

Automatic Annotation of Liver CT Image: ImageCLEFmed 2015

Imane Nedjar¹, Saïd Mahmoudi², Mohamed Amine Chikh¹, Khadidja
Abi-yad³, Zouheyr Bouafia⁴

¹Biomedical Engineering Laboratory, Tlemcen University, Algeria

²Computer Science Department, University of Mons, Belgium

³Department of Pathology, University Hospital Center of Tlemcen, Algeria

⁴Telecommunication Laboratory, Tlemcen University, Algeria

{imane.nedjar, ma_chikh, khadidja.abi-yad, Zouheyr.Bouafia
}@mail.univ-tlemcen.dz, said.mahmoudi@umons.ac.be

Abstract. In this paper, we present the methods that we have proposed and used in the liver image annotation task of ImageCLEF 2015. This challenge entailed the annotation of liver CT scans to generate a structured report. To meet this challenge we have proposed two methods for annotating the liver image. The first one uses a classification approach, which is composed of two main phases. The first step consists of a pre-processing, where a texture and shape based features vector is extracted, in the second phase a classification process is achieved by using random forest classifier with two different sets of features. Our second method uses a specific signature of the liver. Indeed, we have taken a slice from 3D liver CT scans, thereafter we have normalized it into a rectangular block with constant dimensions to account for imaging inconsistencies, and then we have divided the block into small blocks. After applying the 1D Log-Gabor filters transformation, the dominant phase data of each block was extracted and quantized to four levels to encode the unique pattern of the liver into a bit-wise template. The Hamming distance was employed for retrieval. We submitted 3 runs to the liver image annotation task of ImageCLEF 2015 and we obtained the following scores (90.4%, 90.2%, and 91%).

Keywords: Image Annotation, Liver, Random Forest Classifier, Image Retrieval, Computer-Aided Diagnosis.

1 Introduction

The digital imaging revolution in the medical domain over the past three decades has changed the way the present-day physicians diagnose and treat diseases. The focus now includes more effective post-processing, organization, and retrieval. In this context, a major challenge is to support the clinical decision making by retrieving and displaying the relevant cases using all available information, such as structured reports.

The structured reports are highly valuable in medical contexts due to the processing opportunities that they provide, such as reporting, image retrieval, and computer-aided diagnosis systems. However, structured reports are time consuming since they need a lot of time to be created. Furthermore, their creation requires high domain expertise, which is time constrained. Consequently, such structured medical reports are often not found or are incomplete in practice [1].

The aim of the liver annotation task [2] is to improve computer-aided automatic annotation of liver CT volumes by filling in a structured radiology report. The main goal of this task is describing the semantic features of the liver, its vascularity, and the types of lesions in the liver.

One of the major challenges of this work was the limited amount of training data compared to the number of annotations to be recognized. In particular, there were some annotations that did not occur at all in the dataset. Similarly, there were also some instances having the same annotation for all training samples [3]. To overcome this issue we present in this paper two methods based on the visual features and the information extracted from the Liver Case Ontology (LiCO) [4].

This paper is organized as follows: section two presents the principles related works in this area. Section 3 presents the dataset used. The proposed methods are presented in section 4. Section 5 provides a discussion on the experimental results. Finally, section 6 gives a conclusion to the work.

2 Related Works

ImageCLEF is the image retrieval track of the Cross Language Evaluation Forum (CLEF) [5] [5]. In 2014, for the first time, the liver CT annotation task was proposed [6]. Three groups have participated in this challenge, the BMET group from the School of Information Technologies at the University of Sydney (Australia), and the second group is CASMIP from The Hebrew University of Jerusalem (Israel). The third participant was piLabVAVlab from Boğaziçi University (Turkey).

The BMET group [3] proposes two strategies for annotating the liver images. The first method uses multi-class classification scheme where each label has a classifier that is trained to separate it from other labels. They used two stages of classification, each one consisting of a bank of several support vector machines (SVM). The first stage is composed of the 1-vs-all classifiers, and the second stage consisted of the 1-vs-1 classifiers. The second method uses the similarity scores from an image retrieval algorithm as weights for a majority voting scheme, thereby reducing the inherent bias towards labels that have a high number of samples. The BMET group submitted eight runs. Four of them used a classifier based approach, and the remaining used an image retrieval algorithm. All runs achieved high scores (>90%), and they also achieved the highest score of 94.7% out of all the submission done to imageCLEF2014.

On the other side, CASMIP group [7] tried four different classifiers in the learning phase: linear discriminant analysis (LDA), logistic regression (LR), K-nearest neighbors (KNN), and finally SVM, to predict labels. An exhaustive search of every combination of image features is done using leave-one-out cross validation method on

training data for every label and classifier. As result, for each label the best classifier and its related features were learned. The learning step was performed using all labels of the training dataset except cluster size, lobe and segments, which were obtained directly from image features. Python scikit-Learn Machine learning toolbox was used for implementing each classifier with default parameters. As result, for most of the labels they got almost the same performance by using any classifier and any combination of image features. The group submitted one run to the task, and they obtained a score of 93%, which achieved the second best performance.

The piLabVAVlab group [8] proposed an approach based on probabilistic interpretation of tensor factorization models, i.e. Generalized Coupled Tensor Factorization (GCTF). This method can simultaneously fit a large class of matrix/tensor models to higher-order matrices/tensors with common latent factors using different loss functions. piLabVAVlab considered the dataset as heterogeneous data and the GCTF approach was applied to predict labels. They considered KLdivergence and Euclidean-distance based cost functions as well as the coupled matrix factorization models by using the GCTF framework. The group submitted three runs to the task, and their highest score was of 67.7%.

3 Material

The training dataset that we use in our work includes 50 cases, each of them consisting of:

- A cropped CT image of the liver -a 3D matrix. The volumes had varied resolutions (x: 190-308 pixels, y: 213-387 pixels, slices: 41-588) and spacing (x, y: 0.674-1.007mm, slice: 0.399-2.5mm),
- A liver mask that specifies the part corresponding to the liver - a 3D matrix indicating the liver areas with the value 1 and nonliver areas with 0,
- A bounding box (ROI) corresponding to the region of the selected lesion within the liver - as a vector of 6 numbers corresponding to the coordinates of two opposite corners,
- A user expressed features generated by Liver Case Ontology (LiCO) and stored in RDF file.

The test dataset contained 10 CT volumes, with varied resolutions and pixel spacing, cropped to the region around the liver. The test data also included a mask of the liver pixels and a bounding box (ROI).

4 Methods

For this challenge, we proposed two methods for annotating the liver image. The first one uses a classification approach, and the second method uses the signature of the liver. First of all we have extracted the user expressed (UsE) features from the ontolo-

gy of liver cases (LiCO) by using OWL API¹. The two methods are presented in the following sections.

4.1 Method 1:

This method is composed of two main phases. The first step consists of a pre-processing, where a texture and shape based features vector are extracted. In the second phase a classification process is achieved by using random forest classifier.

Features extraction and description : unlike database used in the liver image annotation task of ImageCLEF 2014 that contained a set of 60 computer generated (CoG) features obtained from interactive segmentation software, the database of liver image annotation task 2015 does not include the CoG. For this reason a set of shape and texture features are extracted from both lesion and liver.

The proposed liver descriptor includes:

1. Liver Mean: liver's mean intensity value.
2. Liver Variance: liver's variance intensity value.
3. Liver Skewness : liver's skewness value.
4. Liver Kurtosis: liver's kurtosis value.
5. Liver Solidity: solidity of liver.
6. Liver Convexity: convexity of liver.
7. Haralick's texture features: we have used the following features extracted from the gray level co-occurrence matrix: contrast, entropy, variance, sum mean, correlation, max probability, inverse variance, inertia [9], and also energy, cluster shade, cluster prominence and homogeneity proposed in [10]. We have calculated the 3D gray level co-occurrence matrix for four different directions ($\theta \in \{0^\circ, 90^\circ, 45^\circ, \text{ and } 135^\circ\}$) with a distance $d=1$. Therefore, our GLCM based feature vector includes 48 elements.
8. 3D Gabor wavelet transform: we have used the mean and the standard deviation as texture features, with eight orientations and three scales, so that the feature vector includes 24 elements for the means and 24 elements for deviation. Therefore, the feature vector is composed of 48 features.

The proposed lesion descriptor

A pre-processing step was applied in order to segment the lesion. To do this, we have applied a morphological operation which is dilatation. This operation was done after thresholding the liver and applying the AND operator between lesion bounding box and the liver mask. Thereafter, we extracted the following features:

1. The Euclidean distance between the centroid of the liver and lesion centroid.
2. The distance between the x coordinate centroid of the liver and the x coordinate-centroid of the lesion.

¹Java API for creating, parsing, manipulating and serialising OWL Ontologies.

3. The distance between the y coordinate centroid of the liver and the y coordinate centroid of the lesion.
4. Surface area of the lesion.
5. The perimeter of the lesion.
6. The circularity C1 of the lesion: this measure always takes a value of 1 for perfect circles[11], it is expressed by the following formula:

$$C1 = \sqrt{\left(\frac{Area}{\pi MaxRadius^2}\right)} \quad (1)$$

7. Dispersion property: the irregularity of the mass is estimated from dispersion property, which identifies the irregular shape characteristics [11]. This value is given by the equation below:

$$Dp = \left(\frac{MaxRadius}{Area}\right) \quad (2)$$

8. Elongation property: the regular oval mass can be differentiated from the irregular by using Elongation [11]. Its value is expressed by the following equation :

$$En = \left(\frac{Area}{(2MaxRadius)^2}\right) \quad (3)$$

9. The circularity C2 of lesion: this value presents how a mass is similar to an ellipse. It is useful in differentiating circular /oval masses from irregular masses. This measure always takes a value of 1 for perfect squares, circles [11]. It is calculated by the equation given below:

$$C2 = \sqrt{\frac{MinRadius}{MaxRadius}} \quad (4)$$

The total dimension of the first descriptor is 111(6+48+24+24+9).

We have also investigated the performances of our method by using a second descriptor containing the texture features of the lesion instead of liver texture features.

The two texture features are the gray level co-occurrence matrix (GLCM) for four different directions ($\theta \in \{0^\circ, 90^\circ, 45^\circ, \text{ and } 135^\circ\}$) with a distance $d=1$. Therefore, our GLCM based feature vector includes 48 elements. The second texture feature is Gabor wavelet transforms, we have used the mean and the standard deviation, with sixteen orientations and five scales, the feature vector includes 80 elements for the means and 80 elements for deviation. Therefore, the total dimension of the second descriptor is 223 (6+48+80+80+9).

Classification In the second phase of our experiments we have used a supervised multi-class classifier based on random forest classifier (RF) and a proposed similarity score calculation.

Random Forest (RF) is a machine learning technique that builds a forest of classification trees where each tree is grown on a bootstrap sample of the data, and the attribute at each tree node is selected from a random subset of all attributes. The final classification of an individual is determined by voting over all trees in the forest. Indeed, there are many advantages of using RF method, that make it an ideal approach for the analysis of biological data. First, it can handle a large number of input attributes - both qualitative and quantitative-. Second, it estimates the relative importance of features in determining classification. Third, RF is fairly robust in the presence of etiological heterogeneity and relatively high amounts of missing data.

For the kind of tree we have used Classification And Regression Tree (CART), and ours RF is composed of 500 tree.

The extracted computer generated (CoG) features and the RF are used to predict the following property separately:

Is Central Localized, Is Contrasted, Is Gallbladder Adjacent, Is Peripheral Localized, Is Subcapsular Localized, Has Lesion Quantity, Has Area Density, Has Area Shape, Has Area Margin Type, Has Density, Has Lesion Contrast Uptake, Has Lesion Contrast Pattern, Has composition, Has Lesion Vein Proximity, Is Located In Segment, Is Close To Vein.

The property "Is Located In Lobe "is estimated according the property " Is Located In Segment ".i.e. if segment is {II,III,IV} the lesion lobe is the left lobe, if the segment {V,VI,VII,VIII} the lesion lobe is the right lobe, and caudate Lobe for segment I. The height and the width of lesion are extracted directly from image.

For remaining Use (see Table1), we used the proposed similarity score calculation between the unannotated image (U) and a training image (T) as the distance between their respective features vectors is given as bellow:

$$Sim(U, T) = \sum_{i=1}^d 1 - \frac{|u_i - t_i|}{v_i} \quad (5)$$

Where v_i was the i -th maximum feature in dataset, u_i was the i -th feature in the feature vector of U , t_i was the i -th feature in the feature vector of T , and d was the dimensionality of the feature set.

Thereafter, we selected the five most similar images, the label that has the majority voting will be assigned to the Use.

Table 1. List of UsE features

Group	Concept	Properties
Vessel	Middle Hepatic Vein	HasLumenDiameter
		HasLumenType
	HepaticVein	HasLumenDiameter
		HasLumenType
	LeftPortalVein	HasLumenDiameter
		HasLumenType
		IsCavernousTransformationObserved
	HepaticPortalVein	HasLumenDiameter
		HasLumenType
		IsCavernousTransformationObserved
	RightPortalVein	HasLumenDiameter
		HasLumenType
		IsCavernousTransformationObserved
	RightHepaticVein	HasLumenDiameter
HasLumenType		
LeftHepaticVein	HasLumenDiameter	
	HasLumenType	
HepaticArtery	HasLumenDiameter	
	HasLumenType	
Liver	RightLobe	HasSizeChange
	LeftLobe	HasSizeChange
	CaudateLobe	HasSizeChange
	Liver	HasLiverPlacement
		HasLiverContour
		HasDensity
		HasLiverDensityChange
	HasSizeChange	
Image	HasPhase	
Lesion	Parenchyma	HasDensity
		HasParenchymaDensityChange
		isCalcified
	Lesion	isDebrisObserved
		isLevelingObserved
		hasWallType
	Wall	isContrasted
		hasCalcification
		isCalcified
	Septa	hasSeptaWidth
hasSeptaDiameter		

4.2 Method 2:

Our second method uses the signature of the liver. To do that, we have taken a slice from 3D Liver CT scans localized at lesion center. First, we normalized the liver into a rectangular block with constant dimensions to account for imaging inconsistencies. The size of the normalized block is (200×190). The retrieval process is illustrated in figure1.

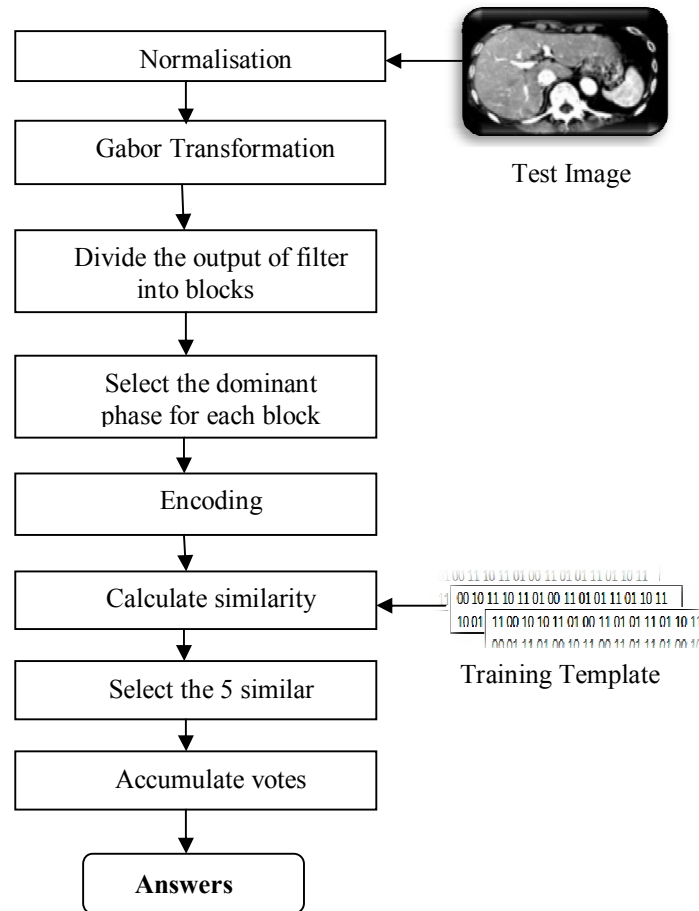


Fig. 1. An overview of the retrieval process

Feature encoding was implemented by convolving the normalized liver pattern with 1D Log-Gabor wavelets. The 2D normalized pattern was broken up into a number of 1D signals, and then these 1D signals were convolved with 1D Gabor wavelets. The frequency response of a Log-Gabor filter is given as:

$$G(f) = \exp\left(\frac{-(\log(f / f_0))^2}{2(\log(\sigma / f_0))^2}\right) \quad (6)$$

Where f_0 represents the centre frequency, and σ gives the bandwidth of the filter.

Thereafter we divided the output of filtering into a small blocks of size (5×5) , consequently the size of template becomes (40×38) .

Finally, the dominant angular direction of each block was extracted and quantized to four levels, using the Daugman method, where each angular direction produced two bits of data [12]. Indeed when going from one quadrant to another, only 1 bit changes. Figure 2 shows the phase quantization.

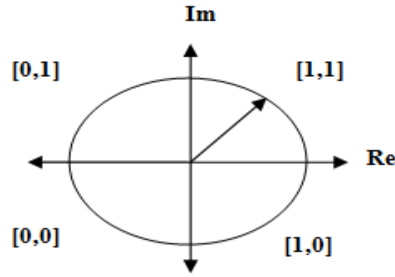


Fig. 2. Phase Quantization

The encoding process produces a bitwise template containing a number of bits of information, the final size of template is 40×76 . For the retrieval task, the Hamming distance was employed. This distance gives a measure of how many bits are the same between two bit patterns.

In comparing with the bit patterns X and Y , the Hamming distance, HD , is defined as the sum of disagreeing bits (sum of the exclusive-OR between X and Y) over N , the total number of bits in the bit pattern.

$$HD = \frac{1}{N} \sum_{j=1}^N X_j (XOR) Y_j \quad (7)$$

Thereafter, we selected the five most similar images, and the label that has the majority voting will be assigned to the Use.

5 Results and Discussion

We submitted three run to the ImageCLEF2015 liver annotation challenge, two of them for the first method with two different feature sets, and the last for the second method. The runs were evaluated on *accuracy*, the percentage of completed questions with a correct answer, and *completeness*, the percentage of question that was answered. Finally, the total score is given as:

$$TotalScore = \sqrt{Completeness \times Accuracy} \quad (8)$$

Table 2. The score obtained for each run

Run	Method	Score
1	1	0.904
2	1	0.902
3	2	0.910

The results show that all of our runs achieved good scores (>90). In general, there were no large difference between the scores, especially for the first method, where we have used two different descriptors (as stated in section 4.1).

Descriptor 1 (run1): include texture features of liver and shape features.

Descriptor 2 (run2): include texture features of lesion and shape features.

The score obtained from the first and the second descriptor are respectively 90.4% and 90.2%, this small difference shows that texture features of liver is more descriptive than the texture features of lesion. We achieved a completeness score of 99% for every run, and accuracy of 84% given by the second method.

Table 3. The test results for each group

Group	Run1	Run2	Run3
Liver	0.925	0.925	0.925
Vessel	1.000	1.000	1.000
LesionArea	0.730	0.746	0.753
LesionLesion	0.470	0.470	0.480
LesionComponent	0.870	0.844	0.889

The results presented in table 4 shows that the scores obtained by the three runs for Liver group and Vessel group are the same. One explanation for this could be that there were instances where all the training samples had the same annotation.

We notice that the second method outperform the first method in the other groups, which shows its efficacy.

For the first method, descriptor 1 (run1) gives a best discrimination of lesion component compared to the descriptor 2 (run2), in this case the texture features of liver is more suitable than texture features of lesion, in the other hand the properties of area lesion were well described by the texture features of lesion.

6 Conclusions

In this paper we have presented the methods submitted to the liver annotation task of imageCLEF2015. In the first method we have used a classification approach, and in the second method we have used a specific signature of the liver.

Our three runs achieved scores (>90), and a completeness level of 99%. There were no large differences between scores, and the best accuracy level is 84% given by the second method.

In our future works we plan to use more image features, and investigate the semantic reasoning by using the information in ONtology of the Liver for RAdiology (ONLIRA).

Acknowledgment

The authors would like to thank the organizers of the ImageCLEF 2015 liver annotation task for making the database available for the experiments.

References

- [1] Marvasti N et al., "ImageCLEF Liver CT Image Annotation Task 2014. In: CLEF 2014 Evaluation Labs and Workshop", Online Working Notes. (2014).
- [2] Marvasti N, Mar Roldan Garcia M, Uskudarli S, Aldana J.F, and Acar B, "Overview of the ImageCLEF 2015 liver CT annotation task," *CLEF2015 Working Notes. Workshop Proceedings.CEUR-WS.org*, no. 1613-0073, September 8-11. 2015.
- [3] Kumar A, Dyer S, Li C, Leong P.H.W, and Kim J, "Automatic annotation of liver ct images: the submission of the bmet group to imageclefmed 2014", in CLEF 2014 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (CEUR-WS.org), September 2014.
- [4] Kökciyan. N, Üsküdarli.S, Yolum. P, Bakir. B, and Acar. B, "Semantic Description of Liver CT Images: An Ontological Approach," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1363 - 1369, 2014.
- [5] Villegas M et al., "General Overview of ImageCLEF at the CLEF 2015 Labs," *Lecture Notes in Computer Science.Springer International Publishing*, no. 0302-9743, 2015.
- [6] Caputo B et al., "ImageCLEF 2014: Overview and analysis of the results", in CLEF proceedings, Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2014.
- [7] Spanier A.B and Joskowicz L, "Towards content-based image retrieval: From computer generated features to semantic descriptions of liver ct scans", in CLEF 2014 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (CEUR-WS.org), September 2014.
- [8] Ermis B and Cemgil A.T, "Liver ct annotation via generalized coupled tensor factorization", in CLEF 2014 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (CEUR-WS.org), September 2014.
- [9] Haralick R M, Shanmugam k, and Dinstein I, "Textural Features for Image Classification," *IEEE Transaction on Systems ,Man and Cybernetics*, vol. 3, no. 6, pp. 610-621, November 1973.
- [10] Soh L.K and Tsatsoulis C, "Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices," *IEEE transactions on Geoscience and Remote Sensing*, vol. 37, no. 2, MARCH 1999.
- [11] Vadivel A and Surendiran B, "A fuzzy rule-based approach for characterization of mammogram masses into BI-RADS shape categories," *Computer in Biology and Medicine*, no. 43, pp. 259-267, 2013.
- [12] Daugman J, "How iris recognition works", Proceedings of 2002 International Conference on Image Processing, Vol. 1, 2002.