Interpretation breast cancer imaging by using ontology

Benmarouf Meriem, Tlili Yamina

Abstract— **Ontology-based software and image processing engine must cooperate in new fields of computer vision like microscopy acquisition wherein the amount of data, concepts and processing to be handled must be properly controlled. Within our own platform, we need to extract biological objects of interest in huge size and high-content microscopy images. In addition to specific low-level image analysis procedures, we used knowledge formalization tools and high-level reasoning ability of ontology-based software. This methodology made it possible to improve the expressiveness of the clinical models, the usability of the platform for the pathologist and the sensitivity or sensibility of the low-level image analysis algorithms.**

Keywords— **Histopathology, image processing, medical imaging, ontology, interpretation.**

I. INTRODUCTION

Similar to satellite imaging, digitized pathology faces new challenges; digitized pathology will raise new major issues both from technological and scientific points of view. On the more technological side of the research scope, storage and networking is dramatically challenging since, unlike mammograms for instance, the amount of data easily scales up to eight gigabytes of pixels for a single biopsy patient case. At a more fundamental level, the high number of meaningful biological concepts to be handled, sometimes implicitly, calls for innovative ways to handle information in huge visual data.

We proved in this work that image processing at the signal level embedded with high-level interfaces to interact with the system can improve not only the ease of use of such systems for the end-user but also the robustness of the results in daily practice, above all in fields where, paradoxically, subjective decision can be taken like in the case of breast cancer grading. Few attempts to bridge the gap between knowledge managing and medical image analysis outcomes have been carried out so far [2]-[3] and mostly about highly atlas-based or informed anatomical structures like the brain. Works on microscopy images with an intrinsically higher content of either explicit or implicit biological objects are much fewer [5].

The paper is organized as follows. Section 2 overviews the challenges in breast cancer grading not only from a purely clinical point of view but also out of digitized pathology images. Section 3 describes the low-level image processing machine while Section 4 elaborates on the high level interface based on ontology capabilities we embedded in our system in order to improve the overall performances of the grading process as assessed in Section 5. Last, Section 6 draws meaningful conclusions about the interactions between the ontology paradigm and computer vision achievements so far leading to major perspectives for the intelligent system designer community.

II. BREAST CANCER GRADING

A. Digitized Pathology

The European Virtual Physiological Human project is an ambitious program to build up a holistic in silico functional model of the human biology system. To do so the design of inter-operable formats for biology system modeling or bio-engineering workflows are of utmost importance. Multi-scale modeling of the human body functions is as a well an expected outcome of this scheme. The new advances in microscopy image acquisitions will be a definite asset to fulfill this goal. Digitized pathology in particular has been recently standardized even at the DICOM level and will provide a bunch of valuable visual insights about the way biological phenomena proceed. Modeling at the cellular level will be eased as well. Ontology should be a key player to share models in this framework.

Even for clinical daily practice, computer vision systems involving knowledge management are the only way to make them really usable at a large scale and even sustainable in terms of design, maintenance and usability. Moreover, reasoning capabilities can leverage image analysis ones to improve the final diagnostic process for instance. Grading breast cancer out of histopathological images is the gold standard for clinicians drawing prognosis reports of such diseases and is known to be still challenging: in terms of reproducibility mostly but also in terms of training and legal assessment due to the lack of traceability and archiving features.

From a learning standpoint as well as from a training perspective, the clinician uses a “mental database” of visual cases that helps him to go straight to the relevant Region of Interest (ROI) over the microscope screening, then switch to the optimal magnification and then to perform the standard computation of scores following the Nottingham Grading System (NGS) for breast cancer grading. The grade is a combination of three scores:

1) *Nuclear Pleomorphism:* if uniform cells (minimal or no nuclear enlargement, long axis diameter ~10µm, minimal or no darkening of chromatin) then score=1; if moderate nuclear size and variation (long axis diameter ~15µm), then score=2; if marked nuclear variation (long axis diameter ≥20µm), then score=3.

2) *Tubular Formation*: if ≥ 75% of the invasive area is forming tubules, then score=1; if 10−75% of the cancer is forming tubules, then score=2, else if ≤ 10% of the cancer is forming tubules then score=3. (Only structures exhibing clear central lumina are counted) .

Then, the final grading mark is a linear combination of the three previous scores: Grade=Score tubule + Score mitoses + Score nuclear pleomorphism. Low grade (I) breast cancers

correspond to a sum of 3-5, Intermediate grade (II) to 6-7 and High grade (III) to 8-9. An ontology support to the cross design of a clinical system is almost an almost inescapable requirement to handle such a broad scope of, sometimes evolving, concepts. To start, we designed breast cancer ontology both anatomical and clinically operational [6].

B. Anatomical And Workflow Ontology: OWL

Ontology is a system of knowledge representation of a domain in the form of a structured set of concepts and relationships between these concepts. Ontology is expressed in the form of a XML graph and produces reasoning through a rule language. Our Breast Cancer Ontology (BCO) is based on two languages: OWL-DL (Web Ontology Language Description Logics) to describe the ontology and SWRL (Semantic Web Rule Language) to write and manage rules for the reasoning part. Technically, OWL and SWRL are specifications of the W3C, OWL is an extension of RDF (Resource Description Framework) used in the description of classes and types of properties, SWRL combines OWL and RuleML (Rule Markup Language) to produce the rules for the reasoning. The annotated images are described with the wide Field Markup Language (WFML) specific to the histopathology field. Finally, the query language SPARQL (Simple Protocol And RDF Query Language) is used for querying in Java. SPARQL has been chosen for its ease of use and the very good integration of the API in Java. A thorough description of this ontology-based platform can be found in [6]-[5].

III. Low Level Image Annotation

The image processing machine provides a priori visual landmarks related to ontological biological concepts like nucleus, mitosis, and tubule. As this low-level processing machine is not the core of our current contribution, technical details will be skipped over while snapshot illustrations of the resulting visual landmarks population over the WSI are provided.

Authors should consider the following points:

Invasive Area segmentation as a pre-attentive processing, the invasive Region Of Interest (ROI) detection is currently casted as a classification problem whereby we exploited the relationship between human vision and neurosciences [1]. An illustration of the focusing step in our platform is provided in Fig. 1.

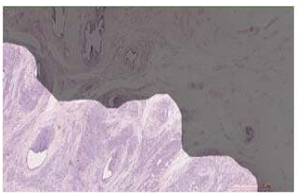


Fig. 1 Invasive area pre-attentive detection step at low magnification x1.2 over the WSI

Nuclei, Mitoses and Lumina extraction: The nuclei detection module is the core image analysis module of the system in the sense that it should be the more robust low level process due to a quite standardized staining process in clinical daily practice. The nuclei detection proceeds following two steps as presented in [6] and the results are illustrated in Fig. 2. As the metric scale is known, most image processing is related to the mathematical morphology toolbox using shape and size criteria. Then geometric and radiometric features can be extracted over each detected nucleus.

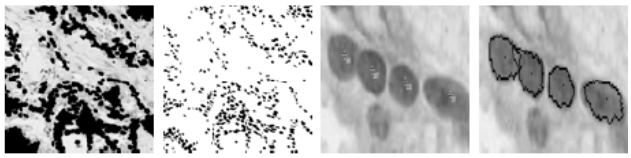


Fig. 2 *Nuclei identification*

The low-level mitosis detection module proceeds by machine learning based on radiometric and geometric features computed out of a ground truth database. Results are illustrated in Fig. 3.

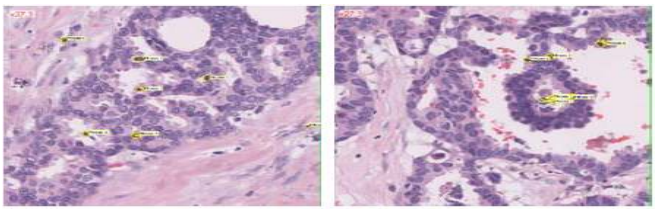


Fig. 3 Mitosis *detection by machine learning based on geometric and radiometric features*

The lumina are void parts in the tissue that let fluid or air pass through. The low-level detection of the lumina uses mathematical morphology tools, represents in Fig. 4, to detect bright blob areas in the WSI. They can be confused with fat matter zone or tubular formation.

Fig. 4 – Lumina detection

IV. High Level Image Annotation

Once all these visual landmarks have been potentially computed by tedious signal processing formalization, learning and numerical implementation of visual characteristics, high-level knowledge representation and handling can enhance the efficiency of the virtual microscope system mostly because the extraction of all the biological concepts by an exhaustive search is not possible in an interactive time.

By designing vision system through the ontology framework, our research work objectives are threefold:

- Consistency checking annotation: to improve the specificity rate.
- Image Analysis Engine Triggering Control: to improve the sensibility rate within a limited response time.
- Smart and Adaptive Interface: Consider the end-user as a key player of the system design and functionality in relation with the two previous objectives.

We leverage both on the knowledge formalization capabilities and the reasoning features of platforms like protégé to achieve higher level of interoperability, usability and potentially robustness of the system.

A. Rules and reasoning

The core numerical object of the enhanced system is the XML-like file that stores the annotations of the WSI in the WFML format. The WFML files are translated into OWL file formats both related to the XML technology. Then, we can use the first order logic machine inference within the protégé environment. For each WSI, the system is able to generate complementary annotation outputs: Rsignal and Rknowledge as described in the current section. We study hereafter the articulation *R*signal×knowledge between the two sets of annotation outputs and illustrate how high-level processing can improve the overall behavior of the virtual microscope system through the three previously mentioned objectives.

1) User-based Consistency Checking Annotation: mitosis detection: In histopathology, the biological concepts are usually expressed as high-level concepts while image analysis modules provide actually implicit definitions of these concepts. Ideally, a fully-fledged smart vision platform should provide a way of checking the consistency of the low-level numerical annotations with their high-level ontological definitions: the famous semantic gap. So anchoring the histopathological concepts in the digitized WSI can benefit of a cross validation between (a) the low-level, implicit, signal based extraction providing a set of results Rsignal , usually by statistical learning and tedious numerical modeling and (b) the explicit high-level description corresponding to a SWRL rule like the one expressed in the Protégé platform in Fig. 5 and potentially providing a set of results Rknowledge, and where Circularity and Roundness are the standard shape features.

→ Nucleus(?x) ∧ hasIntensity(?x, ?value) ∧ swrlb:lessThan(?value, 110.0) ∧ hasCircularity(?y, ?cir) ∧ swrlb:lessThan(?cir, 0.75) ∧ hasRound(?z, ?round) ∧ swrlb:lessThan(?round, 0.65) → Mitosis(?x)

FIG. 5 A SWRL rule for mitosis description in our BCO (Breast Cancer Ontology) within the Protégé platform

Our platform implements these two kinds of mitosis detection: the R(signal, mitosis) as slightly described in Section 3 and the R(knowledge ,mitosis) that relies on the R(signal,nucleus) set of results. The whole annotation updating workflow for the mitosis detection is described in figure 6 and in figure 7 for the WFML-based annotated resulting images 7. Step 2 in the workflow provides R (signal, nucleus), step 8 yields R(knowledge, mitosis) and step 2 outputs R(signal,mitosis). Section 5 elaborates on the *nucleus*, step 8 yields R(knowledge, mitosis) and step 2 outputs R(signal,mitosis). Section 5 elaborates on the synergy between the two interactions modes in order to achieve the Rsignal×knowledge mitosis final objective.

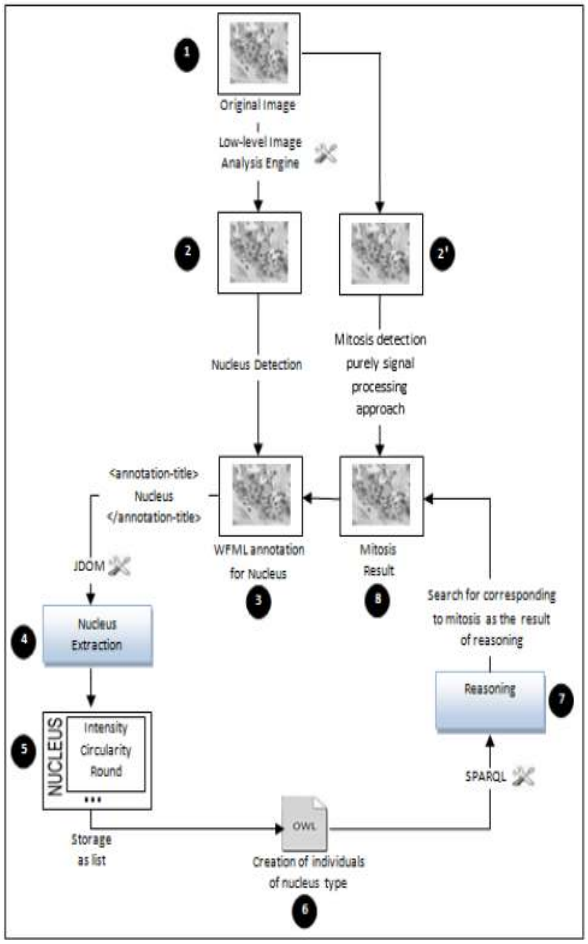
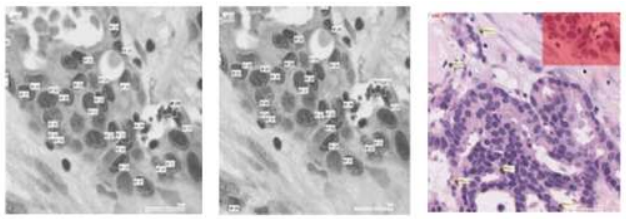


FIG. 6 Mitosis detection process

2) Spatial RelashionShip modeling: Spatial configuration Consistency Checking: tubule detection. Along the same line, tubule detection can be achieved by high level spatial configuration reasoning or constraint checking. Let us assume that a sound definition of a tubule is a lumina

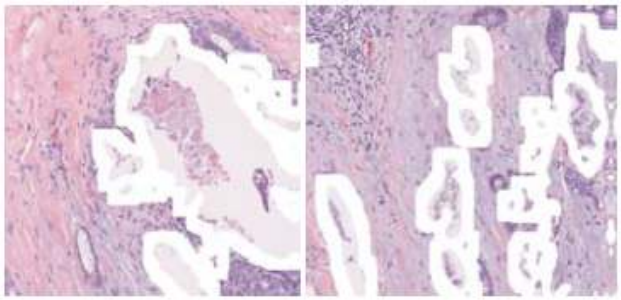
surrounded by two lines of cells as reported in academic books about pathology. This definition permits among others to discriminate between mere fat areas and tubular configurations that both correspond to large bright blob zones in the WSI. From an image analysis point of view, we need to formalize spatial relationships concepts like Surrounded by in a sound, theoretical way. That is done by the use of the mathematical morphology toolbox like in [2] represents in Fig.8. Image Analysis Engine Triggering Control: mitosis detection. Interactive time is a fundamental issue in current image processing systems to really comply with user requirements. In particular in digitized histopathology, we must fit in a ten-minute response time frame on a par with the daily practice clinician time scores. Thence, being able to control the image analysis triggering over the WSI can help to improve the sensibility rate of the platform under the time constraint.



(a) (b) (c)

FIG. 7 (a) WFML before Mitosis Detection corresponding to step 3 in FIG. 6 (b) WFML after Rule-based Mitosis Detection corresponding to step 8 (c) WFML after low-level based Mitosis Detection corresponding to step 2.

The highlighted zone at the top-right corner focuses on the zone in (a) and (b) and no mitosis is detected inside it. Mitoses detected by the low-level engine are annotated in the rest of the image.



(a) (b)

FIG. 8 Formalization of the surrounding areas of a lumina

Referring to pathologists' explicit knowledge, the rule stating that mitoses should be first searched at the periphery of invasive areas can be expressed in the first order Protégé logic represents in Fig.9.The spatial relationship around is closely related to the previous one Surrounded by and its modeling relies on the same mathematical morphology tools. When applicable, the system was able to save between five and ten-fold processing time which is of dramatic importance for WSI exploration. In addition, this kind of spatial relationship rules can help to check the consistency of Rsignal.

$\rightarrow Mitosis\ (?x)$.

FIG. 9 A SWRL rule for the expression of the spatial relationship constraint within the Protégé platform

V. RESULTS

A. *Mitoses Detection*

We chose to thoroughly assess our methodology on the mitosis detection problematic which is determinant according to pathologists. Tables in figure 10 sum up major recognition rates for their detection respectively with ontological intensity and geometrical criteria for the first two ones and with the purely signal processing approach for the last one. The Rsignal processing results over detect the mitoses over the ten frames with 50 detections against 20 for the Rknowledge ones. The ontology-driven detection in-creases the specificity of the system by dramatically reducing the false alarm rate. Rsignal outperforms Rknowledge in the case of nuclear pleomorphism score 3 images that correspond to cases wherein the segmentation process is by far less robust and reliable because of treacherous deformations intrinsically and within the cells (for instance specificity rate of score 3 image NB19413 in comparison with score 1 image NB7824). On the contrary, Rknowledge overcomes Rsignal in the case of nuclear pleomorphism score 1 images. The better the low-level segmentation process is, the better the ontology-driven approach behaves.

Mitosis detection with ontological approach : query with intensity

Image	Frame	Nb of cells	Correct Nb. of mitosis	Nb of mitosis detected	TP	TN	FP	FN	Specificity	Sensitivity
NB19413	f001	257	3	6	3	251	3	0	0,988	1,000
NB19413	f003	314	2	4	2	310	2	0	0,994	1,000
NB19413	f008	296	2	3	2	293	1	0	0,997	1,000
NB7824	f001	244	0	8	0	236	8	0	0,967	1,000
NB7824	f004	242	0	18	0	224	18	0	0,926	1,000
NB7824	f002	297	2	4	2	293	2	0	0,993	1,000
NB19271	f002	261	0	6	0	255	6	0	0,977	1,000
NB19271	f003	214	0	1	0	213	1	0	0,995	1,000
NB19271	f011	229	0	5	0	224	5	0	0,978	1,000
NB19271	f016	234	0	5	0	229	5	0	0,979	1,000
		2588	**9**	**60**	**9**	**2528**	**51**	**0**	**0,980**	**1,000**

(a)

Mitosis detection with ontological approach : query with intensity and geometrical constraints

Image	Frame	Nb of cells	Correct Nb of mitosis	Nb of mitosis detected	TP	TN	FP	FN	Specificity	Sensitivity
NB19413	f001	257	3	2	2	255	0	1	1,000	0,667
NB19413	f003	314	2	3	2	311	1	0	0,997	1,000
NB19413	f008	296	2	3	2	293	1	0	0,997	1,000
NB7824	f001	244	0	0	0	244	0	0	1,000	1,000
NB7824	f004	242	0	0	0	242	0	0	1,000	1,000
NB7824	f002	297	2	3	2	294	1	0	0,997	1,000
NB19271	f002	261	0	5	0	256	5	0	0,981	1,000
NB19271	f003	214	0	1	0	213	1	0	0,995	1,000
NB19271	f011	229	0	3	0	226	3	0	0,987	1,000
NB19271	f016	234	0	0	0	234	0	0	1,000	1,000
		2588	9	20	8	2568	12	1	0,995	0,889

(b)

Mitosis detection based on purely signal processing approach

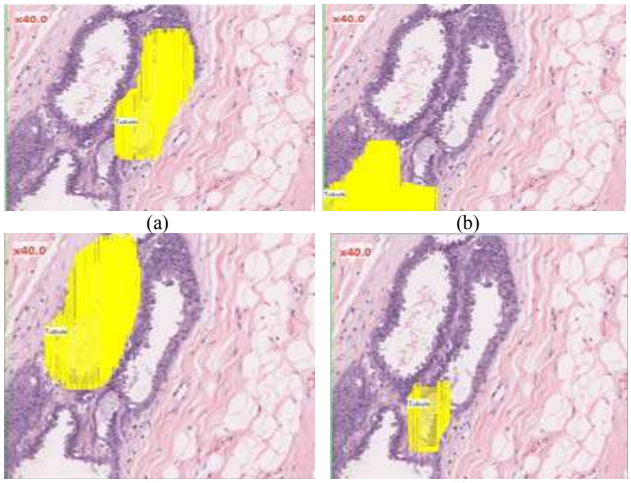
Image	Frame	Nb of cells	Correct Nb of mitosis	Nb of mitosis detected	TP	TN	FP	FN	Specificity	Sensitivity
NB19413	f001	257	3	5	3	252	2	0	0,992	1,000
NB19413	f003	314	2	3	2	311	1	0	0,997	1,000
NB19413	f008	296	2	3	1	293	1	0	0,997	1,000
NB7824	f001	244	0	7	0	237	7	0	0,971	1,000
NB7824	f004	242	0	7	0	235	7	0	0,971	1,000
NB7824	f002	297	2	3	1	294	1	0	0,997	1,000
NB19271	f002	261	0	5	0	256	5	0	0,981	1,000
NB19271	f003	214	0	5	0	209	5	0	0,977	1,000
NB19271	f011	229	0	11	0	218	11	0	0,952	1,000
NB19271	f016	234	0	1	0	233	1	0	0,996	1,000
		2588	9	50	7	2538	41	0	0,984	1,000

(c)

FIG. 10 Mitosis detection major recognition rates with query based on (a) intensity ontological query (b) intensity ontological query and (c) purely signal processing approach

B. *Tubule Detection*

As for tubule detection, as more complex and versatile objects, first results based on spatial relationships sound modelling and appropriate triggering in the field of visual reasoning provide clear insights about applications for exploration of huge images like WSIs. Figure 11 shows how it is possible to annotate regions like tubular formation against fat matter zones or simple lumina over the WSI. In addition, figure 11(e) provides preliminary quantitative assessment of the ontology-driven tubule detection.



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

FIG. 11 An example of (a) tubular detection against (b) mere fat and lumina zones by involving visual reasoning procedure (c) preliminary quantitative assessment of ontology-driven tubule detection

Discussion. From a formal point of view, hypothesizing that the nuclei detection is the basic axiom of our system (that is we exclude it from the Rsignal set of results) or not, the tubular formation module in our system can be considered either as purely a Rknowledge tubule result or actually a Rsignal×knowledge tubule . But as any vision system will need a basic detection module, we consider that the tubule detection module is purely an Rknowledge tubule result. If a low-level, implicit, signal processing was able to provide a Rsignal tubule (which might be very challenging according to our experience) then a collaborative synergy between the two approaches could help improve the interaction with and the robustness of the system. This general discussion applies to the mitosis detection case, and in general to all complex biological concepts to be extracted, even discovered, over biological visual material. Thence, the global formalization of Rsignal×knowledge object extends itself to any kind of complex, versatile object detection over huge image like in current high-resolution satellite images database.

VI. CONCLUSION

While developing a new paradigm of virtual cognitive microscope for the exploration of high-content microscopy images, we proved that articulating knowledge management capabilities with low-level image analysis modules can not only improve the design of the system but as well increase the performance of the system. In particular, when the high level reasoning module can rely on fair low-level segmentation outcomes, the knowledge - and so the user - in the loop can increase the specificity rate of the system for mitosis detection for instance. In addition, high-level semantic queries are made available by the formalization of spatial relationships between biological objects. This kind of spatial reasoning is a definite asset to discriminate structures from a structural point of view much more than from a purely radiometric one like tubular against fat zones or, as perspective, subtle differences related to Ductal Carcinoma in Situ.

REFERENCES

[1] R. Miikkulainen, J. Bednar, Y. Choe, and J. Sirosh, "Computational Maps in the Visual Cortex," Springer, 2005.

[2] C. Hudelot, J. Atif, and I. Bloch, "A spatial relation ontology using mathematical morphology and description logics for spatial reasoning," ECAI Workshop on Spatial and Temporal Reasoning, pp. 21–25,2008.

[3] A. Mechouche, X. Morandi, C. Golbreich, and B. Gibaud, "A hybrid system using symbolic and numeric knowledge for the semantic

annotation of sulco-gyral anatomy in brain MRI images," IEEE Transactions on Medical Imaging, pp. 1165–1178, 2009.

[4] L. Roux, A. Tutac, N. Lomenie, D. Balensi, D. Racoceanu, W.-K. Leow, A. Veillard, J. Klossa, and T. Putti., "A cognitive virtual microscopic framework for knowlege-based exploration of large microscopic images in breast cancer histopathology," IEEE Engineering in Medicine and Biology Society, Minneapolis, Minnesota, USA, Sept.2009.

[5] A. E. Tutac, D. Racoceanu, W. K. Leow, J. R. Dalle, T. Putti, W. Xiong, and V. Cretu, "Translational approach for semiautomatic breast cancer grading using a knowledge-guided semantic indexing of histopathology images," In MIAABMICCAI Workshop, New-York, USA, Sept. 2008.

[6] J.-R. Dalle, H. Li, C.-H. Huang, W.-K. Leow, D. Racoceanu, and T. C. PuttiNuclear, "Pleomorphism scoring by selective cell nuclei detection," IEEE Workshop on Applications of Computer Vision, Snowbird, Utah, USA,2009.