: Total number of segments and number of positive segments classified to belong to pharynx in axial MRI datasets. The average (AVG) and standard deviation (SD) are provided.

dataset	1	2	3	4	5	6	7	8	9	10	**AVG**	**SD**
all	10,245	11,360	9,571	10,760	10,106	8,533	7,250	9,555	9,024	10,490	**9,689**	**1,132**
positive	26	12	12	21	32	17	21	22	21	27	**21.1**	**6**

© The Eurographics Association 2015.