High Resolution Digital Tissue Image Processing Using Texture Image Databases

Gábor Kiss[1]([✉]), Orsolya Eszter Cseri[1], Ádám Altsach[1],
István Imre Bándi[1], Levente Kovács[1], and Miklos Kozlovszky[1,2]

[1] Óbuda University/Biotech Knowledge Center, Budapest, Hungary
{kiss.gabor,cseri.eszter,altsach.adam,bandi.istvan,
kovacs.levente,kozlovszky.miklos}@nik.uni-obuda.hu
[2] MTA SZTAKI/Laboratory of Parallel and Distributed Computing, Budapest, Hungary

Abstract. Texture based image databases integrated with effective searching algorithms are useful solutions for many scientific and industrial purposes. Medical image processing of high resolution tissue images is one of the areas, where the cell/tissue classification can rely on such solutions. In this paper we are describing the design, development and usage of a specialized medical texture image database. Our primary aim with this texture database is to provide Digital Imaging and Communication in Medicine (DICOM) compatible texture image dataset for cell, gland and epithelium classification in histology. Our solution includes a Picture Archiving and Communication System (PACS) subsystem, which is mainly provide a communication interface (texture image searching and retrieval) and enables image processing algorithms to work more effectively on high resolution tissue slide images. In this paper we describe how our Local Binary Pattern (LBP) based algorithm benefits texture database usage when solving image processing problems in histology and histopathology.

Keywords: PACS · DICOM · LBP · Texture based image database · Medical image processing · Digital microscopy

1   Introduction

With the development of high resolution scanning and informatics, the practitioner's workflow has changed, and high resolution digital microscopy becomes possible. Slide sharing, in digital world is just easy as forwarding images via Internet, also visualization and archiving the images become simple tasks. Unfortunately, the migration into digital world did not solve all problems, and slide assessment tasks still remain complex and time consuming in the workflow. Practitioners - independently of whether they use traditional methods or new digital slides-, are sharing the same level of difficulties to do cancer grading and determine efficient treatments. In many countries (due to the large amount of samples) patients are waiting in long queues for their diagnosis result. Such waiting queues need to be avoided, because during waiting time, health problems may get worse.

2   State of the Art

We have done a market screening of the available high resolution digital microscopy image assessment software solutions. Our evaluated systems were the followings: the TissuemorphDP [1] image analyzer from Visiopharm, which is designed mainly for the population of nuclei examination, HER2-CONNECT [2] targeting tumor cells detection, Virtuoso, DeveloperXD [3] and the MediaCybernetics [4] doing image of tissue analysis.

We have examined also some image databases namely: ASH Imagebank [13], Image After [14], TinEye [15], RevIMG [16]. The ASH Imagebank is a hematology database with key word search. ImageAfter is a texture based approach with the string-based search method. Finally TinEye and RevIMG are Content Based Image Retrieval systems (CBIR). All the examined systems support only full size images instead of plain texture images.

MorphCheck is our medical image analysis platform working on high resolution digital microscopy images. It enables the load of high resolution tissue images, automatically detects the basic structures of tissue (nuclei, glands, surface epithelium) and provides quantitative and qualitative parameter measurements Recently we have added texture based image analysis to improve its object recognition accuracy. The users can do texture-base search and object recognition without manual selection.

Most of the image acquisition modalities are using the DICOM standard combined with PACS. We also have enabled the support of a DICOM/PACS based medical database system [5] in our system. DICOM and PACS technologies guarantee the standard storage for medical images and the software makes the additional extension of its abilities later. PACS server is being used by virtual servers in a cloud. We are using IaaS (Infrastructure as a Service), because we can scale to any amount of computer resources dynamically based on usage and other parameters. The scaling process is supported by the EC2 compatible cloud middleware.

The system measures a large number of morphological parameters on the detected image objects and store them in the database, all higher level analysis and decision is based on the archived database parameters. Image processing and classification tasks demand high computing capacity, thus the algorithms are able to run on various hardware architectures and support GPUs to enable higher performance [6], [7]. Our aim is to increase the efficiency of the software with new texture-based approaching.

3   Architecture

MorphCheck consists of two well separable parts. One of these parts is the plain algorithmic methods working on processing and detection. The other part is using the texture-based approach doing the image processing tasks (shown in Fig. 1.) The user may define workflows and freely use a combination of both parts of the image processing methods. A workflow contains tasks (e.g.: texture based and classifier algorithms) and can be arbitrary complex. The system also supports manual image annotation, where the pathologist makes the annotation of the tissue sample and uploads it into the database. The annotated images will be stored in the PACS system, and algorithms can reuse later on their data during classification processes.
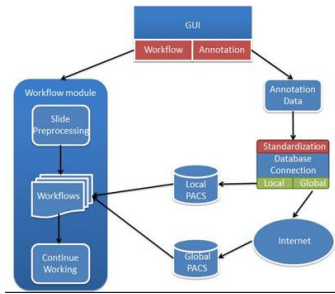
Fig. 1. High-level system architecture of MorphCheck

4   Texture Based Search Algorithm

Our implemented LBP algorithm is based on the Local Binary Pattern [9] (shown in Fig. 2. and Eq.1). The value of LBP code of a pixel is given by:

$$s(x) = \begin{cases} 1, & if \ x \geq 0; \\ 0, & otherwise \end{cases} \tag{1}$$

Fig. 2. part A shows a sample image with intensity pixels and the center pixel is being marked by blue color. Fig. 2. part B shows the differences between each pixels and the marked center pixel, and the differences are being written to the place of the original pixel. For example: The center pixel is 70 and the top left pixel is 47. The difference is -23, and it is written to the top left pixel place at the part B of the Figure 2. The last calculation is being based by Eq.1. Take each pixel on B image and if the current pixel greater than or equal 0, at C image the same pixel will be marked as 1. Otherwise it will be written 0. Finally the LBP code can be red by the clockwise direction.
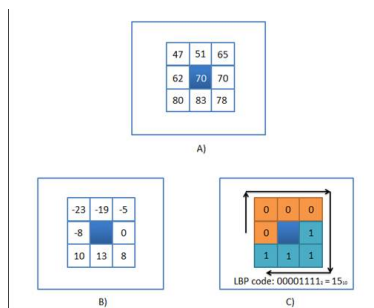

Fig. 2. Local Binary Pattern. A) Sample intensity pixels with center pixel, B) Differences from center pixel, C) LBP code in binary number and decimal system.

The task is to separate and classify the natural textures on the images. The input of the algorithm is the area where the examination happens, and those classes should be later identified on images (more pictures may belong to one class).

At first the Simple Linear Iterative Clustering (SLIC) algorithm slices images into pieces of texture [10], [11]. In the SLIC algorithm it is necessary to set two parameters: the size of the superpixel and how uniform it should be. The second parameter means how homogenous the colors should be in the superpixel. At high parameters we get inhomogeneous, compact and not random size superpixels (shown in Fig. 3.).
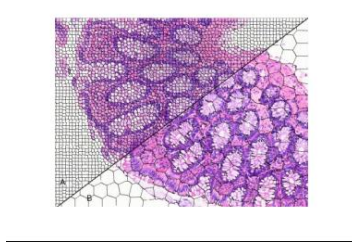

Fig. 3. A high resolution tissue image split into superpixels

After these steps we examine all superpixel in terms of belong in to a class. This process happens to the help of two parameters: Simple histogram and rotation invariant Local Binary Pattern histogram. With both parameters we calculate distance measure for all images. For the case of histogram we calculate the difference with the Earth Mover's Distance [4] metric. In the case of the LBP histogram we calculate the absolute difference of both histograms. Based on LBP to all superpixel we calculate a classification, after that we compare all superpixels classification with other. The robust classified superpixels are which superpixels classified same class, on the basis of both histograms (LBP and simple), all others are indeterminate superpixels. Currently, to avoid the false negative hits we classify them to the unclassified class, but later with the stable neighbors we will assign ranks to them.

5   Measurement and Test Environment

To measure image object classification efficiency, we have adapted some metrics from the literature. It is necessary to introduce some concepts to this:

- Reference result set: The pixels, which are marked in the reference image by a skilled expert.
- Test result set: The pixels, which are marked by the algorithms.
  The selected pixel-based parameters from the literature are [12]:
- False negative (FN): Reference result set contains the current pixel, but the test result set does not.
- False positive (FP): The test result set contains the currently pixel, but the expert is not marked it on the reference result set.

- True positive (TP): The pixel is contained in both reference result set and test result set.
- True Negative (TN): Neither set contains the current pixel.
- Accuracy (TP+TN)/(TP+TN+FP+FN): This is the measure between 0% and 100%, where 100% means the reference result set and the test result set are equal. Thus the doctor and algorithms marked different pixels on the images.
- Recall (TP/(TP+FN)): This is a measure number indicates how much it found from the reference result set. If found all of them 1, else converge to zero.
- Precision (TP/(TP+FP)): Measurement of the hit. If the algorithm marked all pixels marked, the precision value is 1, else converge to zero.

6   Performance and Accuracy

In this section we show the performance of our LBP based algorithm, according to next aspects: runtime, accuracy, recall, precision. We use anonym tissue samples to do the efficiency measurements. To the tests we use 20 digital samples, which consist of 50 ROI (Region of Interest). From among these 32 ROI were healthy, 18 unhealthy. Table 1 contains all the data which has been produced by the tests.
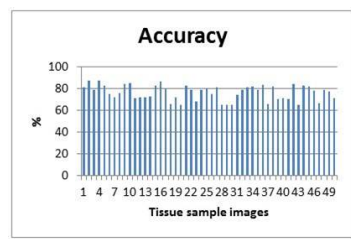

Fig. 4. Result of accuracy examination

We run the examined algorithms under the tests, then we have calculated on all images the defined metrics (TP,TN,FP,FN). The accuracy is being shown by the Fig. 4. and Table 1. This metric of our implemented algorithm is in the range of 64.3- 87,3% and the average is 75,99%.

The other important derived result is the so called "recall" (shown in Fig. 5.). Mixed results arise in the course of the tests. The LBP based algorithm tries to avoid the false positive hits. The area will be marked as undetermined if there is small confidence. This is the main reason of low recalls in the test result.

The third examined parameter was the precision (shown in Figure 6.). Our implemented algorithm is able to find pixels between 0,6 and 0,8 precision on all reference images. Among the examined 50 test, the founded pixels in 24 cases were correct, which means over 80% precision (which is very good if using such samples).
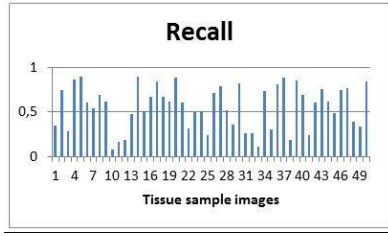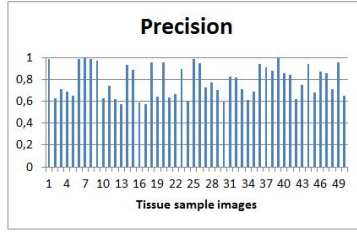
Fig. 5. Result of recall examination


Fig. 6. Result of precision examination

Table 1. Runtime, accuracy, recall and precision values of each image

Image	Runtime	Accuracy	Recall	Precision (0-1)
1.	48032 ms	80,5561 %	34,989983 %	0,987352781
2.	50503 ms	87,3210 %	74,7283407 %	0,627021984
3.	45815 ms	78,4133 %	27,5994314 %	0,707845967
4.	53523 ms	86,8289 %	86,2483716 %	0,69109668
5.	48907 ms	82,3787 %	89,4566917 %	0,652158323
6.	45962 ms	74,5288 %	60,1550501 %	0,98606324
7.	49185 ms	71,5572 %	54,2035412 %	0,997744071
8.	51919 ms	75,7369 %	68,6987809 %	0,984041932
9.	47646 ms	83,8932 %	61,719451 %	0,968283442
10.	46471 ms	84,8415 %	72,857387 %	0,626680672
11.	46991 ms	70,9533 %	16,0547004 %	0,739163416
12.	50978 ms	71,2926 %	18,0046956 %	0,621332388
13.	56035 ms	71,4357 %	47,9457611 %	0,571567614
14.	53736 ms	72,6450 %	89,063646 %	0,934838713
15.	45531 ms	82,1188 %	50,8651431 %	0,883274726
16.	51051 ms	85,9209 %	67,1163571 %	0,591183075
17.	45935 ms	79,3216 %	83,922605 %	0,574389702
18.	57948 ms	65,2986 %	66,5004848 %	0,956795831
19.	50487 ms	71,7855 %	61,324101 %	0,646012997
20.	46289 ms	64,3528 %	87,9682645 %	0,957433946
21.	55556 ms	82,6466 %	59,9626072 %	0,632806935
22.	42149 ms	78,7552 %	31,8393793 %	0,666676572
23.	42207 ms	68,0102 %	49,1480765 %	0,896028264
24.	53199 ms	78,2759 %	50,4916674 %	0,594485946
25.	48327 ms	79,9055 %	23,9523506 %	0,986024626

Table 1. (*Continued*)

26.	54416 ms	74,5399 %	70,6376023 %	0,948759786
27.	53046 ms	80,5891 %	78,4338479 %	0,722898764
28.	46335 ms	64,9316 %	52,0194195 %	0,775494037
29.	58761 ms	64,8349 %	36,1403908 %	0,705592692
30.	52995 ms	64,8009 %	82,1711576 %	0,593973493
31.	44834 ms	74,0871 %	26,2567997 %	0,824267399
32.	42018 ms	78,9241 %	25,836577 %	0,819038487
33.	49489 ms	81,2256 %	10,4580788 %	0,710334421
34.	48039 ms	81,3128 %	72,8684392 %	0,609290772
35.	48893 ms	78,9584 %	30,04997 %	0,689207987
36.	51592 ms	83,4439 %	80,8248287 %	0,944038789
37.	43584 ms	65,4951 %	87,9658039 %	0,913664024
38.	53240 ms	81,6129 %	18,0776139 %	0,876442238
39.	49503 ms	70,2107 %	84,9993829 %	0,992048124
40.	52690 ms	71,1560 %	68,9798564 %	0,854984643
41.	43347 ms	70,2632 %	23,2500211 %	0,843804389
42.	56431 ms	84,0216 %	60,8157887 %	0,623285695
43.	54683 ms	64,8362 %	75,8693771 %	0,749807238
44.	56764 ms	82,4279 %	613156813 %	0,941009703
45.	58503 ms	81,9611 %	48,3010939 %	0,684079944
46.	44213 ms	77,9225 %	74,623875 %	0,874604003
47.	51478 ms	66,5134 %	76,3992459 %	0,856582055
48.	47767 ms	78,8671 %	39,228208 %	0,709422414
49.	46275 ms	76,8262 %	33,2365511 %	0,952316072
50.	43277 ms	71,0800 %	84,0721253 %	0,651843577

Fig. 7 shows some test images. During visualization we are using some false colors by our algorithm: Yellow is a correct hit, black is undetermined area, blue is a background and green is a not examined tissue type. On the sample images there are some tissue structure: purple circle are glands, pink area is the epithelium. In this test we are searching for glands. In the B and D images the avoidance of the false positive hits showed by black color. Finally there are lots of good hits in a gland, epithelium and the background, among the black undetermined areas.
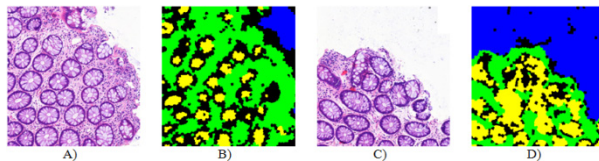
Fig. 7. A) Test image 1, B) Results of test image 1, C) Test image 2, D) Results of test image 2

The key of the algorithm's accuracy is the chosed distance measure technique. In the future we can probably enhance the accuracy by changing the current histogram based approach to the GLCM (Gray-Level Co-Occurance Matrix) based approach.

7   Summary

We have successfully implemented into the MorphCheck system an LBP-based texture classifier algorithm, which significantly provides more efficient object classification and segmentation then the previous solutions. After the integration we have done efficiency examination of our algorithm. We have used pre-defined metrics for the algorithm assessment. For the texture based algorithm we have combined MorpCheck with a DICOM/PACS based system and annotated a large number of tissue images. We used 1039 texture in PACS database and the LPB algorithm was working on these annotated textures. The new texture based algorithm and the combined PACS has been validated and facilitates diagnostic works significantly.

References

1. Visiopharm for digital pathology. Visiopharm (March 2013). http://www.visiopharm.com/pdf/visiomorphdp-factsheet.pdf
2. Her2-connect. Visiopharm (March 2013). http://ww1.prweb.com/prfiles/2010/11/18/4253634/Imageforpressrelease.bmp
3. Developerxd. Definiers (March 2013). http://www.definiens.com/tissue-diagnostics.html
4. Bickel, E.L.P.: The earth mover's distance is the mallows distance: Some insights from statistics. In: ICCV 2001, IEEE 12th International Conference on Computer Vision, pp. 251–256 (2001)
5. Digital imaging and communications in medicine, National Electrical Manufacturers Association (2004). http://dicom.nema.org
6. Reményi, A., Szénási, S., Bándi, I., Vámossy, Z., Valcz, G., Bogdanov, P., Kozlovszky, M., Sergyán, S.: Parallel biomedical image processing with GPGPUs in cancer research. In: 3rd IEEE International Symposium on Logistics and Industrial Informatics, Budapest, Hungary, August 2011, pp. 225–248 (2011)
7. Kozlovszky, M., Szénási, S., Vámossy, Z.: GPGPU-based data parallel region growing algorithm for cell nuclei detection. In: CINTI 2011, 12th IEEE International Symposium on Computational Intelligence and Informatics, pp. 493–499 (2011)
8. Fugal, D.L.: Conceptual wavelets in digital signal processing. Space & Signals Technical (2009)

9. Xiaoyu Wang, S.Y., Han, T.X.: An hog-lbp human detector with partial occlusion handling. In: ICCV 2009, IEEE 12th International Conference on Computer Vision, pp. 1550–5499 (2009)
10. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slicsuperpixels, Institute of Electrical and Electronics Engineers, Tech. Rep. (2010)
11. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**, 2274–2282 (2012)
12. Yasnoff, J.W.A., Mui, J.K.: Error measures for scene segmentation. Pattern Recognition **9**, 217–231 (1977)
13. ASH ImageBank, hematology imagebank. (Marc 2013). http://imagebank.hematology.org/
14. ImageAfter texture database (March 2013). http://www.imageafter.com/index.php
15. TinEye Reverse Image Search (March 2013). https://www.tineye.com/
16. RevIMG reverse visual search (March 2013). http://www.revimg.net/