Biomedical Image Mapping: Issues and Challenges

Nurzi Juana Mohd Zaizi[1] and Albert Burger[1,2]

[1]Department of Computer Science
Heriot-Watt University
EH14 4AS, Edinburgh, Scotland, UK
[1]*Email: nz27 {at} hw.ac.uk*

[2]MRC Human Genetics Unit
Edinburgh, EH4 2XU, UK

Abstract—**Biomedical imaging has become an essential tool in both basic research and the clinical sciences. Significant advances in imaging modalities such as computed tomography (CT) or magnetic resonance imaging (MRI) in recent years have contributed to a substantial increase in the quantity and quality of such images. Being able to integrate and compare such image-based data is useful for epidemiological studies, educational uses, monitoring the clinical progress of a patient or translational science purposes. The work presented in this paper uses examples from biomedical atlases. This paper discusses the need for biomedical image data integration, the shortcomings of existing solutions and proposes the opportunity to build a new technique.**

Keywords-component; spatial relations, image region mappings, biomedical atlases, ontologies, image processing algorithms

I. INTRODUCTION

A biomedical atlas consists of a graphical model, the ontology associated with the graphical model and a mapping between those two. The ontology contains a collection of anatomical domains and relations between those domains. The graphical model is a digital image of an object (e.g., of a human or animal body) along with the identified anatomical domains. Three of the main atlases in the current domain of interest are the e-Mouse Atlas of Gene Expression (EMAGE) [1], the Allen Developing Mouse Brain Atlas [2], and the GENSAT Brain Atlas [3]. These atlases are the data resources for gene expression information. Gene expression information describes whether or not a gene is expressed in a location of a particular anatomical structure relating to a model organism [4]. This research focuses on the mouse embryo as the model organism.

The Allen Developing Mouse Brain Atlas is a data source storing gene expression data across seven developmental stages of the mouse brain [5]. EMAGE [6] is another example of a mouse atlas covering gene expression data for anatomical structures corresponding to the EMAP Anatomy Ontology [7]. Gene expression data for the mouse brain is also available from EMAGE. Another example of a mouse atlas that provides gene expression data for the mouse brain is the GENSAT brain atlas. GENSAT is a gene expression atlas of both the developing and adult mouse, and stores gene expression data for anatomical structures corresponding brain and spinal cord [3].

Although the EMAGE, the GENSAT brain atlas and the Allen Developing Mouse Brain Atlas are fundamental resources, they cannot be considered complete. There are a number of explanations for this, including differences in the experimental designs and various interpretation of results [8]. In addition, different update routines may cause data from these atlases to remain incomplete. As a consequence, these atlases may provide different results even for the same gene expression query.

In addition to incompleteness, these atlases can suffer from inconsistency [8]. Inconsistency implies when one biomedical atlas publishes an annotation suggesting the gene is expressed in a particular structure, and a second annotation suggesting it is not [9]. Such variability is often associated with the complexity of the underlying experiments, including unrecognised differences in experiments, and human error on the part of the resource's curators [9]. All resources must be exploited to generate a full and complete query results. Therefore, the integration of anatomical space in the context of same-species atlases can provide the solution to facilitate the sharing of biomedical data from many resources.

The mapping of anatomical space across different model organisms [10], such as the linking of human and the mouse embryo model organisms, can be useful to facilitate the integration of biomedical data between cross-species atlases. Integrating data from cross-species atlases is useful for the expression of homologous genes experimental field [11]. Corresponding organs and genes in different species can be homologous. Homologous genes are the same form of genes expressed in a particular body part of two different species. These sets of genes can be compared meaningfully to facilitate analysis, modelling and prediction in biomedicine [12; 13; 14]. For example, integrating data between the mouse atlas and the chick atlas to facilitate the comparison of homologous genes between these organisms assists biologists in understanding which genes are important in foetal development [15]. Consequently, integration of biomedical atlases is vital both within and across species boundaries.

Although there are many different methods to integrate data from these resources, the use of image-based data integration is considered in this research. Image-based data integration typically involves anatomical space mapping. Given two

images *C1* and *C2*, mapping one image onto another means that, for each anatomical region in image *C1*, we try to find a corresponding region with the same intended meaning, in image *C2*. These anatomical regions occupy unique anatomical spaces in the corresponding images. Of the existing solutions to this problem, ontology-based solutions tend to lack spatial precision. Image processing-based solutions have difficulties when the underlying morphologies are too different.

This paper discusses the issues related to image representation and mappings. An overview of image mapping problems is provided in Section II. Section III explains the existing mapping solutions and describes how these solutions may be used to deal with image mapping issues. The problems to be addressed by this research are discussed in Section IV. Conclusions of this research are presented in Section V.

II. IMAGE MAPPING PROBLEMS OVERVIEW

A. Image Mapping Classiffication

There are two categories of image mappings. The first case involves a single query image and a set of potential target images and the matching process should return those images from the target set that match the query image by some notion of equivalence. The second case is given a drawn query region in one image, find the corresponding region in another image that matches the query region. Either of these cases can be classified into two subcases. The first subcase is the images with painted domains. The second subcase refers to images with no painted domain.

B. Image Segmentation

Image segmentation involves breaking an image down into its basic components or regions. Before we can use the anatomy ontology to label the painted domains, we must perform image segmentation. The common methods of image segmentation are segmenting an image based on colour, boundary and shape. We use the words *painted* and *segmented* interchangeably. The term paint denotes a technique of drawing on an image region. The purpose of painting is to tag a region according to an anatomy ontology. MAPaint [16] software provides the painting mechanism. MAPaint is a 3D painting program based on the woolz image processing library. The software allows the definition and saving of arbitrarily complex 2D spatial domains by painting over the appropriate region. The basic idea of painting is for image segmentation. The purpose of segmenting an arbitrary region is to partition a 2D image space into multiple segments. The result of image segmentation is a set of regions that collectively cover the entire image.

C. Variation of Morphologies

In general, the two cases of image mappings can be carried out between images that have exactly the same morphologies, as well as between images that have different underlying morphologies. Morphology concerns the different formation of a particular anatomical structure in terms of scale, orientation and position. Fig. 1 depicts two image slices of an

embryo. These images are good representatives for the non-identical images with different morphologies. Both images contain the following structures: liver, heart, and lung. However, these structures are different in terms of their scales and positions. For example, in terms of scale, the liver in Fig. 1(a) appears slightly smaller compared to the liver in Fig. 1(b). In terms of position, the heart in Fig. 1(a) is located above the liver and the lung. In contrast, the heart in Fig. 1(b) is located between the liver and the lung.
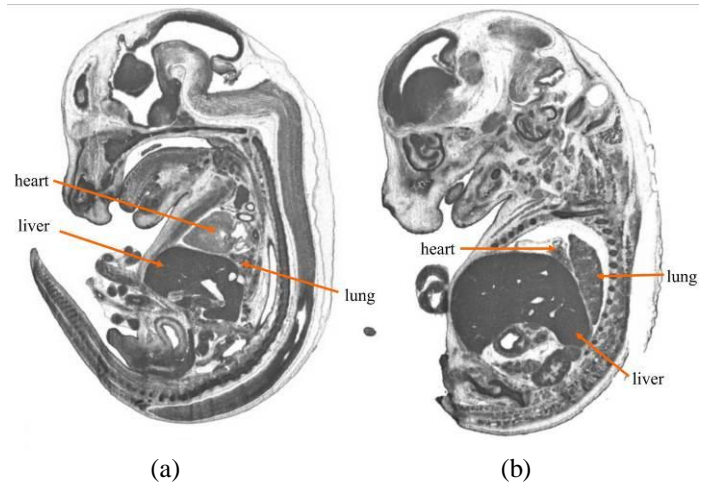

(a)   (b)

Figure 1. Two non-identical mouse embryo images with the same structures: liver, heart, and lung. However, these structures are different in their morphologies.

Fig. 2 depicts two images of an embryo across different visual domains. The image in Fig. 2(a) is the clip art graphic version of the image in Fig. 2(b). The visual content of both images is similar at the higher scene level, but both images are entirely different at the pixel-level. These images are good representation of the non-identical images with the same morphology.
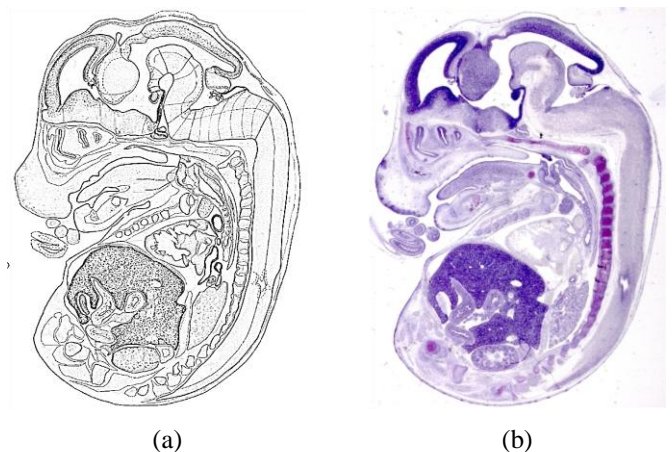

(a)   (b)

Figure 2. Two non-identical mouse embryo images with morphologically the same set of structures in terms of scales,

orientations and positions. The visual content of both images is only similar at the higher scene level, but both images are entirely different at the pixel-level.

Although the two images are dissimilar at the pixel-level, both images are morphologically the same set of structures in terms of their scales, orientations and positions.

D. Criteria for Matching Images/Regions

In a case in which images are lacking painted domains, the criteria to consider whether two images match or two regions match can be based on the equivalence of their low-level image features. Commonly extracted features include colour, texture and shape. These low-level image features can either represent a whole image or a specific region. A notion of equivalence is quantified using either the combination of several low-level features or making use only on a particular feature.

Colour information can be extracted from an image and represented as colour histogram [17; 18]. A global colour histogram can be used to define the number of pixels that have colours in each of a fixed list of colour ranges for a whole image. Two images match when their colour histograms match. Similarly, image regions matching can also be performed using colour histograms. A local colour histogram can be used to define the colour information for a specfic region of an image. An image can be divided into several regions and a colour histogram can be created for each region. Two regions match when their local colour histograms match.

Texture is defined as properties related to the appearance and feel of a surface. It can be categorised into two types. The first is stochastic, meaning rough, grainy and irregular; and second is structural, or having a regular and smooth surface. The main characteristics of the texture are distinctive and repetitive over a region. Texture features are useful in the comparison of equivalence between two images. The equivalence between two images can be compared based on texture matching. Texture matching is performed by extracting the texture features from the images and the images are segmented into regions; each containing a homogeneous texture pattern [19; 20]. Two images match or two regions match when they are homogeneous in terms of their texture patterns.

Shape matching is performed by comparing the region based edge features. Edge features extracted from images are considered as point sets. For a point on the first image, it is expected to find the best matching point on the second image according to the edge features. Two images match when their feature point correspondences match. Similarly, image regions matching can also be performed by comparing the region-based edge features. Edge features extracted from one specific region are considered point sets. A region matches another when a point on a region in one image is found on the region in the other image, according to the edge features.

In the case of images with painted domains, the criteria to consider whether two images match or two regions match can be based on the equivalence of their high-level semantic classifications based on the painted regions; an image or a specific region in one image can be classified into categories which are intended to distinguish semantically meaningful differences [21]. Spatial knowledge representations are useful for describing the spatial relationships among the painted regions in an image. Commonly use spatial knowledge representation is the topological properties. Topological properties include the number of subregions and the relationships between the properties of the subregions and the regions [22]. These topological properties are examples of meaningful semantics which may be derived from the content of an image. For example, if a structure is seen, it can be connected to some previously learned spatial concepts (e.g. lies to the right of the stomach and overlies the gallbladder), which can be used to recognise this structure is a liver. Besides spatial concepts, the types of shared semantic attributes which might describe common structure properties are such as parts of a structure (e.g. has four chambers, two superior atria and two inferior ventricles which can be used to recognise this structure is a heart), common materials (e.g. material with positive mass which can be used to recognise anatomical structures with positive mass, such as liver and brain), and common immaterial (e.g. material with no mass which can be used to recognise anatomical structures with no mass such as the cavity of the stomach).

In the context of finding a matching image, two images match when a region matching scheme that integrates the semantic properties of all the regions in one image matches the semantic properties of all the regions in the other image. Similarly, in the context of finding a matching region, two regions match when the semantic properties of a region match with the semantic properties of the other region.

Since the gene expression data in biomedical atlases are queried by painting or drawing arbitrary regions in the anatomical images, this research focuses on image region mappings. Therefore, the case of mapping to find a matching image is not further explored. In general, there is no literature that provides the definition of criteria to determine if two regions match biologically. Therefore, in the case of region mappings between non-identical images, we proposed to use the mappings from biologists as the golden standard.

III.   OVERVIEW OF MAPPING SOLUTIONS

In general, image mappings can be carried out using image processing and ontology-based methods. In particular, this research focuses on the ontology-based mapping via the use of spatial relations between anatomical regions and the image processing-based mapping via the use of fiducial points.

A. Image Processing Algorithm

There are several examples of image processing algorithms, which perform mapping based on fiducial points. These are the work as proposed by Cai et al. [15], Yu and Morel [23], Park et al. [24], Wong and Orchard [25], You et al. [26], and Zeng et al. [27]. There is also one exemplary work of image processing algorithm, which makes use of semantic concepts. Liu et al. [28] presented an algorithm, which segments an image into

different regions. Each region is extracted together with its low-level features. These features are linked to the semantic concepts obtained in a proposed decision tree-based learning algorithm. The matching algorithm combines both query by keyword and query by region of interest. However, this algorithm does not include the notion of fiducial points, which makes the method irrelevant to the scope of this research.

An image processing algorithm may have difficulty to cope with the mapping of regions between similar images but which are not identical in their morphologies. Differences in morphologies and regions between two images can cause different pixel intensity distributions, which the image processing technique is unable to cope with. An exemplary case is when the mapping is between image slices of an embryo. This also applies to non-identical images that are not identical in their morphologies. Similarly, an image processing algorithm may have difficulty to cope with the mapping of regions between non-identical images with same morphology. The morphology of an animal can be represented in different set of pixel distributions. This is the case when the visual content of both images is only similar at the higher scene level, but entirely different at the pixel-level.

Nevertheless, in the case of mapping a region from one image onto another where the two images have exactly the same morphologies and these morphologies have exactly the same pixel intensity distribution, the image processing algorithm may provide image region mappings with good precision.

B. Ontologies

The concepts of spatial relations have been well employed in ontologies by both FMA (Foundational Model of Anatomy) [29] and OBO (Open Biomedical Ontologies) [30] to describe anatomical space in a biomedical domain. Other commonly used ontologies include RadLex and SNOMED-CT. In general, spatial relations between anatomical entities are described using relationships from the following categories: Mereological Relations, Topological Relations and Location Relations. Mereological relations describe the concept of parthood between the whole and its parts, for example, finger is part of hand, hand is part of the arm and so forth. Topological relations describe the concept of connectedness among entities, for example, two entities are externally connected if the distance between them is zero and do not overlap; one example is in human major parts of the joint, where the synovial cavity is externally connected to the synovial membrane [30]. Location relations describe the concept of relative location between entities that may coincide wholly or partially without being part of one another, for example, the brain is located in (but not part of) cranial cavity. A full exposition of spatial relations for biomedical domains can be found in [31, 32].

In the case of mapping image regions using the same ontology, this method may provide a mapping with good precision if the regions being mapped are segmented in both images and the ontology has a definition for this particular region. This applies to the images with exactly the same morphology as well as the images that are morphologically different.

In the case of mapping image regions using different ontologies, this method may provide a mapping with good precision if the regions being mapped are segmented in both images, these regions have their definitions in the ontologies, and the term associated with each region can be mapped from one ontology to another one. This applies to the images with exactly the same morphology as well as the images that are morphologically different.

In general, an ontology-based method may provide an image region mapping low in precision if the region to be mapped is visible in one image but not in the other. This applies in both cases either using the exact same ontology or using different, but aligned ontologies.

IV. DISCUSSION

This research proposes to develop a new mapping technique by addressing many of the limitations of existing solutions. For example, the ontology-based method requires the image to be painted according to anatomical regions. The new proposed method should not require the image to be painted in order to perform the mapping. Hence, mappings should work regardless of one region is segmented in one image and not in the other, or when one term is defined in one ontology but not in the other.

In addition, the technique should be able to perform mapping between images that are morphologically different. Hence, the mappings should work regardless of having varying orientation and position of the structure nor having a different number of regions in the two images. These are the issues with existing solutions, which should be considered in the development of a new mapping technique. This technique will be useful for integrating image-based data towards the integration of biomedical atlases. This technique is expected to provide an alternative to the image processing-based solution that may fail when the images have different underlying morphologies, or when the mapping is between images acquired in different modalities, as that image processing is not capable with these. Finally, this technique is expected to overcome the drawbacks of an ontology-based solution that may not be available for images without painted domains, or when there are no matching ontologies. Although we can argue that a biologist can paint an image according to its anatomical regions, this is a hard task for biologists and is expensive to acquire.

V. CONCLUSION

Image-based data integration is useful to facilitate the sharing of biomedical data across biomedical atlases. Existing image mapping techniques have many limitations. Ontology-based solutions often lack spatial precision and do not work when the images are not annotated with the ontological concepts. Image processing-based solutions may fail when the images have different underlying morphologies nor when mapping involved images with the same morphology but are taken from different imaging modalities. This research suggests that a new technique should be developed to overcome the shortcomings of existing solutions.

REFERENCES

[1] EMAGE. http://www.emouseatlas.org/emage/home.php. Accessed: 15/09/2014

[2] Allen Developing Mouse Brain Reference Atlas: Technical Report. (2010). Allen Developing Mouse Brain Reference Atlas

[3] Gensat. Gensat Brain Atlas of Gene Expression. GENSAT Brain Atlas of Gene Expression in EGFP Transgenic Mice. http://www.gensat.org/index.html. Accessed: 15/09/2014

[4] S. Andrew and K. McLeod, Gene Co-Expression in Mouse Embryo Tissues, 1st CUBIST Workshop, 2011, vol. 753, pp. 1-10.

[5] Allen Brain Atlas: Developing Mouse Brain. http://developingmouse.brain-map.org. Accessed: 15/09/2014

[6] J.H. Christiansen, Y. Yang, S. Venkataraman, L. Richardson, P. Stevenson, N. Burton, R.A. Baldock, D.R. Davidson, "Emage: a spatial database of gene expression patterns during mouse embryo development," in Nucleic Acids Research, vol. 34, pp. 637-641, 2010.

[7] R.A. Baldock, J.B. Bard, A. Burger, N. Burton, J. Christiansen, G. Feng, B. Hill, D. Houghton, M. Kaufman, J. Rao, J. Sharpe, A. Ross, P. Stevenson, S. Venkataraman, A. Waterhouse, Y. Yang, D.R. Davidson, "Emap and emage - a framework for understanding spatially organized data," in Neuroinformatics, vol. 4, pp. 309-325, 2003.

[8] K. McLeod and A. Burger, "Towards the use of argumentation in bioinformatics: a gene expression case study," in Bioinformatics, vol. 24, pp. 304-312, 2008.

[9] K. McLeod, G. Ferguson and A. Burger, "Argudas: lessons for argumentation in biology based on a gene expression use case," in BMC Bioinformatics, vol. 13, pp. S8, 2012.

[10] A. Burger and J. Bard, "Xspan- a cross-spcies anatomy network," in Anatomy Ontologies for Bioinformatics: Principles and Practise, vol. 1, A. Burger, D. Davidson, and R. Baldock, Eds. Springer, London, 2008, pp. 163-175.

[11] H. Nie, R.P.M.A. Crooijmans, A. Lammers, E.M.V. Schothorst, J. Keijer, P.B.T Neerincx, J.A.M Leunissen, H.J. Megens, and M.A.M Groenen, "Gene expression in chicken reveals correlation with structural genomic features and conserved patterns of transcription in the terrestrial vertebrates," in PLoS One, vol. 5, no.8, 2010.

[12] W.Z. Wang, F.M. Oeschger, J.F. Montiel, F. Garcia-Moreno, A. Hoerder-Suabedissen, L. Krubitzer, C.J. Ek, N.R. Saunders, K. Reim, A. Villalon, and Z. Molnar, "Comparative aspects of subplate zone studied with gene expression in sauropsids and mammals," in Cereb Cortex, 2011.

[13] P. Coumailleau and D. Duprez, "Sim1 and sim2 expression during chick and mouse limb development", in International Journal of Developmental Biology, vol. 53, pp. 149-157, 2009.

[14] The University of Bath Press Release, 2008. Chicks to give scientists clearer picture of fetal development http://www.bath.ac.uk/news/2008/11/14/chick-atlas.html. Accessed: 28/06/2013

[15] J. Cai, Z.J. Zha, Y. Zhao, and , Z. Wang, "Evaluation of histogram based interest point detector in web image classification and search," in ICME, pp. 613-618, 2010.

[16] R. Baldock. EMAP Technical Report: A 3D Paint Program for the Mouse Atlas and Gene Expression Database. MRC Human Genetic Unit, 2004.

[17] W. Jia, H. Zhang, X. He, and Q. Wu. A comparison on histogram based image matching methods, IEEE International Conference on Video and Signal Based Surveillance, AVSS '06, 2006, pp. 97-97.

[18] A. Hanbury and B. Marcotegui, Colour adjacency histograms for image Matching, Proceedings of the 12th International Conference on Computer Analysis of Images and Patterns, CAIP'07, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 424-431.

[19] W. K. Leow and S. Y. Lai, Invariant matching of texture for content-based image retrieval, Proc. Workshop on Texture Analysis in Machine Vision, 1999, pp. 119-125.

[20] W. K. Leow and S. Y. Lai, "Invariant matching of texture for content-based image retrieval," in Texture Analysis in Machine Vision, World Scientific, pp. 151-163, 2000.

[21] J. Z. Wang, J. Li, and G. Wiederholdy, "Simplicity: Semantics-sensitive integrated matching for picture libraries," in Advances in Visual Information Systems, Lecture Notes in Computer Science, vol. 1929, R. Laurini, Eds. Springer Berlin Heidelberg, 2000, pp. 360-371.

[22] S. Todorovic and N. Ahuja, "Region-based hierarchical image matching," in International Journal of Computer Vision, vol. 78, no. 1, pp.47-66, 2008.

[23] Y. Guoshen and J. M. Morel, "Asift: An algorithm for fully affine invariant comparison," in Image Processing On Line, 2011.

[24] H. Park, G. R. Martin, and A. Bhalerao, "Local affine image matching and synthesis based on structural patterns," in Trans. Img. Proc., vol. 19, no. 8, pp. 1968-1977, 2010.

[25] A. Wong and J. Orchard. Efficient fft-accelerated approach to invariant optical-lidar registration, IEEE T. Geoscience and Remote Sensing, 2008, vol. 46, no. 11, pp. 3917-3925.

[26] J. You, E. Pissaloux, T. Dahlinm, and W. Zhu, "The parallel detection of interesting points in distance transform for image matching," in MVA, pp. 151-154, 1994.

[27] Y. Zeng, D. Samaras, W. Chen, and Q. Peng,"Topology cuts: A novel min-cut/max-ow algorithm for topology preserving segmentation in n-d images," in Comput. Vis. Image Underst., vol. 112, no. 1, pp. 81-90, 2008.

[28] Y. Liu, G. Zhang, and D. Lu, "Region-based image retrieval with high-level semantics using decision tree learning," in Pattern Recogn., vol. 41, no.8, pp. 2554-2570, 2008.

[29] C. Rosse, and J.L.V. Mejino, "The foundational model of anatomy ontology," in Anatomy Ontologies for Bioinformatics: Principles and Practise, vol. 1, A. Burger, D. Davidson, and R. Baldock, Eds. Springer, London, 2008, pp. 59-117.

[30] T. Bittner, M. Donelly, L.J. Goldberg, and F. Neuhaus, "Modeling principles and methodologies – spatial representation and reasoning," in Anatomy Ontologies for Bioinformatics: Principles and Practise, vol. 1, A. Burger, D. Davidson, and R. Baldock, Eds. Springer, London, 2008, pp. 307-326.

[31] N.J.M Zaizi and D.N.A. Iskandar, "Using image mapping towards biomedical and biological data sharing," in GigaScience, vol. 2, no. 12, 2013.

[32] W. M. Dahdul, H. Cui, P. M. Mabee, C. J. Mungal, D. Osumi-Sutherland, R. L. Walls and M. A. Haendel, "Nose to tail, roots to shoots: spatial descriptors for phenotypic diversity in the Biological Spatial Ontology," in Journal of Biomedical Semantics, vol. 5, 2014.