



HAL
open science

Knowledge-Guided Semantic Indexing of Breast Cancer Histopathology Images

Adina Eunice Tutac, Daniel Racoceanu, Thomas Putti, Wei Xiong,
Wee-Kheng Leow, Vladimir Cretu

► **To cite this version:**

Adina Eunice Tutac, Daniel Racoceanu, Thomas Putti, Wei Xiong, Wee-Kheng Leow, et al.. Knowledge-Guided Semantic Indexing of Breast Cancer Histopathology Images. BMEI2008, International Conference on BioMedical Engineering and Informatics, May 2008, Sanya, Hainan, China. hal-00342275

HAL Id: hal-00342275

<https://hal.science/hal-00342275v1>

Submitted on 27 Nov 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Knowledge-Guided Semantic Indexing of Breast Cancer Histopathology Images

Adina Eunice Tutac^{1,5}, Daniel Racoceanu^{1,6}, Thomas Putti², Wei Xiong^{1,4},
Wee-Kheng Leow^{1,3} Vladimir Cretu⁵

¹IPAL - UMI CNRS 2955 Singapore

²NUH, Singapore, ³NUS, Singapore, ⁴I2R, A-STAR, Singapore,

⁵Politehnica University of Timisoara, Romania,

⁶University of Besançon, France

newlifeadit@gmail.com, danielr@comp.nus.edu.sg

Abstract

Narrowing the semantic gap represents one of the most outstanding challenges in medical image analysis and indexing. This paper introduces a medical knowledge – guided paradigm for semantic indexing of histopathology images, applied to breast cancer grading (BCG). Our method improves pathologists' current manual procedures consistency by employing a semantic indexing technique, according to a rule-based decision system related to Nottingham BCG system. The challenge is to move from the medical concepts/rules related to the BCG, to the computer vision (CV) concepts and symbolic rules, to design a future generic framework- following Web Ontology Language standards - for an semi- automatic generation of CV rules. The effectiveness of this approach was experimentally validated over six breast cancer cases consisting of 7000 frames with domain knowledge from experts of Singapore National University Hospital, Pathology Department. Our method provides pathologists a robust and consistent tool for BCG and opens interesting perspectives for the semantic retrieval and visual positioning.

1. INTRODUCTION

Within the last decade, histological grading [1] has become widely accepted as a powerful indicator of prognosis in breast cancer. Most grading systems currently employed for breast cancer combine criteria in nuclear pleomorphism, tubule formation and mitotic counts. In general, each grading criteria is evaluated by a score of 1 to 3 (3 being associated to the most serious case) and the score of all three components are added

together to give the "grade". Breast Cancer Grading (BCG) [1] [2] requires time and attention while classifying 100 cases/ day, each of them having around 2000 frames, as the pathologists usually do. Currently, BCG is achieved by visual examinations of pathologists. Such a manual work is time-consuming and inconsistent. According to those issues, developing an automatic grading system represents a strong medical requirement.

Such an automatic grading system should naturally be able to semantically index the images in line with the medical domain knowledge, and inspired from their real content. Content-based image indexing [3], [5] has been subject of significant researches in the context of medical imaging domain [4] [6]. Solving the issue of the semantic gap [7] [8] between the low level features [9], [10] and the high level semantic concepts [11] represents the cutting edge research [12], [13].

In this paper, we propose a solution to meet pathologist needs for automatic BCG. Beyond this, we further model the BCG-related medical knowledge into reasoning rules. These rules are embedded in semantic indexing approaches.

The proposed method provides pathologists a robust and consistent tool, as a second opinion, for breast cancer grading, using the Nottingham grading system [1] [2]. The effectiveness of the proposed approach has been validated in experiments over six breast cancer cases consisting of 7000 frames with domain knowledge from pathologist experts.

The paper is organized as follows. In section 2, domain knowledge analysis is introduced by describing a synthesis of the breast cancer grading standard system and showing the importance of grading in breast cancer detection. Section 3 presents our grading approach model with the medical image indexing inspired rules, with the computation of local and global grading. The

semantic indexing of image features to give the local and the global grading is presented in section 4. Section 5 contains experiments and results leading to understanding semantic breast cancer image analysis, thus, to achieve the grading aim. Finally, the results and approaches are analyzed and conclusion/perspectives are indicated.

2. DOMAIN KNOWLEDGE ANALYSIS – COMPUTATION OF LOCAL AND GLOBAL GRADING

Breast cancer refers to a malignant tumor that has developed from cells within the breast. Breast cancer is a leading cause of death among women, and its incidence is rising. Although curable, especially when detected at early stages, breast cancer is expected to account for 28% of incident cancer and 20% of cancer deaths in women. A powerful marker in breast cancer detection is the breast cancer grading. Among the standard grading systems used all over the world, Nottingham Grading System (NGS) is preferred for the reason of providing more objective criteria for the three component elements of grading and specifically addresses mitosis counting in a more rigorous fashion.

The three component NGS criteria are briefed below (see Table 1, Figure 1):

- Tubule Formation (TF) - are referred as white blobs (lumina) surrounded by a continuous string of cell nuclei. The assessment of tubular/acinar differentiation applies to the neoplasm overall (over the whole tumor) and requires examination of several sections at scanning magnification.
- Mitoses represent dividing cells and the Mitosis Count (MC) score is assessed in the peripheral areas of the neoplasm and not in the sclerotic central zone. Mitoses are abundant in areas of poor tubule formation. Although, the Oncologic Standards Committee considers that mitosis count per square millimeter is most accurate, the NGS uses a scoring system based on the number of mitoses per 10 High Power Field's (HPF's).
- Nuclear Pleomorphism Score (NPS) - categorizes cells nuclei based on two features: size and shape.

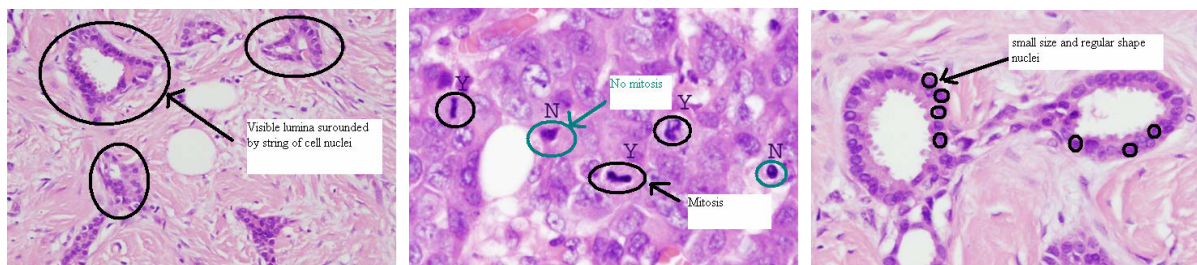


Figure 1. NGS synthesis a) Tubule Formation with more than 75% of neoplasm having a tubular pattern b) Mitosis differentiation - the black arrow indicates mitosis; the green arrow indicates non-dividing cells c) Small size and regular shape nuclei.

The scores for the three separate parameters (tubules, nuclei and mitoses) are summated and the overall grade of the neoplasm is determined. The summation is usually done on 10 frames chosen visually by the pathologist, according to his expertise (see Table 2). In the technical reports [1], only the way of choosing frames for mitosis count is mentioned. In our approach, we propose the use of a simple summation fusion between the 3 component criteria by frame, to automatically choose the top ten hyperfields able to compute the global grading.

Table 1. CRITERIA Score – Breast Cancer Grading

Criteria Score	Hyperfield (Frame) Score		
	TF	NPS	MC
1	>75% neoplasm tubules	Small size and regular shape	< 9
2	10 - 75 % neoplasm tubules	Medium size and variation shape	10-19
3	<10% neoplasm tubules	Big size and irregular shape	> 19

Table 2. COMPOSITE score –Breast Cancer Grading

Composite score /10 frames	Global score (TF+NPS+MC)
----------------------------	---------------------------------

Composite score /10 frames	<i>Global score (TF+NPS+MC)</i>
Grade I (well differentiated)	3- 5
Grade II (moderately differentiated)	6-7
Grade III (poorly differentiated)	8-9

3. BREAST CANCER GRADING RULES-BASED SYSTEM MODELLING

The purpose of this section is to introduce the approach used to step from the medical concepts and rules related to the breast cancer grading, to the computer vision (CV) concepts and symbolic rules.

The aim is to move towards a future generic frame for an assisted semi-automatic generation of CV rules and (in future) computer programs, starting from specific medical queries.

The modeling demarche has been led according to the Ontology Web Language (OWL) developed in the Semantic Web Protégé framework [14].

This section is structured in three parts, according to the main steps of the proposed approach: development of the correspondence between medical concepts and computer vision concepts with respect to the OWL standard; definition of intermediate CV rules and generation of the final Symbolic Rules by combination of the CV concepts and intermediate CV symbolic rules.

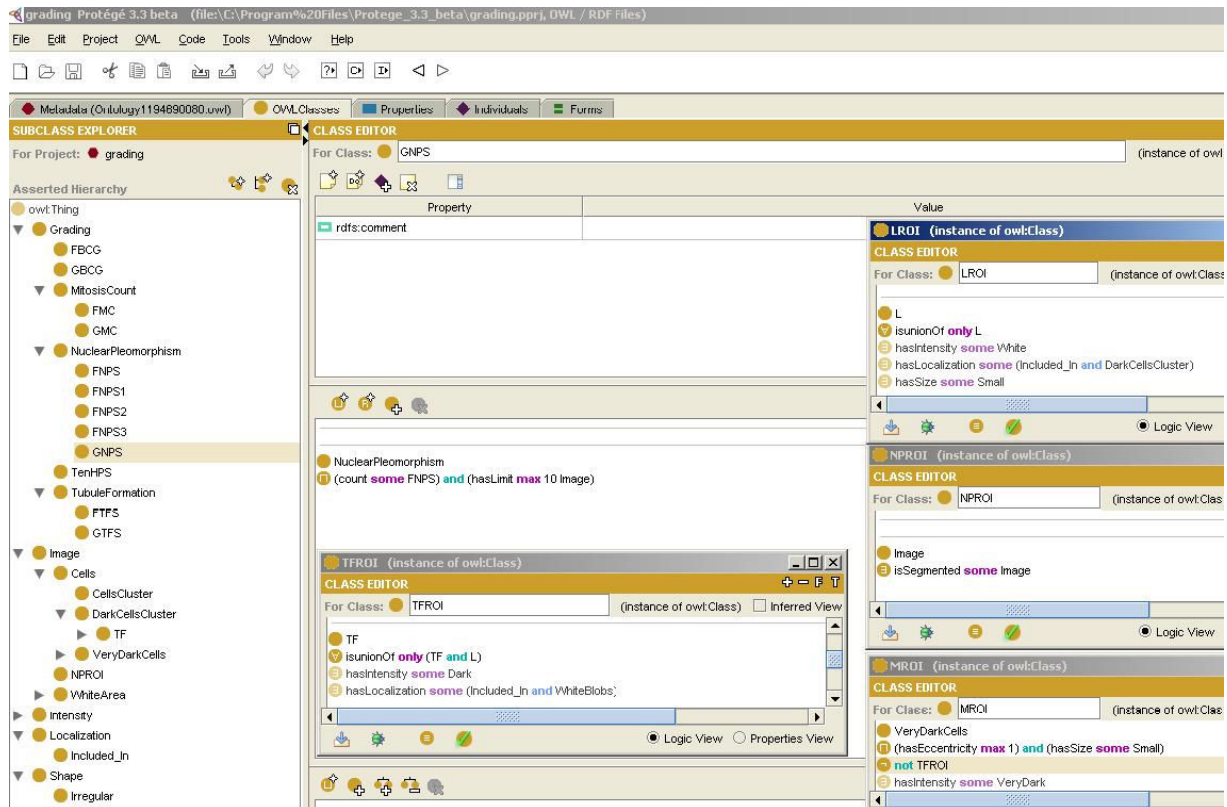


Figure 2. PROTEGE model of the Breast Cancer Grading computer vision symbolic rules and concepts

3.1 Correspondance between the Medical concepts and adequate Computer Vision concepts

According to the NGS synthesis, we proceed to the tubule formation extraction as the nuclear pleomorphism and mitosis count parameters computation in order to create the rule-based method

able to automatically generate the grading (see Table 3). Therefore, to clearly define the rules, medical concepts are transformed into computer vision concepts. The computer vision concepts are then modeled as Protégé concepts, in standard types as classes and properties. Various instances for different classes are created as individuals, where specific

values are assigned for classes and properties, respectively.

Table 3. CONCEPTS correspondence (Medical – CV Protégé)

Medical Concepts	CV concepts	Protégé concepts type
Slide	Image (digitized)	Super class
Grading	Grading	Super class
Cells	Cells	Class inherited from Image
CellsCluster	Union of Cells	Class inherited from Cells
DarkCellsCluster/ VeryDarkCells Cluster	Union of Cells	Class inherited from Cells with <i>hasIntensity</i> (property) Dark/VeryDark (instances of Intensity class)
Lumina	White compact segments of the Image included in union of dark Cells	Class inherited from WhiteBlobs with <i>hasIntensity</i> (property) White (instance of Intensity class) <i>hasSize</i> (property) Small (instance of Size class) , <i>hasLocalization</i> (property) Included_In (instance of Localization class) DarkCellsCluster (instance of Cells)
Tubule Formation	Union of Cells	Class inherited from Grading
Mitosis	Dividing Cells nuclei	Class inherited from Grading
Nuclear Pleomorphism	dimension and shape characteristics of nuclei	Class inherited from Grading
Local Grading FBCG	Computation of grading for Tubule Formation , Mitosis number, Nuclear Pleomorphism for a single image (FTFS, FMC, FNPS, FBCG)	Class inherited from TubuleFormation/ NPS MC
Global Grading GBCG (10HPFs)	Computation of grading for Tubule Formation , Mitosis number, Nuclear Pleomorphism for ten images (GTFS, GMC, GNPS, GBCG)	Class inherited from TubuleFormation NPS MC
Intensity	Color spectrum [White...VeryDark] e.g. White	Super class for White, Dark and VeryDark
Size/Shape	Dimension/Sha	Super classes for Small,

	pe range [small...large]/ [regular...irregular] e.g. small e.g. regular	Medium, Large Classes / Regular, Variation, Irregular Classes
Localization	Position indicator	Supper class for Included_In class and <i>hasLocalization</i> property
	Math operators	count, round, sum, divide property

CV concepts are defined in terms of Protégé language as super classes, classes and properties. The inheritance strategy is applied in the building of the classes' structure. Combining these elements enables modeling the symbolic rules related to each breast cancer grading criterion.

Figure 2 presents a part of the Protégé model designed for Breast Cancer Grading with computer vision concepts implemented as classes, properties and symbolic rules.

3.2 Intermediate rules

To obtain the symbolic tubule formation rule, we create intermediate rules for each domain concept used for this criterion.

- *DarkCellsCluster* is defined as containing group of cells with intensity property value setup between *VeryDark* and *White* limits.

$$DarkCellsCluster = \{Cells \mid VeryDark < i < White\}$$

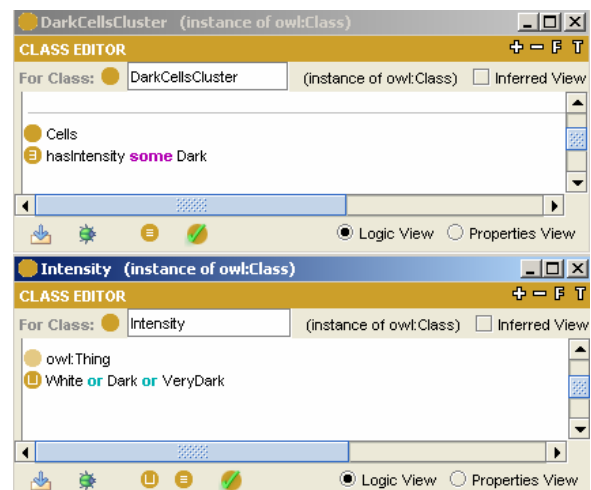


Figure 3. PROTEGE *DarkCellsCluster* symbolic rule

In terms of Protégé (see Figure 3), this rule is defined as: Cells with *hasIntensity* (property) some Dark, which is an instance of Intensity class. *Some* Dark may

take different values when creating individuals for *DarkCellsCluster* class.

WhiteBlobs rule points to the blobs inside the white image area that are white.

- $WhiteBlobs = \{morph(WhiteArea)\}$

In practical image processing/analysis, the detection *DarkCellsCluster*, *VeryDarkCells* or *WhiteBlobs* becomes a simple intensity-based segmentation method.

WhiteBlobs intermediate rule helps us to define the Lumina (L) rule. Practically, lumina defines white blobs included in the existing *DarkCellsCluster*.

$L = \{WhiteBlobs(i,l) \mid \exists DarkCellsCluster(i,l) \supset WhiteBlobs(i,l)\}$

LROI intermediate CV rule is a union of all lumina detected in the image.

- $LROI = \cup L$

Following the same idea, intermediate rules are defined for the mitosis count symbolic rule.

Similar to *DarkCells*, detection of *VeryDarkCells* is obtained by an intensity- based segmentation method with a setup thresholding.

- $VeryDarkCells = \{Cells \mid i \leq VeryDark\}$

- $ecc(VeryDarkCells)$ intermediate rule represents

an eccentricity determination operation computed for the *VeryDarkCells*. We use a deterministic approach for the *ecc* concept but the use of fuzzy logic membership sets obtained from the training database is applicable in the same way, according to the training database consistency.

- $size(VeryDarkCells)$ intermediate rule applies a size detection threshold onto the *VeryDarkCells*.

Nuclear Pleomorphism rule takes into consideration the size and shape of cells nuclei. There was no need for an intermediate rule to be generated.

3.3 Generation of the final symbolic computer vision rules

Considering the tubule formation criteria given by the pathologist:

- Pathologist rule:

“Tubule = white lumina blobs surrounded by string of dark cells nuclei.”

the correspondent symbolic rule is created involving the white visible part (Lumina) surrounded by continuous string of dark cells nuclei:

- Symbolic rule (used in our algorithm):

TF intermediate symbolic rule specifies that if there are *WhiteBlobs* included in the *DarkCellsCluster*, the pathologic criterion is satisfied.

$TF = \{DarkCellsCluster(i,l) \mid$

$\exists WhiteBlobs(i,l) \subset DarkCellsCluster(i,l)\}$

where: *TubuleFormationROI* is defined by:

$TF_{ROI} = \{U(DarkCellsCluster(i,l), L)\}$

The TF_{ROI} symbolic rule creates the union of all *DarkCellsCluster* - clusters dependent of intensity and localization- and *L*, resulting in tubule formation region of interests. The close $\cup DarkCellsCluster$ detection is operated using morphologic operators. The \cup operator from the symbolic rule is translated by object filtering techniques.

The result of this operation is to index the medical image by the *TubuleFormationROI*. This is an important point of our approach, since we are able to associate to each frame precise ROI structure corresponding to the detected tubule formations. Each ROI will have also additional attached information. In this sense, information like tubule formation area can be associated with the grading score and used afterwards for the FTFS computation by dividing it to the neoplasm area. This area is computed as a subtraction of the white parts area from the frame one.

For the mitosis rule we consider all very dark dividing cells that are not contained in the tubule formation areas.

Pathologist Rule: Mitosis = very dark dividing cells nuclei from the peripheral area of neoplasm

- Symbolic Rule:

MitosisROI:

$M_{ROI} = \{\cup(VeryDarkCells) \mid ecc(VeryDarkCells), size(VeryDarkCells), Cells \notin TF_{ROI}\}$

The important aspect here is that *VeryDarkCells* structures must not be contained in the tubule formation area (TF_{ROI}), specified in the rule by the \notin operator.

Thus, M_{ROI} rule is defined as a union of all *VeryDarkCells* dependent of particular *ecc*, *size* and *localization* values.

Following the same idea, we establish nuclear pleomorphism rules consisting of a detection of size and shape vectors of nuclei, contained in the frame.

- Pathologist Rule : Size and Shape features of nuclei
- Symbolic Rule :

NucleiROI:

$NP_{ROI} = \{segment(Im)\}$ where

$segment(Im) = \{size(DarkCellsCluster), shape(DarkCellsCluster)\}$

NPOI symbolic rule is defined as a segmentation image processing operation applied to the input image

to obtain the nuclei regions of interest, segmentation based on size and shape thresholds.

Nuclei *size* and *shape* property values are computed using an object filtering method based on the variation of the shape and size features.

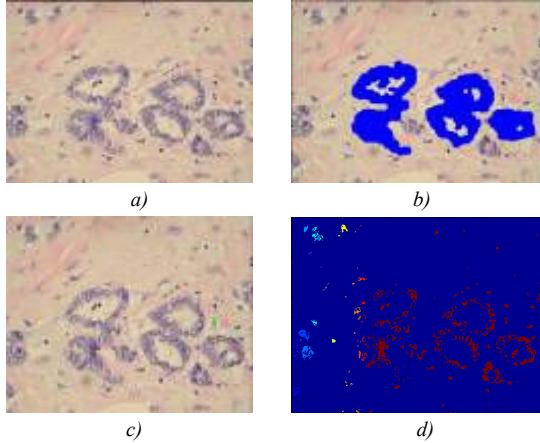


Figure 4. GRADING Rules Modeling Approach:
a) Original image b) TubuleFormationROI detection(blue) c) MitosisROI detection (green) d) NucleiROI detection

4. SEMANTIC INDEXING APPROACH. BREAST CANCER GRADING COMPUTATION

This approach intends to overcome the drawbacks of classical indexing methods. The conceptual annotations are rule-based defined in the grading model for every particular frame and globally transmitted in a structure for the entire case. The algorithm segments images and processes the object recognition phase followed by the semantic classification criterion rules modeling. Thus, it is created a correspondence between the visual features and the semantic image labeling, in terms of *Mitosis*, *Nuclei* and *Tubule Formation* regions of interest - ROIs.

Image segmentation with gray scale conversion and adaptive thresholding obtains a collection of such ROI, meaningful for breast cancer grading and – more generally – for breast cancer evolution diagnosis/prognosis. The region selection is correlated with the model rules. Features extraction step includes color, shape and size-based characteristics along with morphological operations and labeling ensued for the image analysis and semantic indexing (see Figure 4).

Semantic indexing of concepts extracted from the image give us the means to create the rules for the computation of local grading – the three criteria scores-over a single frame, global grading – the three criteria scores applied to 10 frames, respectively.

4.1 Local grading computation

Frame Tubule Formation Score (FTFS) :

$$FTFS = \{f(\text{Area}(TFROI) / \text{Area}(\text{DarkCellsCluster}))\}$$

$$f_{FTFS}(x) = \begin{cases} 1, x > 0.75 \\ 2, 0.10 < x < 0.75 \\ 3, x < 0.10 \end{cases}$$

The result of $f_{FTFS}(x)$ function gives the score for the tubule formation as reported in the pathologist rule (see Table 1).

Frame Mitosis Count (FMC)

$$FMC = \{f(\text{count}(MROI))\}$$

The result of the $f_{FMC}(x)$ function gives the grade for mitosis count according with the pathologic criterion (see Table 1).

FMC rule applies the *count* operator to M_{ROI} (mitosis regions of interest).

$$f_{FMC}(x) = \begin{cases} 1, x < 9 \\ 2, 10 < x < 19 \\ 3, x > 19 \end{cases}$$

For the nuclear pleomorphism score we use two functions: $f(\text{Size})$ and $g(\text{Shape})$ respectively, computed for all the nuclei. Their sum gives the full value of pleomorphism criterion of all nuclei counted with *count* operator used as the upper limit of the sum, related to the medical grading rule (see Table 1):

Frame Nuclear Pleomorphism Score (FNPS) =

$$FNPS = \{\text{round}(\sum_{i=1}^{\text{count}(NPROI)} (f(\text{Size}) + g(\text{Shape})) / \text{count}(NPROI))\}$$

The local breast cancer grading (FBCG)

$$FBCGi = \{f(FTFSi) + FMCi + FNPSi, i = \text{frameID}\}$$

as specified in the domain analysis consists of the sum of the three values computed for each criterion of the NGS.

4.2 Global grading computation

The global breast cancer grading is computed by taking into account the 10 HPFs, for each local score (see Table 2, Figure 5). The 10 HPFs specification appears as the upper limit at each computation of sum in the rules.

The rules thus, are defined as a computation of the *GTFS*, *GMC* and *GNPS* applied to 10 frames.

$$GTFS = \left\{ f_{TF} \left(\frac{\sum_{j=1}^{10} Area(TF_{ROI_j})}{\sum_{j=1}^{10} Area(Im_j)} \right) \right\}$$

$$GMC = \left\{ f_{MC} \left(count \left(\sum_{j=1}^{10} M_{ROI_j} \right) \right) \right\}$$

$$GNPS = \left\{ \frac{f_{NP} \left(\sum_{j=1}^{10} \left(\sum_{k=1}^{count(NP_{ROI_j})} (f(Size_{kj}) + gShape_{kj}) \right) \right)}{\sum_{j=1}^{10} count(NP_{ROI_j})} \right\}$$

$$GBCG = \{f(GTFS_j + GMC_j + GNPS_j), j = \{1, \dots, 10\}\}$$

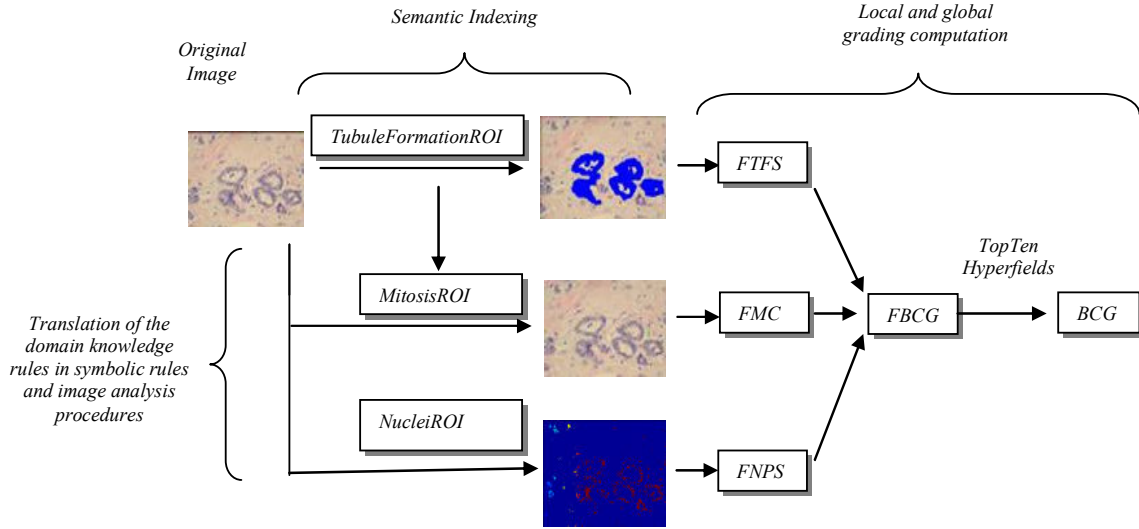


Figure 5. SEMANTIC Indexing in BCG Context

5. EXPERIMENTS & RESULTS

The experimental part consist in analyzing and indexing pathologic images of six breast cancer cases, consisting of 7000 frames scanned from the tumor tissue slides obtained through collaboration with the Pathology Department of the of the National University Hospital. The database is composed by two sets: 1400 frames are used for the training algorithm phase and 5600 frames are used for the testing and validation phase. The slides were scanned on a sequence of frames at 10X40 (400X) magnification with a 1080 x 1024 resolution.

The set of histopathology slides, labeled by our medical partners, have been digitized into a number of hyperfields (frames). Each frame is then analyzed and a local grading is computed. According to this local grading, top ten are automatically retrieved to provide a slide global grading.

Table 4. PATHOLOGIC visual grading and configuration of the training and testing database

Data type	Case ID	Tubule score	Nuclear score	Mitosis count	BCG (path)
Training database (1400 images)	1000	1	1	3	1
	2000	1	2	1	1
	4895	3	3	3	3
Testing database (5600 images)	5020	2	3	3	3
	5042	3	3	2	3
	5075	3	2	1	2

Table 5. AUTOMATIC grading results

Data type	Case ID	Tubule score	Nuclear score	Mitosis count	automatic BCG
Training database	1000	1	1	3	1
	2000	2	2	1	1
	4895	3	2	3	3
Testing database	5020	3	2	3	3
	5042	3	2	3	3
	5075	3	2	1	2

Table 6. COMPONENT scores and global grading errors

Data base	Tubule score	Nuclear score	Mitosis count	Component scores error	Global BCG error
Training errors	11%	11%	0	7,33%	0
Testing errors	11%	22%	0	11%	0

We use Matlab programming environment to develop the method, particularly the Image Processing toolbox for the designed algorithms in this study. The program is tuned to take into account the scale [15] of the images, given by the microscope in the automatic acquisition phase. Local errors were registered in the training base for the tubule score in one case (2000) and for the nuclear score in another case (4895). In the testing database, local errors were obtained at the tubule score and nuclear score for the same case (5020) and only for the nuclear score in another case (5042). Note that for the mitosis count there was no registration in either training or testing database which gives us a good confidence degree in the detection of mitosis. (100% automated detection). Within the 6 cases, the nuclear encountered 50% matched results followed by tubule score with 66% matched results.

The most interesting fact is that, when computing the BCG for training and testing database respectively, local errors are not propagated to the global level. Therefore the Global BCG error is 0, while the component errors are 7.33%, 11% respectively computed by a simple formula of matches from the total items.

The good results obtained on the global grading are promising and allow us to envisage interesting generic perspectives of this approach.

6. DISCUSSIONS, CONCLUSION AND PERSPECTIVES

Even if strongly related to a particular application field and specific medical domain, the presented semantic labeling approach has a generic character. Indeed, in association with localization and quantitative information, this meaningful indexing allow to design semantic query content-based medical image retrieval systems, usable in evidence based medicine framework. Those types of CBIR systems will certainly replace in the near future the actual query by example ones, based only on visual features.

In the context of virtual microscope platforms, automatic semantic-query based visual positioning systems [16] present also a strong interest for the medical technicians and doctors. Transmitting simple

semantic (textual or vocal) requests, allow in this case saving precious time to those professionals by automatically positioning the system according to a specific request (e.g. search for the hyperfields of the slide with the most irregular nuclei shape, most important number of mitosis, and so on). In this sense, beside doctors and technicians, the students, teachers and researchers will also be interested into the use of such a system for a deeper study of a given pathology.

Finally, the purpose of generating computer vision (CV) concepts and symbolic rules from medical concepts/rules related to the breast cancer grading, with the respect of the Ontology Web Language (OWL) and the Semantic Web is expected to future generic perspectives for an assisted semi-automatic generation of CV rules and computer programs, starting from specific medical queries/rules.

ACKNOWLEDGMENT

This project is partially supported by the ONCO-MEDIA¹ project. We would like to thank to Pr. Teh Ming, head of the Pathology Department of the National University Hospital from Singapore for his constant support in this joint study.

REFERENCES

- [1] A. Tutac, "Histological Grading on Breast Cancer", *IPAL internal report 2007*, MIIRAD/IPAL – BCG, 2007
- [2] I.Marandet, A.Tutac, "Smart Microscope User Guide", *IPAL internal report 2006*, MIIRAD/IPAL -μ-MediSearch, 2006
- [3] H. Muller, N. Michoux, D. Bandon, and A. GeissBuhler, "A Review of Content-Based Image Retrieval Systems in Medical Application- Clinical Benefits and Future Directions", *International Journal of Medical Informatics*, vol. 73, pp. 1-30, 2004
- [4] S. Petushi, F. Garcia, M. Haber, C. Katsinis, and A. Tozeren, "Large- Scale Computation on histology images reveal grade- differentiating parameter for breast cancer", pp. 1-11, 2006
- [5] G. Carneiro, A. Chan, P. Moreno, and N. Vasconcelos, "Supervised Learning of Semantic Classes for Image Annotation and Retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.29, no.3, 2007
- [6] T.Lehman, M.Guld, C.Thies, B.Fischer, K.Spitz, D.Keysers, H.Ney, M.Cohnen, H.Schubert, and B.Wein, "Content-based image retrieval in medical applications", *Methods of Information in Medicine*, vol. 43, no. 4, pp.354-361, 2004
- [7] P.Enser, C.Sandom, P.Lewis, and J.Hare, "The Reality of the Semantic Gap in Image Retrieval", *tutorial held in*

conjunction with the 1st International Conference on Semantics and Digital Media Technologies, 2006

[8] Y.K.Alfoglou, S.Dasmahapatra, D.Dupplow, B.Hu, P.Lewis, N. Shadbolt, "Living with the Semantic Gap: Experiences and Remedies in the Context of Medical Imaging", *1st International Conference on Semantics and Digital Media Technologies*, 2006

[9] P.Duygulu, K. Barnard, J.F.G de Freitas, D.A.

Forsyth, "Object Recognition as machine translation: Learning a lexicon for a fixed image vocabulary", *Proceedings of the 7th European Conference on Computer Vision*, part IV, pp. 97-112, 2002

[10] J.Leon, V.Lavrenko, and R.Manmatha, "Automatic image annotation and retrieval using cross-media relevance models", *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM Press, pp.119-126, 2003

[11] S.Little and J.Hunter, "Rules-By-Example- A Novel Approach to Semantic Indexing and Querying of Images", *International Semantic Web Conference ISWC*, pp.534-548,

2004

[12] Y.Liu, N.Lazar, W.Rothfus, F. Dellaert, A.Moore, J.Schneider, and T.Kanade, "Semantic - based Biomedical Image Indexing and Retrieval", *Trends and Advances in Content- Based Image and Video Retrieval*", Shapiro, Kriegel and Veltkamp ed., pp. 1-20, in press, 2004

[13] H.Tang, R.Hanka, and H.Ip, "Histological Image Retrieval Based on Semantic Content Analysis", *IEEE Transaction on Information Technology Medicine*, vol. 7, no. 1, 2003

[14] D.L. McGuinness and F.van Harmelen, "OWL Web Ontology Language W3C Overview", pp. 1-26, 2004

[15] P. Van Osta, J.M. Geusebroek, K. Ver Donck, L. Bols, J. Geysen, and B.Romeny, "The Principles of Scale Space applied to structure and color in light microscopy", *Proceedings RMS*, vol. 37, no. 3, 2002

[16] G. Begelman, M. Lifshits, and E. Rivlin, "Visual Positioning of Previously Defined ROIs on Microscopic Slides", *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, 2006