Multi-organ Localization Combining Global-to-Local Regression and Confidence Maps

Romane Gauriau[1,2], Rémi Cuingnet[1], David Lesage[1], and Isabelle Bloch[2]

[1] Philips Research MediSys, Paris, France
[2] Institut Mines-Telecom, Telecom ParisTech, CNRS LTCI, Paris, France

Abstract. We propose a method for fast, accurate and robust localization of several organs in medical images. We generalize global-to-local cascades of regression forests [1] to multiple organs. A first regressor encodes global relationships between organs. Subsequent regressors refine the localization of each organ locally and independently for improved accuracy. We introduce *confidence maps*, which incorporate information about both the regression vote distribution and the organ shape through probabilistic atlases. They are used within the cascade itself, to better select the test voxels for the second set of regressors, and to provide richer information than the classical bounding boxes thanks to the shape prior. We demonstrate the robustness and accuracy of our approach through a quantitative evaluation on a large database of 130 CT volumes.

1 Medical Motivation and Overview

With the ever growing size and complexity of $3D$ medical acquisitions, automatic, robust and accurate anatomy localization is of prime interest. First, it enables faster data navigation and visualization of target structures. Secondly, organ localization is a key initialization step for tasks such as segmentation. It is, overall, a crucial component to complex workflows such as treatment follow-up.

General object detection has been deeply studied in computer vision. However algorithms proposed for natural $2D$ scenes are usually not efficient enough (exhaustive scanning of the image) or not even applicable (from $2D$ to $3D$) to the case of anatomical objects. Moreover, medical images often hold specific contextual information, which entails to design specific methods. In the literature we can mostly find three types of approaches for multi-organ localization: classification, regression and atlas-based approaches. As shown in [2], regression-based methods are computationally less expensive (about 25 times less) than atlas-based ones, and then more adapted to clinical contexts. A good overview of the different classification and regression approaches proposed so far can be found in [3]. In this paper we focus on regression-based methods, as their speed and accuracy [2] make them well adapted to clinical contexts. The idea of these approaches is to learn a regression function which relates a voxel and its associated image features to a set of parameters that we want to predict (e.g. organ bounding box). We say that a voxel votes for a set of parameters. The votes from several voxels form a distribution from which we can infer the final result. In [2]

P. Golland et al. (Eds.): MICCAI 2014, Part III, LNCS 8675, pp. 337–344, 2014.
© Springer International Publishing Switzerland 2014

the authors have developed a multivariate regression approach where the organs bounding boxes positions are predicted from voxel locations. The authors of [1] enriched this approach by performing a cascade of locally trained regressors. In both works, bounding boxes give a very rough approximation of the target organs (e.g. the liver). In [4] the authors increase the organs parameterization complexity. They perform a joint anatomical landmarks detection and then align shape models. This method gives very good and fast results on shapes such as the lungs or the kidneys, but its application on organs like the stomach or the gallbladder remains questionable, as specific landmarks may be difficult to define.

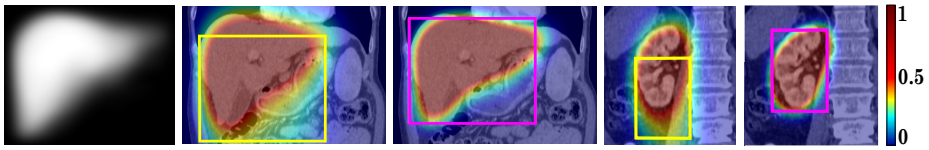
In these methods based on predictions, the authors often use the maximum, the median or the mean of the votes as a final result [2,5]. The vote distributions are not taken into account even though they hold precious information. In this work we intend to make a deeper use of the vote distributions. Following the idea of [6] we also propose to condition the image voxels membership to a global shape prior represented by probabilistic atlases. In Sect. 2 we develop this idea by introducing the concept of *confidence maps* that can be seen as weighted vote distributions associated with shape priors, and propose a fast implementation to compute them. Then we present our global-to-local prediction framework taking benefit of these maps. In Sect. 3 we show different aspects of our contribution: the benefit of performing a greedy parameter optimization and the evaluation of our approach on a large database, thus demonstrating the interest of the *confidence maps* as part of the localization framework and as a result in itself.

2 Methodology

Merging Shape Priors and Vote Distributions with Confidence Maps. In the main works on organ localization with regression, the spatial vote distributions are not fully exploited. However, as shown with the Hough forests [7], vote aggregation can give more information than a single measure of the distribution. For this purpose we introduce the notion of *confidence map*, which encodes the confidence in finding a target organ at a given location. It is built through an aggregation process, making use of both the spatial distribution of regression votes and of organ shape priors through probabilistic atlases.

Probabilistic Atlas. To compute a probabilistic atlas of an organ we first register the binary masks of several samples of this organ. For this purpose let $\{M_i\}_{i\in[\![1,N]\!]}$ be the set of N different cropped binary masks of the organ such that $\forall i\ M_i : \Omega \to \mathbb{R}$ (Ω being the image volume) and where M_r ($r \in [\![1, N]\!]$) is a mask of reference chosen arbitrarily. We transform each mask with a transformation T_i (rigid and anisotropic scaling) in order to scale up the masks M_i to the same size as M_r. The probabilistic atlas A is then computed as an average of these masks, given by $A(x) = \frac{1}{N}\sum_{i=1}^{N} M_i \circ T_i(x)$. Each value $A(x)$ evaluates the probability of a voxel x to belong to the organ (see Fig. 1a).

Confidence Map. Let us consider a regressor $\boldsymbol{R}_\theta$ which, for a test voxel v and its associated image features θ, can give a prediction of the bounding box position



(a) (b) (c) (d) (e)

Fig. 1. Atlas of the liver (a), localization of the liver and the right kidney: images with confidence map as overlay (images have been cropped) and predicted median box after global step (b,d) and local step (c,e) (best viewed in color)

and scale of an organ with a confidence score α. We parameterize a bounding box by $\mathbf{b} = (b_{min}, b_{max}) \in \mathbb{R}^6$ where b_{min} and b_{max} are its extremal vertices. We denote $X_{\mathbf{b}}$ the set of voxels included in the bounding box with parameterization $\mathbf{b}$. We introduce $\{\mathbf{b}_k\}_{k\in[\![1,K]\!]}$ an ensemble of K votes with confidence scores $\{\alpha_k\}_{k\in[\![1,K]\!]}$. The confidence map C for a given organ is built by localizing and scaling the organ's probabilistic atlas A according to each vote k and by accumulating the result in C with weight α_k. The map C gives a confidence score about the presence of the organ at a given location in the image. The pseudo-code is given in Algo. 1. Some examples of confidence maps are given in Fig. 1(b-e). Figures 1(b,d) show that the maps capture the ambiguity of the vote distribution, as we observe that some voxels were correctly voting for the box upper wall position, an information that the median was not able to capture. The computation of these maps may be expensive. Therefore we propose a fast implementation which considerably reduces the computation time while not degrading significantly the accuracy (see Sect. 3). The idea is to uniformly discretize the space of predicted bounding boxes dimensions. As detailed in Algo. 2, considering we have J discrete values, for each j^{th} discretization value we convolve the volume C'_j, where associated center of votes have been set to their confidence score, with the probabilistic atlas A_j resized to the corresponding sample dimension. The final map C is computed as the sum of each C'_j (we normalize by the maximum). With a uniform discretization on 27 boxes dimensions (3 per spatial dimension, which is good trade-off between speed and precision), the computation is about 30 times faster.

Localizing Organs with Regressors and Confidence Maps. Our approach consists of two steps in a global-to-local fashion. A first regressor, aiming at capturing the spatial relationships between organs, is learned using global information: voxels of the whole image can vote for the positions of all the organs simultaneously. Then, new regressors, dedicated to each organ, are learned using more local information. The benefit of the cascade approach has been already shown in [1]. Here we propose to introduce the use of confidence maps for refining the votes in the cascade of regressors. The selection of voxels which vote in the local step may benefit of the information given by the confidence maps, that is to say the vote distributions and the shape prior. Notice that this method can be applied with any multi-variate regressor. We denote by $o \in [\![1, N_{\text{org}}]\!]$ the

$C \leftarrow 0$;
for $k \leftarrow 1$ **to** K **do**
  $[b_{min}, b_{max}, \alpha] \leftarrow \boldsymbol{R}_\theta[v[k]]$;
  $X_b \leftarrow \texttt{computeXb}(b_{min}, b_{max})$;
  $c \leftarrow \frac{1}{2}(b_{min} + b_{max})$;
  $d \leftarrow \texttt{norm}(b_{min},b_{max})$;
  $A_t \leftarrow \texttt{resizeToBoxDim}(A,d)$;
  foreach $x \in X_b$ **do**
    $C[x] \leftarrow C[x] + \frac{\alpha}{K} A_t(x - c)$;

Algo. 1. Confidence map computation

$C \leftarrow 0$, $d_{max} \leftarrow 0$, $d_{min} \leftarrow$max ;
for $k \leftarrow 1$ **to** K **do**
  $[b_{min}, b_{max}, \alpha[k]] \leftarrow \boldsymbol{R}_\theta[v[k]]$;
  $c \leftarrow \frac{1}{2}(b_{min} + b_{max})$;
  $d[k] \leftarrow \texttt{norm}(b_{min},b_{max})$;
  $d_{max} \leftarrow \texttt{max}(d_{max},d)$;
  $d_{min} \leftarrow \texttt{min}(d_{min},d)$;
$\hat{d} \leftarrow \texttt{quantize}(d_{max},d_{min},J)$;
foreach $\hat{d_j} \in \hat{d}$ **do**
  $indices \leftarrow \texttt{getIndicesOfVotes}(\hat{d_j},d)$;
  $C' \leftarrow 0$;
  foreach $ind \in indices$ **do**
    $C'[ind] \leftarrow C'[ind] + \alpha[ind]$;
  $A_j \leftarrow \texttt{resizeToBoxDim}(A,\hat{d_j})$;
  $C' \leftarrow \texttt{convolve}(C',A_j)$;
  $C \leftarrow C + C'$;

Algo. 2. Fast confidence map computation

indices of the N_{org} organs to localize, and we describe the two steps more precisely hereafter.

Global Step. In the first step, a random subset of K_g voxels $\{v_k\}_{k\in[\![1,K_g]\!]}$ of the image I will vote for the bounding boxes parameters of organ o $\{\mathbf{b}_{k,o}\}_{k\in[\![1,K_g]\!]}$. Votes are performed according to long-range features computed from the image (see Sect. 3 for specific application with regression forest as regressor). These features are chosen to encapsulate global information from the image. Notice that the regressor should be designed such that the relationships between the organs are implicitly embedded during learning (by preserving correlation information between organs positions). Then we compute the confidence map C_o for each organ o using Algo. 1 or 2 and given a probabilistic atlas A_o.

Local Step. The second step aims at improving the previous localization. We relocalize each organ o individually by first computing the binary mask B_o built from the map C_o at a threshold t_g (see Sect. 3). Then we select a random subset of K_l voxels $\{v_k\}_{k\in[\![1,K_l]\!]}$ such that each voxel $B_o(v_k) = 1$. Each voxel v_k votes for the parameters of organ o using a regressor specifically trained for this organ. Contrarily to the previous step this predictor is now learned using short-range features (see Sect. 3) and computed in the vicinity of the organ o thanks to the confidence map so that we give more importance to local information. We then use the votes to compute new and more accurate confidence maps C'_o for each organ. Figures 1(c,d) show the benefit of this local step.

3 Experiments and Results

To validate our approach we propose to localize 6 abdominal organs: the liver, the kidneys, the gallbladder, the spleen and the stomach from various types of 3D CT volumes and using a regression forest as a regressor.

Using the Multivariate Regression Forest as a Regressor. Regression forests have been introduced in [8] and recently popularized in [2] for the purpose of multi-organ localization. This method has proved to be very fast and

quite accurate. As in [2] we use the random forest to regress the parameters **b** of each organ. Random trees are learned from a subset of test voxels from the training images. Each node of each tree contains a $1D$ feature and in each leaf the distributions of the parameters to regress are stored (here multivariate Gaussian distributions). We refer the reader to [2] for more details on the method. This approach is well suited to rough localization. We are able to reinforce its robustness and accuracy using our global-to-local approach with confidence maps. For the first step of the algorithm we use long-range features computed from the image after Gaussian smoothing. As in [1,2], a $1D$ feature corresponds to the difference of mean intensities in two $3D$ patches of random sizes and locations in a certain range. The statistical information stored in the leaf of the random forest regressor allows us to compute confidence scores for each vote. Notice that in practice votes with low confidence scores are discarded. Here we point out that the forest implicitly encodes the organs relationships, as each tree has been built considering the entire set of parameters from all the organs. In the second step we use one random regression forest per organ while using the same kind of features in a more local range (see Sect. 3). Each of these regression forests is learned from test voxels in the vicinity of the corresponding organ thanks to the confidence maps built after the global step. Votes are thus explicitly restricted to a certain neighborhood around each organ.

Database Description and Implementation. Our database is composed of 130 3D CT images coming from 112 patients with diverse medical conditions (healthy and pathological subjects, no organ missing). It includes volumes with varied fields of view, body shapes, resolution and use or not of contrast agents. Slices and inter-slices resolution ranges from 0.5 to 1 mm and from 0.5 to 3 mm, respectively. All the organs have been manually segmented in these 130 volumes. The dataset has been split randomly into 50 and 80 volumes for training and testing, respectively. Our method was implemented in C++ and running times are given for a machine with four 2.3 GHz cores and 8 Go RAM.

Off-line Training. To reach the best performances and analyze each aspect of the algorithm, we performed an extensive greedy optimization of the algorithm parameters. Learning one tree with depth 12 takes about 2 minutes. Before learning we decorrelate the data with a whitening transformation.

Greedy Parameters Optimization. For each parameter we performed a 3-fold cross-validation on the training set. The accuracy of the algorithm is measured as the mean distance of the predicted box to the ground truth bounding box. We first initialized every parameter arbitrarily. Then we optimized each parameter one-by-one by grid-search and we replaced its value by the optimized one. Concerning the training parameters we optimized the tree depth, the threshold of the confidence map t_{g1} and the range of the features. We got the best performances with tree depths of 14 and 12 for the global and local steps respectively, $t_{g1} = 0.4$ and the range of features $\Delta_g = [0, 70]^3$, $W_g = [0, 70]^3$ and $\Delta_l = [0, 40]^3$,

$W_l = [0, 40]^3$ for global and local steps respectively (where Δ is the range of distances from the test voxel to the $3D$ patches and W is the range of sizes of the $3D$ patches, all sizes given in millimeters). Regarding the testing parameters, we looked for the best number of votes (with $K_g = 30000$ and $K_l = 10000$) for the final prediction and the best threshold t_{g2} of the confidence map after the global step. Setting $t_{g2} = 0.5$ and keeping respectively 3% and 1% of the votes with best confidence gave the best results.

Learning Phase. The best parameters found after the above optimization were used for the final forests learning on the 50 training volumes. Atlases of each organ were learned on the same dataset. As in [8] we perform bagging for learning all the forests (uniform random draw with replacement). Node optimization was performed with 30 feature tries as a good compromise between speed and accuracy. Hereafter, if not specified, we used 3 trees as it achieves a good compromise between computation time and accuracy.

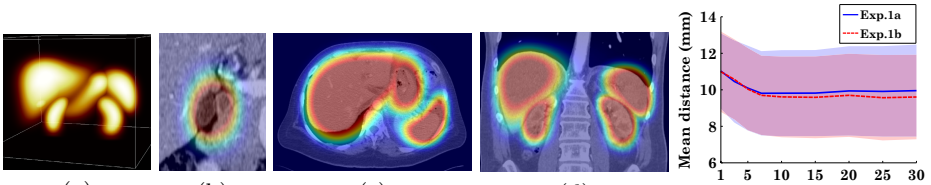
Evaluation of the Localization Approach. The first objective of our evaluation is to show that the confidence maps can be used as a localization result itself, giving more consistent information than the bounding boxes alone. Confidence maps are computed at a 5 mm isotropic spacing. We give some examples of localization results in Fig.2(a-d). An exhaustive visualization of the results can be seen in the **supplementary material**[1]. The lines of Table 1 starting with 'MD' give the mean distance of the thresholded maps contours ($t = 0.5$) to the ground truth contours, using Algo. 1 and 2. The statistics of the results (median and standard deviation) confirm that our method is robust to the variety of test images. We also show that our fast implementation degrades the overall performance very slightly. This allows us to think that our approach can be very useful in various contexts such as segmentation initialization, added to the fact that it runs in about 10 seconds which makes it adapted to clinical applications (code optimization may still improve the computation time). Moreover the confidence maps give much more information than a simple binary mask or contour. For that reason we propose adapted evaluation measures taking into account the fuzziness of the maps. If C denotes a confidence map and B a binary mask of the organ ground truth, then the true positive values are defined as $TP = \sum_{x\in\Omega} B(x)C(x)$, the false negative values as $FN = \sum_{x\in\Omega} B(x)(1 - C(x))$ and the false positive values as $FP = \sum_{x\in\Omega}(1 - B(x))C(x)$. Following the definitions of the sensitivity $S = TP/(TP + FN)$ and the precision $P = TP/(TP + FP)$, we are able to compute the weighted versions of these measures. The corresponding results are reported in Table 1. We notice lower performances for the stomach and the gallbladder, which are challenging organs due to their shape variability. However the figures show that they are still correctly detected.

The second objective of our evaluation comprises two experiments in which we only change the way of selecting the test voxels $\{v_k\}_{k\in[\![1,K_l]\!]}$ for the second local step: exp.1a: from the predicted boxes after the global step, exp.1b: from

[1] `http://perso.telecom-paristech.fr/~gauriau/files/MICCAI14_SupMat.pdf`

Table 1. Results with confidence maps (5 mm isotropic spacing) with Algo.(1, 2) with measures MD: mean distance (mm), S: weighted sensitivity (%), P: weighted precision (%) (mean±standard deviation (median))

		Liver	L. Kidney	R. Kidney	Spleen	Gallbladder	Stomach	**All organs**	Time
Algo.1	MD	9.6 ± 3(9)	5.6 ± 3(5)	6.1 ± 3(6)	7.9 ± 4(7)	9.4 ± 6(7)	13.5 ± 6(12)	8.7 ± 3(8)	
	S	76.8 ± 5(77)	65.7 ± 8(68)	64.6 ± 8(66)	66.3 ± 11(69)	45.2 ± 19(48)	49.7 ± 9(51)	61.4 ± 15(65)	300s
	P	78.8 ± 7(79)	85.2 ± 7(87)	85.1 ± 5(87)	80.8 ± 10(82)	52.4 ± 22(59)	68.3 ± 12(69)	75.1 ± 17(80)	
Algo.2	MD	9.8 ± 3(9)	6.0 ± 4(5)	6.3 ± 3(6)	8.5 ± 4(7)	9.6 ± 4(7)	13.8 ± 6(13)	9.0 ± 3(8)	
	S	75.9 ± 6(77)	64.4 ± 9(66)	63.8 ± 9(66)	64.9 ± 11(67)	44.1 ± 19(46)	49.3 ± 9(50)	60.4 ± 15(64)	10s
	P	78.2 ± 7(79)	84.5 ± 7(86)	85.0 ± 5(87)	80.2 ± 10(82)	52.1 ± 21(57)	67.4 ± 13(68)	74.6 ± 17(80)	



(a) (b) (c) (d) (e)

Fig. 2. 3*D* MIP of final confidence maps (a), some results with Algo.2 with overlaid confidence maps on the images (b,c,d) and results of exp.1(a,b) (e)

the confidence map using Algo. 2. Figure 2e shows that above 10, the number of trees does not significantly improves the results. Then we keep this number of trees for exp.1(a,b). In Table 2 we report the results of [1,2] and ours after global step and exp.1(a,b). The performances after the global step show that a fine optimization of the parameters helps reaching better results than in [2]. Exp.1a shows the benefit of the cascade approach (additional iterations did not show significant improvements of the results), compared to the single step one. Exp.1b shows the difference between our method and the original cascade approach. The improvement is specially significant for the liver. Moreover Fig. 2 shows that the use of confidence maps tends to rather decrease the bias of the results with an increasing number of trees. We also get similar results to those of [1] for the kidneys, which shows that the cascade approach is scalable and that an increasing number of organs does not degrade the average performance.

Table 2. Box walls mean distances per organ (mean distance (mm) ± standard deviation (median)), per method and per experiment

Method		Liver	L. Kidney	R. Kidney	Spleen	Gallbladder	Stomach	**All organs**	Time(∼)
	[2]*	15.7 ± 15	13.6 ± 13	16.1 ± 16	15.5 ± 15	18.0 ± 15	18.6 ± 16	16.3	4s**
	[1]*	-	7 ± 10(5)	**7 ± 6(6)**	-	-	-	-	3s**
Local	Global	12.5 ± 4(12)	13.6 ± 7(13)	13.8 ± 5(12)	14.3 ± 6(14)	13.9 ± 6(12)	14.3 ± 6(14)	13.9 ± 6(13)	1s
	Exp.1a	11.8 ± 4(11)	6.9 ± 6(5)	**7.2 ± 3(7)**	9.6 ± 7(8)	**9.8 ± 5(9)**	13.6 ± 5(13)	9.8 ± 6(9)	3s
	Exp.1b	**10.5 ± 4(10)**	**6.8 ± 6(5)**	7.3 ± 3(7)	**9.6 ± 6(8)**	10.0 ± 5(8)	**13.5 ± 5(13)**	**9.6 ± 5(8)**	4s

* results are given for other datasets than ours ** times are given for different machines than ours

4 Conclusion

In this article we proposed a fast, robust and accurate method for the localization of multiple organs. We extended the idea of cascade of regressors while introducing the concept of confidence map, which models the vote distributions with the addition of shape prior. We showed that the confidence map, with a proposed fast implementation, can enhance the consistency and accuracy of multi-organ localization with a limited additional computational cost. It is a generic tool with promising potential, which can be used with any type of regressor and which is adaptable to different modalities (e.g. CT, MRI). Moreover its fuzziness property may be useful in many types of clinical applications, such as segmentation (for initialization) or visualization (to target the objects of interest for $3D$ rendering) for instance. Therefore the perspectives are numerous. We also showed that an extensive optimization of the regression parameters significantly improves the localization results. Finally, the consistency and accuracy of our method may still be improved with the use of multiple probabilistic atlases per organ and the regression of the rotation parameters.

Acknowledgments. Work supported in part by an ANRT grant (008512012).

References

1. Cuingnet, R., Prevost, R., Lesage, D., Cohen, L.D., Mory, B., Ardon, R.: Automatic detection and segmentation of kidneys in 3D CT images using random forests. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 66–74. Springer, Heidelberg (2012)
2. Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., Siddiqui, K.: Regression forests for efficient anatomy detection and localization in computed tomography scans. Medical Image Analysis 17(8), 1293–1303 (2013)
3. Zhou, S.: Discriminative anatomy detection: Classification vs regression. Pattern Recognition Letters (in press)
4. Lay, N., Birkbeck, N., Zhang, J., Zhou, S.K.: Rapid multi-organ segmentation using context integration and discriminative models. In: Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., Zöllei, L. (eds.) IPMI 2013. LNCS, vol. 7917, pp. 450–462. Springer, Heidelberg (2013)
5. Zhou, S.K., Comaniciu, D.: Shape regression machine. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 13–25. Springer, Heidelberg (2007)
6. Sun, M., Kohli, P., Shotton, J.: Conditional regression forests for human pose estimation. In: CVPR, pp. 3394–3401 (2012)
7. Gall, J., Lempitsky, V.: Class-specific hough forest for object detection. In: CVPR, pp. 1022–1029 (2009)
8. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)