Automated lipid droplets recognition in human steatotic liver: some preliminary results

Michele Sciarabba[a*], Maurizio Vertemati[a], Claudia Moscheni[a], Mara Cossa[a] and Laura Vizzotto[a]

[a]Department of Human Morphology and Biomedical Sciences "Città Studi", University of Milan, Milan, Italy

Abstract. The assessment of the degree of steatosis in routine liver biopsies represents an important task in different clinical situations, such as alcoholic steatohepatitis, non-alcoholic fatty liver disease, viral hepatitis, and evaluation of the viability of the graft in liver transplantation. Despite the advances in imaging techniques, microscopic examination remains the gold standard for the assessment of hepatic steatosis. In this study, we developed an automated approach for hepatic steatosis assessment in routine liver biopsies stained with Hematoxylin-Eosin (HE) from patients affected by hepatitis C. We performed a multi-step procedure by using a clustering technique, a two-levels thresholding and three shape parameters - solidity, elongation and roughness - to correctly distinguish fat droplets from other not stained objects like sinusoids. Lastly, we validated our results comparing them with those obtained by a pathologist via stereological point counting. We found a high agreement in the results, with a detection power of 91.01% and a false positive ratio of 4.49%.

1 Introduction

Steatosis is characterized by the abnormal accumulation of lipid droplets within the cytoplasm of hepatocytes. Extensive fat accumulation occurs in a large number of hepatic disorders: alcoholic steatohepatitis (AFL), related to alcohol abuse, and non-alcoholic fatty liver disease (NAFLD), associated to obesity and to metabolic disorders of insulin resistance [1,2]. Other clinical conditions characterized by steatosis are viral hepatitis, nutritional disorders other than obesity, hepatic ischemia and metabolic or endocrine disorders. Moreover, steatosis is the most prevalent condition underlying in liver grafts available for liver transplantation; in this setting, the assignment of moderately steatotic grafts remains controversial [3], particularly when associated with additional risk factors, such as prolonged ischemia or advanced donor age.

Considering that accurate quantification of hepatic fat is not provided by imaging studies, microscopic examination remains the gold standard. Up to now different methods for the estimation of liver steatosis have been proposed: semiquantitative evaluation, manual stereological techniques like point counting, image thresholding and automated morphometry. Semiquantitative analysis appears to overemphasize the amount of liver steatosis and it is influenced by large inter- and intra-observer variations [4]. Point counting is simple, very reliable, and has a good reproducibility [5]; however it is time consuming and labour intensive. Image analysis thresholding requires the manual exclusion of all those structures that – together with fat droplets – are not stained by Hematoxylin-Eosin (HE), or the use of special stains (i.e. osmium tetroxide), not used in routine biopsy. An automated morphometric method seems more suitable. Recently some first approaches using morphological operators such as erosion and dilation, and shape features like circularity and eccentricity, have been proposed [6,7,8]; however deeper analysis and validation are still needed. In this study we developed an automated approach for hepatic steatosis assessment in routine liver biopsies stained with HE from patients with hepatitis C. We also evaluated the accuracy of our method via stereological point counting.

2 Materials and methods

Four routine liver biopsy specimens from patients with chronic hepatitis C from the archives of the Department of Pathology of Niguarda Hospital – Milan were evaluated. All specimens were paraffin embedded, and 4 μm thick sections were stained with HE. A light microscope (Axioplan-Zeiss), equipped with a digital camera, an auto-focusing software and a motorized stage, was used to image capturing and meander scanning. In order to examine the microscopic fields, a 20x objective was used to obtain, together with the camera magnification, images having a 0.264 μm resolution. Each slice was acquired as a grid of tiles, by using a 10% overlap between contiguous tiles, in order to correctly stitch them and to completely include each fat droplet in at least one tile. A slice is a collection of about 100 tiles of 2584 x 1572 pixels in size.

2.1 Overview of the method

Lipid droplets can be considered in digital images as blobs of not stained (white) pixels having a quite circular shape. However, not all the pixels in vesicles are completely white, but they can change from white to pink (the background

* Department of Human Morphology and Biomedical Sciences, University of Milan, Via Mangiagalli 31 20133 Milan, Italy , sciarab@libero.it

colour) or to a grey-blue colour (as a consequence of the refraction of the light on the fat tissue inside vesicles). Moreover, the possible different brightness during the acquisition procedure should be taken into account. Lastly, there are other structures that show the same chromatic characteristics of lipid vesicles and that must be discarded (i.e. sinusoids, portal veins and centrilobular veins). In order to consider all these factors that are involved during image analysis, we performed the following multi-step procedure: firstly pixels of not stained objects are identified by a clustering procedure followed by a two-level thresholding, then blobs in the two binary masks obtained according to the different thresholds are analyzed using shape-related features, in order to recognize fat droplets.

2.2 Colour clustering

To minimize the impact of different brightness conditions, in order to detect which pixels must be marked as "white", we used a clustering technique that automatically fixes the threshold according to the characteristics of the image. We applied the algorithm described by Uchimaya and Arbib [9] for colour image segmentation using a competitive learning clustering. According to the chromatic characteristics of the samples, the pixels are grouped around three reasonably well-defined colours, respectively corresponding to white (the portion surrounding the tissue in the slice, the lipid vesicles, the portal spaces and the lumen of sinusoids), to light red (cytoplasm of hepatocytes) and to dark violet (nuclei of hepatocytes). The clustering algorithm identifies three values (C_1, C_2, C_3) that correspond to the barycentres of the three clusters in the RGB space (i.e. $C_1=[R_1,G_1,B_1]$, and so on). Each pixel of the image is associated to the cluster to which it is closest in l^2 metric. White pixels are the ones associated to C_1. Distances between C_1, C_2 and C_3 have been analyzed to correctly detect tiles with no or few material. In the first case all clusters are related to white pixels, therefore distances are close to zero. In this case the tile was discarded. In the second case it may happen that two clusters (C_1 and C_2) are related to white pixels and just the remaining one to the tissue (both pink background and darker nuclei). In this case a new clustering using only two classes was performed.

2.3 Two-levels thresholding and morphological processing

The clustering procedure gave us a binary mask, M, that marks all the "white" pixels, i.e. those associated with the first cluster. The starting RGB image was then modified as follows: all pixels where M was equal 1 and having colour with a saturation close to zero or a hue turned on blue more than on pink/red was saturated to white (these are some of the pixels inside the large fat droplets, cf. Fig 1b).

Then the resulting image was converted from RGB to greyscale and we fixed two grey level thresholds. The first one, τ_1, was computed as the mean value of the greyscale conversions of C_1 and C_2. The second one, τ_2, was experimentally set as τ_1+20. These thresholds were used to compute two new binary masks, M_1 and M_2, containing only those pixels of the greyscale image that were lighter than τ_1 and τ_2, and where M was equal to 1. M_1 is similar to M, since τ_1 is the "equivalent threshold", in the greyscale space, of the starting clustering. On the other side, M_2 is included in M_1 (i.e. $M_2(x,y)=1$ implies $M_1(x,y)=1$), since it contains only pixels where the grey level is higher than a higher threshold (i.e. lighter pixels, if considering 255 as white and 0 as black).

M_1 and M_2 were then modified by applying the following morphological operators [10]: 1) Opening (erosion followed by dilation), using a circular filter having a 2 pixels (about 0,5 μm) radius. 2) Deletion of small objects (area < 50 pixels, equal to about 3 μm^2). 3) Closure (dilation followed by erosion), using a circular filter having a 2 pixels radius (about 0,5 μm). 4) Filling of the detected objects.

We used these two masks in order to disjoin droplets of fat from other objects, as described in the following sections.

2.4 Evaluating objects' shape

We assigned a global shape index to each object contained in M_1 and M_2: the index "2" is related to objects with a good shape, the index "1" to objects with fairly good shape and the index "0" to objects without a suitable shape.

This global shape index was evaluated using three parameters:

1. Solidity: the area of the object divided by the one of its convex hull.

2. Elongation: the ratio between the minor and the major axis of an ellipse having the same normalized second central moments of the object; it is a measure of the length-width relationship.

3. Roughness: the ratio between the perimeter of the object and the perimeter of its convex hull (a shape measure that quantifies the jaggedness of an object's edges).

We empirically fixed, analyzing a set of twenty images, two sets of thresholds, and we assigned: i) global shape index "2" to objects with solidity>0.85, elongation>0.6 and roughness <1.2; ii) global shape index "1" to other objects with solidity>0.65, elongation>0.45 and roughness<1.4, and iii) global shape index "0" to the remaining objects.

2.5 Lipid droplets recognition

Since M_2 is strictly included in M_1, the same relation holds for each object in M_2: it is included in an object of M_1. By this way, for each object in M_2 there is exactly one object in M_1 containing it. We will call this object "parent". In turn each object in M_1 may have 0, 1 or more objects of M_2 contained in it. We will call these objects "sons".

We now performed, in this order, the following operations:

1. From M_1 all objects that do not have any son were deleted. This rule was used to delete blobs in M_2 that did not contain a white core, so to avoid computing as fat droplets low stained background portions.

2. Each one of the remaining objects, in M_1, having global shape index "2" and an area larger than 150 pixels (around 10 μm^2), was replaced by its convex hull, and it was stored as fat vesicle. These are the most regularly shaped fat droplets (Fig. 1, droplets depicted in cyan).

3. All objects in M_2 with a parent having shape index equal to "0" were deleted, and their parent was stored as sinusoid. This rule avoided classifying regular and whiter portions of sinusoids as small fat droplets (Fig. 2a).

4. All objects in M_2 having a parent already stored as fat vesicle or as sinusoid were deleted, and all objects having global shape index equal to "0" were stored as sinusoids (Fig. 1, sinusoids depicted in red). This rule avoided double detections from the two masks.

5. The remaining objects in M_2 – i.e. those having shape bigger then 0 and not included in an already detected parent - were also stored as fat vesicles (Fig. 1, droplets depicted in blue).

6. Lastly, large epathic arteries, portal veins and centrilobular veins, and tiles regions with no tissue were removed with an upper area threshold, according to their dimension, since they are larger than the largest fat vesicle (Fig. 2c). This threshold was set as 50,000 pixels (equal to about 3,500 μm^2, or to a circular blob with radius 33 µm, where the largest droplets have radius close to 20 µm).

2.6 Stitching different tiles

In order to analyze the whole slice, the last step consisted in merging the results obtained by the analysis of each single tile. In each tile we discarded all detected vesicles touching one of the tile boundaries, in order to avoid partial detections. For vesicles included in the overlapping region, we observed if they were twice recognized in two adjacent tiles. If the answer was yes we retained only the vesicles recognized in the right or lower tile.

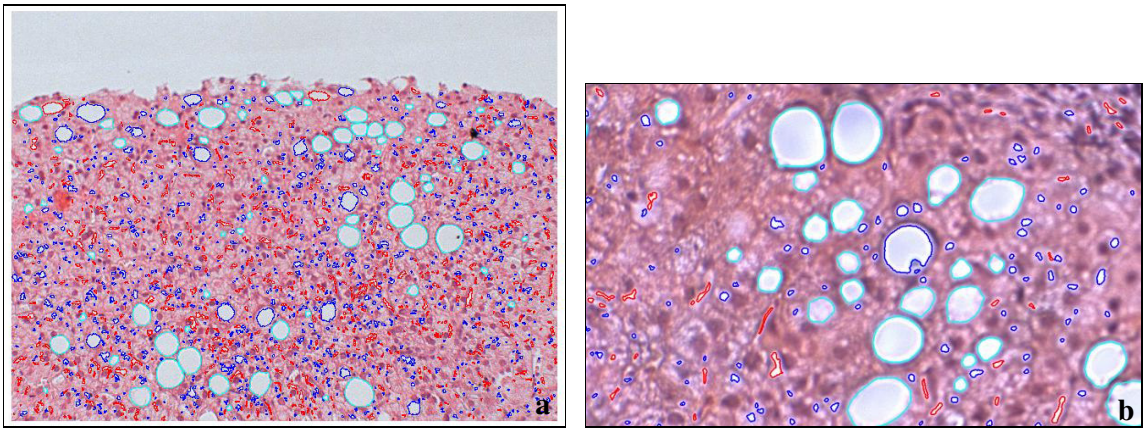

Figure 1. In cyan large and regular shaped droplets (detected by rule 2), in blue other droplets (detected by rule 5), in red sinusoids (detected by rule 3 and 4). (a) Objects identified in a whole tile. (b) A particular from another tile. In the two upper large fat droplets it is visible how pixels inside may be grey-blue instead of white. It is also visible how the detection of the smallest objects can be somehow inaccurate, as we are close to the resolution limit.

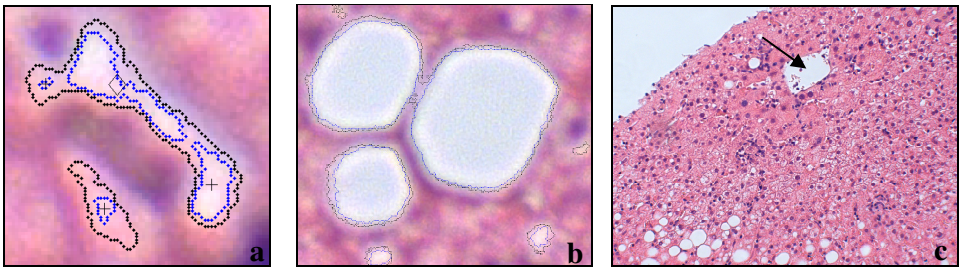

Figure 2. Inner blue lines are the boundaries of the objects in M_2, outer black lines of the objects in M_1 (a) Two sinusoids. In this case, if using only the M_2 mask, some false droplets, marked with a "+", would have been detected. They have been discarded because they was recognized, thanks to rule 3, as parts of a sinusoid. The object marked with a "◊" would anyway have been discarded as it has irregular shape (b) It is visible how using M_1 the two bigger droplets are not disjoined, while M_2 gives the correct detection. It was achieved thanks to rule 5, as the parent object in M_1 had not got a shape good enough to be detected by rule 1 nor an enough bad shape to activate rule 3 (c) A tile with a centrilobular vein, marked with a black arrow.

3 Validation

We randomly selected 15 tile images from three different slices for validation, excluding the 20 images already used for thresholds assessing and all the images with no or very few tissue inside. A counting grid with approximately 600 points, that corresponds to a distance of about 22 μm between adjacent points, was superimposed on each tile. For each point of the grid the classification achieved by our algorithm was computed, using three possible outputs: 1) steatosis, if the point falls inside a lipid droplet, 2) background, if it falls in background or in a sinusoid, and 3) external, if it falls in regions with no tissue or inside portal spaces (detected white objects whose area is bigger than the upper area threshold). A pathologist manually did the same, using the same grid of points. Since our algorithm does not recognize lipid droplets touching the tile boundaries (they are recognized when analyzing the adjacent tile) the pathologist was asked to do the same thing. In all the tiles the steatosis % evaluated by our algorithm was very close to the one obtained applying manual point counting. We also achieved a good detection power (that is the number of the steatosis points correctly detected by the algorithm over the number of the steatosis points detected by the human expert) of 91.01%, with a low false positive ratio (the number of the points wrongly detected as steatosis by the algorithm over the number of the points detected as steatosis by the human expert) of 4.49%. Results are reported in Table 1 and Table 2.

Tile N.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	**Mean**
Auto	5,36	8,11	4,52	7,33	2,37	5,16	1,29	9,68	3,80	11,94	4,48	30,79	16,83	3,81	9,52	**8,33**
Human	5,89	8,61	5,27	7,51	3,16	5,16	0,72	9,88	3,80	13,26	5,37	32,38	17,94	3,33	9,37	**8,78**

Table 1 Steatosis % for each tile, via point counting according to our algorithm and to human expert.

Human / Auto	Ext	Back	Steat
Ext	1871	0	0
Back	0	7018	56
Steat	1	27	567

Table 2 Confusion matrix. All points of the counting grid of each tile are classified according to how they have been marked by the human expert and by the algorithm. From this matrix we achieve a detection power of 91.01% (567/(567+56)) and a false positive ratio of 4.49% ((1+27)/(567+56))

4 Discussion

4.1 Clustering, morphological processing and two-levels analysis

We used colour clustering because, by this way, we can correctly classify the grey-blue pixels inside large lipid droplets as white pixels instead of as background pixels, since in the RGB space they are more distant from the

background colour (pink) than if they were converted to greyscale. In fact, once they are correctly classified, we have to preliminarily saturate them to white, if we want to discard colour information without misclassifying them. The two level analysis is needed because anyway lipid droplets pixels and background pixels may have mixed intensity; as a consequence, independently from the threshold level we choose, there are pixels that can't safely be classified in one way or in the other one. These mistakes can lead us to errors in white blobs classification: for example, if we choose a too high (that means: too white) threshold, a sinusoid can be broken and its fragments can be classified as lipid droplets (Fig. 2a), or some lipid droplets can be detected only partially. On the other side, if a lower threshold is selected, really close steatosis vesicles can merge (Fig. 2b). As described in section 2.5, with such a two-level analysis we were able to analyze together information coming from the grey intensity and information coming from the shape of the objects we were recognizing with the two different thresholds. The 20 grey levels difference between thresholds is empirical, and could surely be improved with further analysis.

Morphological processing is another useful task. Opening breaks weakly connected objects. In particular, when combined with the second operation (area thresholding), it is useful to delete smaller sinusoids. In fact, since they are thin and long, they are firstly broken by opening, and then deleted by area thresholding. Area thresholding also deletes circular objects having diameter less than 2 micron. Micro steatosis smaller than this threshold is anyway invisible. If we wanted to detect it we should use thinner slices and higher magnification. With our slices and magnification factor the objects that are smaller than 50 pixels are more likely light background pixels than vesicles. Lastly closure and filling are useful to correctly mark all pixels in vesicles, completing them.

4.2 Shape descriptors

Elongation, like eccentricity, is a good shape parameter in order to discard objects having a too high length/width ratio. We used elongation, instead of eccentricity (as used in [8]), since it is more directly understandable when looking at the images. Anyway there is a fixed relation between these shape descriptors ($ecc = \sqrt{(1-elong^2)}$), therefore they are equivalent. Both of them are able to correctly detect shapes only when used with other shape descriptors like solidity. In fact they are well suited to classify convex objects, but U-shaped or L-shaped sinusoids may also have small elongation (or eccentricity): they are discarded by the solidity threshold, as they are far from convex. Roughness is used to discard too jagged objects, since they are more likely small sinusoids or few stained background. Anyway, to avoid false deletions, we used a high threshold. Perimeter related shape descriptors are in fact less stable than area related descriptors. For this reason we also avoided to use another classical shape descriptor, circularity (4π*Area/perimeter^2), as reported in [7]. Moreover we tested over the twenty tile images used for thresholds tuning that circularity would have added few or nothing to the combination of elongation and solidity.

5 Conclusions and further developments

In this work we validated the accuracy of the algorithm in steatosis droplet detection. Next steps will be using it on a larger set of biopsies (actually just four biopsies were used), thresholds fine tuning and an analysis of the size distribution of fat droplets, analyzing how it varies in relation with steatosis %. Another related problem we will try to solve is automatic detection of fibrosis regions.

References

1. C.S. Lieber "Alcoholic fatty liver: its pathogenesis and mechanism of progression to inflammation and fibrosis", *Alcohol*, 34(1):9-19, 2004.
2. R. Vuppalanchi, N. Chalasani "Nonalcoholic fatty liver disease and nonalcoholic steatohepatitis: Selected practical issues in their evaluation and management", *Hepatology*, 49(1):306-317, 2009
3. A. Nocito, A.M. El Badry, P.A. Clavien "When is steatosis too much for transplantation?", *J Hepatol*, 45:493-499, 2006
4. L.E. Franzen, M. Ekstedt, S. Hechagias et al. "Semiquantitative evaluation overestimates the degree of stetosis in liver biopsies: a comparison to stereological point counting", *Mod Pathol*, 18:912-916, 2005
5. E. R. Weibel. *Stereological Methods - vol. 1*. Point counting methods pp.101-161 - Acad. Press Inc., London, 1979
6. H. Marsman, T. Matsushita, R. Dierkhising et al. "Assessment of donor liver steatosis: pathologist or automated software?", *Human Pathology*, 35(4): 430-435, 2004
7. V. Roullier, C. Cavaro-Menard, C. Guillaume at al., "Fuzzy algorithms to extract vacuoles of steatosis on liver histological color images" in *Proceedings of the 29th Annual International Conference of the IEEE EMBS*, pp. 5575-5578, Lyon, 2007
8. G.E. Liquori, G. Calamita, D. Cascella et al., "An innovative methodology for the automated morphometric and quantitative estimation of liver steatosi", *Histol Histopatol*, 24:49-60, 2009
9. T. Uchimaya and M.A. Arbib "Color image segmentation using competitive learning", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(12):1197–1206, 1994
10. R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, third edition, Pearson Education, 2008