Unsupervised Content Classification Based Nonrigid Registration of Differently Stained Histology Images

Y. Song, D. Treanor, A. J. Bulpitt, N. Wijayathunga, N. Roberts, R. Wilcox, and D. R. Magee*

***Abstract*—Registration of histopathology images of consecutive tissue sections stained with different histochemical or immunohistochemical stains is an important step in a number of application areas, such as the investigation of the pathology of a disease, validation of MRI sequences against tissue images, multiscale physical modeling, etc. In each case, information from each stain needs to be spatially aligned and combined to ascertain physical or functional properties of the tissue. However, in addition to the gigabyte-size images and nonrigid distortions present in the tissue, a major challenge for registering differently stained histology image pairs is the dissimilar structural appearance due to different stains highlighting different substances in tissues. In this paper, we address this challenge by developing an unsupervised content classification method that generates multichannel probability images from a roughly aligned image pair. Each channel corresponds to one automatically identified content class. The probability images enhance the structural similarity between image pairs. By integrating the classification method into a multiresolution-block-matching-based nonrigid registration scheme (N. Roberts, D. Magee, Y. Song, K. Brabazon, M. Shires, D. Crellin, N. Orsi, P. Quirke, and D. Treanor, "Toward routine use of 3D histopathology as a research tool," *Amer. J. Pathology*, vol. 180, no. 5, 2012.), we improve the performance of registering multistained histology images. Evaluation was conducted on 77 histological image pairs taken from three liver specimens and one intervertebral disc specimen. In total, six types of histochemical stains were tested. We evaluated our method against the same registration method implemented without applying the classification algorithm (intensity-based registration) and the state-of-the-art mutual information based registration. Superior results are obtained with the proposed method.**

***Index Terms*—Computerized diagnosis, digital pathology, histology registration, image analysis, multistains, mutual information, virtual slides.**

Manuscript received April 9, 2013; revised June 14, 2013; accepted August 5, 2013. Date of publication August 8, 2013; date of current version December 16, 2013. This work was supported by WELMEC, Center of Excellence in Medical Engineering funded by Wellcome Trust and EPSRC under Grant WT088908/Z/09/Z. *Asterisk indicates corresponding author.*

Y. Song was with the School of Computing, University of Leeds, Leeds, LS2 9JT, U.K. He is now with the Centre for Medical Image Computing, University College London, London, WC1E 6BT, U.K. (e-mail: yi.song@ ucl.ac.uk).

D. Treanor is with the Leeds Institute of Molecular Medicine, University of Leeds, Leeds, LS2 9JT, U.K., and also with Leeds Teaching Hospitals NHS Trust, Leeds, LS9 7TF, U.K. (e-mail: darrentreanor@nhs.net).

A. J. Bulpitt is with the School of Computing, University of Leeds, Leeds, LS2 9JT, U.K. (e-mail: a.j.bulpitt@leeds.ac.uk).

N. Wijayathunga and R. Wilcox are with the School of Mechanical Engineering, University of Leeds, Leeds, LS2 9JT, U.K. (e-mail: V.N.Wijayathunga@leeds.ac.uk; R.K.Wilcox@leeds.ac.uk).

N. Roberts is with the Leeds Institute of Molecular Medicine, University of Leeds, Leeds, LS2 9JT, U.K. (e-mail: N.J.Roberts@leeds.ac.uk).

*D. R. Magee is with the School of Computing, University of Leeds, Leeds, LS2 9JT, U.K. (e-mail: d.r.magee@leeds.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2013.2277777

I. Introduction

THE histological examination of adjacent tissue sections prepared with different stains or biomarkers (e.g., histochemical, immunohistochemical, and special stains) can provide valuable information to aid understanding of the physical or functional properties of tissue [1]. For example, liver cirrhosis is characterized by dense fibrous scars traversing the tissue and dividing it into nodules. The hematoxylin and eosin (H&E) stain (the most common histological stain) enables cell nuclei and cytoplasm to be identified. The Sirius Red, reticulin, and Masson trichrome stains highlight the fibrous scarring by staining different histochemical components of the scar tissue (Collagen subtypes). Combining multiple stain modalities allows a complete picture of the relationship between the causative liver disease and the resulting scarring to be built up. In addition to its utility in medical diagnosis, histological examination of differently stained tissue sections has seen increasing usage in biomedical engineering [2]. In this domain, one of the objectives of our work is to help build multiscale physics based models of the intervertebral disc. The Alcian blue stain can characterize proteoglycan-rich extra cellular matrix as a blue color—a component necessary for maintaining spinal flexibility under different loading scenarios, whereas the Elastic Picro Sirius Red (EPSR) stain shows collagen-rich structural areas within and at the periphery of the disc. Combining this information allows an improved understanding of the biomechanical behavior of the disc, leading to better numerical evaluation of treatment methodologies [3].

Traditionally, pathologists and researchers view glass microscopy slides serially with a microscope and combine the information mentally to derive an opinion or diagnosis. Although this works for global assessment of tissue sections, detailed assessment and measurement requires more detailed side-by-side comparison. However, due to the nature of the slide preparation and the fact that different stains characterize different substances, the tissue sections do not have the same morphology, appearance, or spatial alignment, making it a nontrivial task to even find the same region on adjacent slides. The introduction of digital pathology (scanning tissue sections and digitizing them into images) allows the development of automatic computer-aided registration algorithms [4]–[10] to assist pathologists and researchers quantitatively analyzing the spatial co-occurrence of structural and functional elements in different modalities.

Our study on multistain histology image registration is motivated by the research demands of both medical research (liver disease) and biomedical engineering (intervertebral disc) applications. The common interest is to spatially align differently stained adjacent tissue sections for information integration. The

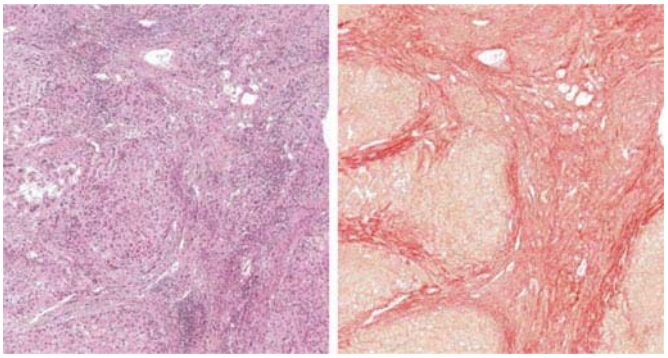

Fig. 1. Dissimilar appearance between differently stained adjacent sections. Slides are viewed at 1.25× magnification. Adjacent liver sections are stained with H&E (left) and Sirius Red (right).

experiments in this paper cover various stain types including H&E, Sirius Red, Alcian blue, EPSR, CK7, and FAST. Although discussions are mainly based on liver and intervertebral disc specimens, the proposed method makes no assumption of the existence of particular anatomic structures and could be applied to any similarly produced histological sections. The main challenges for registering histology images with different stains are:

1) Dissimilar appearances: Different stains highlight different substances in tissues. Consequently, the same tissue structures on adjacent tissue sections often appear very different. For example, the H&E stain allows cell nuclei and cytoplasm to be identified at high magnifications, e.g., 20×. However, as shown in Fig. 1 left, the content at low magnification (1.25×) appears as a rather uniform texture of cellular components. On the other hand, at the same magnification, the fibrous tissue (e.g., scarring of liver tissue) in an adjacent section is clearly selectively highlighted by the Sirius Red stain (see Fig. 1 right).
2) Nonrigid distortions: Introduced by slice preparation, such as bending, shearing, stretching and tearing, etc. At micron resolutions, even minor deformations appear conspicuous.
3) The large size of digital microscopy images. For example, one uncompressed three-channel, 8-bit digital microscopy image of an intervertebral disc at 40× magnification used in our study has file size ∼20 GB.
4) The image pair is not the exact same piece of tissue. They are usually at best serial sections 5μm apart.

In this paper, we address these challenges with:

1) A multiresolution-block-matching-based nonrigid registration scheme, which aims to tackle the large image size and nonrigid distortion.
2) A novel two-dimensional (2-D) unsupervised content classification method (the main contribution of this paper, detailed discussion in Section II). The objective of this method is to enhance the structural similarity between the differently stained image pairs. The output from the algorithm is multichannel probability images (the default is three-channel in this paper). Each channel corresponds to an emergent appearance (content) class that may or may not map to a single anatomical tissue type [see Fig. 2(c)–(e)].

Our multiresolution-block-matching-based registration scheme has been presented in our previous work on registering consecutive histological sections stained with the same stain [9]. For the multistained case presented in this paper, we extend this scheme to register the multichannel probability images obtained from the unsupervised content classification method presented in this paper (see Figs. 3 and 4).

A. Related Work

There are a limited number of previous studies related to the registration of microscopy data with different stains that largely concentrate on specific feature matching. For example, Can *et al.* [4] rigidly registered bright field and fluorescent microscopy images by extracting a nuclear image from each modality, and Cooper *et al*. [5] extracted features based on color that are matched and combined by fitting a polynomial transform. Earlier, Cooper *et al*. [6] presented a more general framework for nonrigid registration that involved detecting anatomic features based on high local pixel variance and using normalized cross-correlation to match these at specific locations within the image. However, for some tissue specimens, anatomic features such as blood vessels, etc., are distributed sparsely across the image, giving rise to insufficient information for calculating a precise deformation field for nonrigid registration. There are also intensity based approaches. The mutual information maximization based method, which is widely adopted for registration of multimodal radiology images [18], has been extended to register differently stained histology images where they exhibit a certain structural similarity [7], [15]. However, such similarity is not necessarily guaranteed for some differently stained pairs, as exemplified in Fig. 1. Therefore, in this paper, we propose an unsupervised content classification method to enhance the structural similarity between image pairs that consequently improves the registration accuracy.

In addition to the proposed classification method, our current registration framework builds on two other categories of related work: multiresolution registration and block matching based nonrigid registration. Multiresolution schemes have been widely adopted in image registration to increase speed and avoid local minima and thus increase the likelihood of convergence to the correct solution [13]–[15]. This is part of the reasoning for our choice of this approach; however, more importantly, the multiresolution approach ensures we start with roughly aligned subimages at each scale, which is essential in our unsupervised content classification method as well as in regularizing the set of displacement vectors. Block matching methods have been proposed in the context of radiology image registration [16], [17], which search reference images to find the best block matching to floating image blocks. The deformation field is then represented by the set of block centroid displacements. While there is similarity between these approaches and our work, it is of note that the block sizes used in our work are substantially larger[1]

[1]From our cross-validation study, the best results are given at 256 × 256 pixels per block.

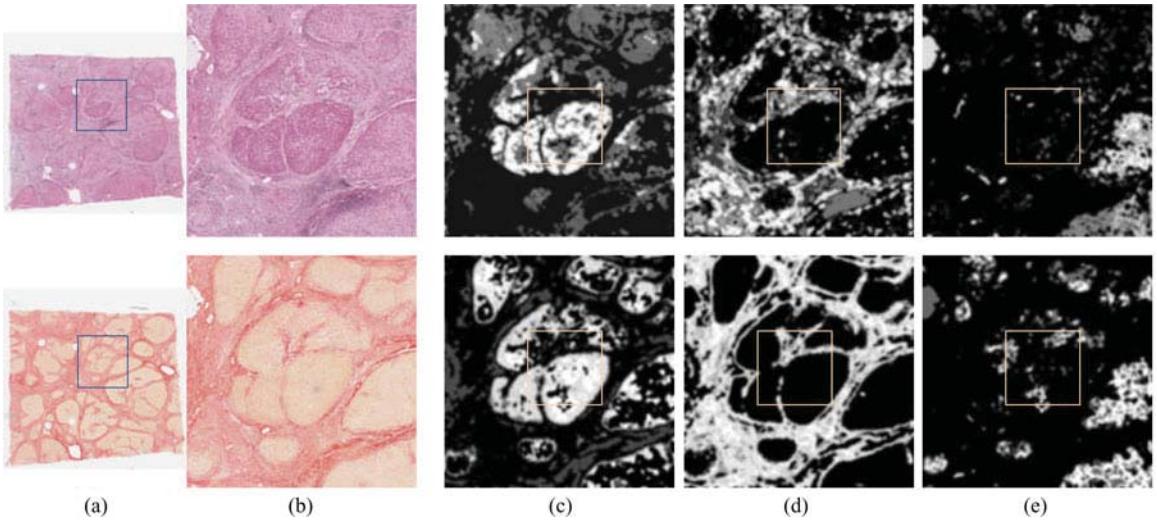

Fig. 2. (a) Images of two adjacent liver tissue sections stained by H&E (reference image) and Sirius Red (floating image) separately. The pair has been aligned using the transformation estimated at resolution 1.25×. (b) The subimages marked in (a) at resolution 2.5×. (c–e) Three channel probability images obtained by applying the 2-D unsupervised tissue classification method on the subimages. Each channel corresponds to one appearance (content) class emerging from the subimages.

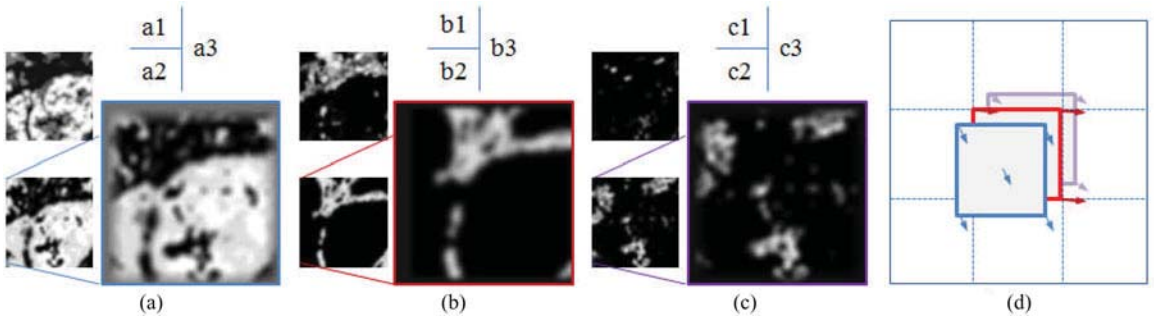

Fig. 3. Registration of multichannel probability blocks by calculating displacement vectors. We use the central blocks highlighted in Fig. 2(c)–(e) to demonstrate. (a1–c1) are reference blocks, (a2–c2) are floating blocks, and (a3–c3) are the transformed floating blocks (images are enlarged for clarity). (d) The registration of the floating block is represented by $5 \times n$ vectors (i.e., displacement vectors) for a n channel probability image ($n = 3$ in this figure).

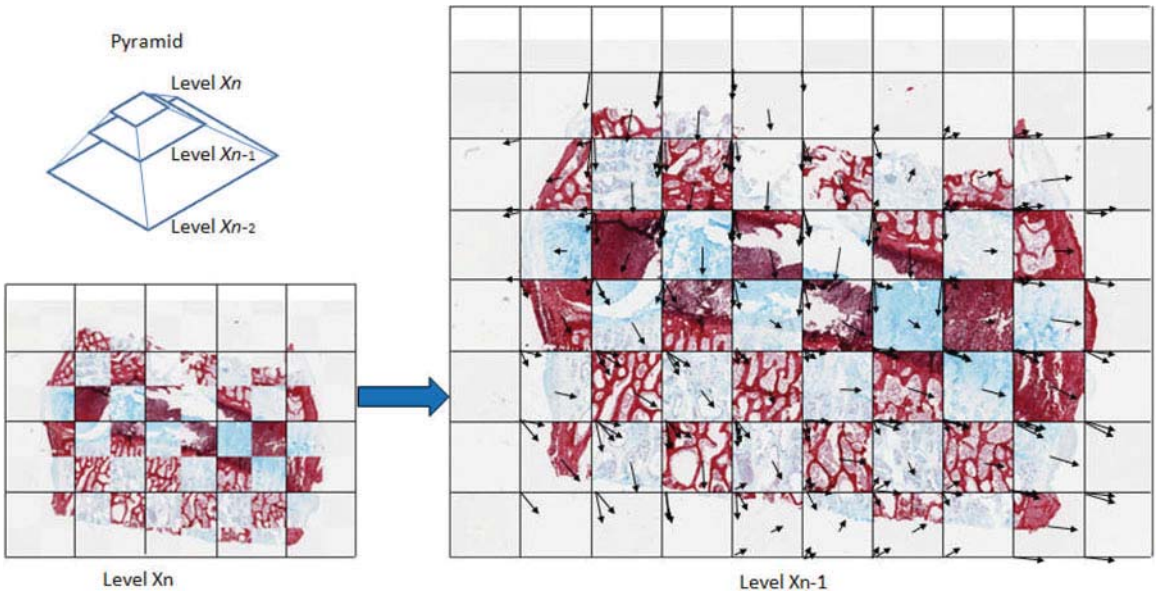

Fig. 4. Illustration of our multiresolution-block-matching-based nonrigid registration scheme applied to intervertebral sections stained by Alcian blue and EPSR. The transformation estimated at level X_n is refined at level X_{n-1}. The reference and floating images are automatically partitioned into blocks by regular lattices of constant size (256×256 pixels). For clarity, the lattice is not visualized with 50% overlapping as in our implementation and the set of displacement vectors (exemplified at level X_{n-1}) is plotted based on only one channel.

than that often used in block matching in the literature. The advantage of this is that a greater amount of detail can be used in the registration process, increasing robustness. The disadvantage of using larger blocks is that a sparser set of displacement vectors is produced. However, our use of five vectors per block (thus taking into account block rotation; see Fig. 3), 50% block overlapping, and B-spline approximation alleviate this to an extent.

There is a rich literature on employing unsupervised classification algorithms (e.g., clustering) for applications such as gene expression analysis [23], data mining [24], information retrieval [25], image segmentation [26], etc. However, unlike previous works that process images individually, our method processes image pairs simultaneously to capture the statistically significant structures shared by both images. This approach has not been seen in the literature to the best of our knowledge.

The rest of this paper is organized as follows: An overview of our registration implementation integrated with the 2-D unsupervised content classification method is presented in Section II. In Section III, we present the classification algorithm in five steps: (A) appearance feature selection and generating feature vector sets; (B) clustering the appearance feature vector set of each image; (C) partitioning the obtained two sets of clusters into content classes common to both images; (D) iteratively refining the emergent content classes by introducing spatial features; (E) generating probability images based on the refined content classes. The datasets and experimental results are described in Section IV. Conclusions are made in Section V.

II. Overview on Registration Implementation

Our registration implementation is based on a rigid block matching approach that we have previously presented for the same stain registration [9]. The novel extension here is generating multiple channel probability images from images of different stains using an unsupervised content classification method to increase the similarity of the two images (see Section III). The sequence of processes is as follows.

1) Transform each image into a multichannel probability image, where the individual channels of each image correspond (see Section III).
2) Perform block matching based registration for each probability channel separately (see Fig. 3(a)–(c), the rigid transform of the block is estimated by using phase correlation, a combination of [11] and [12] to recover rotation and translation).
3) Represent the rigid transform of each block by five vectors [located at the four corners and center of the block, Fig 3(d)]. The results from different channels are combined into a displacement vector set.
4) Compute a single approximating nonrigid transform (cubic B-spline) from the displacement vector set (see Fig. 4) by using a regularized least squares difference minimizing method (see [9] for details).

The approach is applied in a multiscale manner from coarse to fine. In addition, except for the initialization (which is performed by rigidly registering whole grayscale images at the coarsest resolution), the registration problem is decomposed into numerous small block matching problems. So we bypass the infeasible practice of directly handling gigabyte images.

TABLE I
Notation

Symbol	Descriptions
I_R, I_F	Reference and floating histopathological images.
U_R, U_F	Appearance feature vector sets where each vector represents colour and/or texture information at a pixel in I_R, I_F.
N_R, N_F	The number of clusters associated with I_R and I_F respectively.
I_R^C, I_F^C	Cluster-label image for I_R and I_F, in which each pixel (x,y) is associated with a cluster label.
N_T	Number of content classes (common to both images).
I_R^T, I_F^T	Content-label images, in which pixel (x,y) is associated with a content class label.
S_R, S_F	Spatial feature vector sets representing the coordinates of pixels in a specified cluster in $I^C{}_R$ or $I^C{}_F$.

III. Unsupervised Content Classification

In this section, we present our novel 2-D unsupervised content classification method, which generates multichannel probability images from a roughly aligned image pair. The purpose of this method is to increase the structural similarity of the two images and thus increase registration performance. Each derived probability channel corresponds to an emergent "content class" [see Fig. 2(c)–(e)], where a pixel value in a specified channel image is the probability that it belongs to the corresponding content class. *It should be noted that the automatically identified content classes do not necessarily map to real anatomical tissue classes as they are solely emergent from co-occurrence statistics between image pairs*. The key idea is that we capture the statistically significant structures common in images, by studying two images simultaneously.

Without loss of generality, we explain the proposed classification method based on images at level X_i (which have previously been registered at lower resolution level X_{i+1}) in the following context. Since it is impracticable to directly classify a whole image when the images are very large, we divide the image pair into a set of subimages and apply the proposed content classification algorithm on each subimage pair separately. This also has the advantage of being more robust to color variations over the images. The multiresolution registration scheme ensures similar tissue regions are present in the corresponding subimage pairs. The size of subimages is balanced between being big enough to be representative, but small enough to allow for stain variation. We have used subimages sized 1024×1024 pixels in our experiments. Notation used in the following discussion is listed in Table I. The five-step unsupervised classification algorithm is outlined in algorithm I. In step D of the algorithm, spatial location features are introduced as complementary to the appearance features. This overcomes the problem where the appearance features alone (for some stains) are not sufficient to separate certain regions on an image that have similar appearance, but belong to different content classes.

We detail the implementation of the algorithm step by step in the following Sections III-A–E.

Algorithm I. Unsupervised content classification algorithm

Input: RGB colour image pair, i.e. reference image I_R, floating image I_F.
Output: two multi-channel probability images.
A. Create appearance feature vector per pixel $u_R(x,y)$, $u_F(x,y)$, which form feature vector sets $U_R=\{\boldsymbol{u}_R(x,y)\}$, $U_F=\{\boldsymbol{u}_F(x,y)\}$
B. Cluster feature vector sets U_R, U_F into N_R, N_F clusters respectively, obtaining cluster-label images I_R^C, I_F^C.
C. Find the best partition functions $\phi^{(n)}(I_R^C)$, $\psi^{(m)}(I_F^C)$ which partition N_R, N_F clusters into N_T common content classes, by maximising Mutual Information $\boldsymbol{I}\left(\phi^{(n)}(I_R^C);\ \psi^{(m)}(I_F^C)\right)$.
D. Iteratively refine the content classes by introducing spatial features until the convergence condition is met (a threshold on the partition error).
E. Convert the input images I_R and I_F into two N_T-channel probability images.

A. Appearance Feature Vector Sets (U_R and U_F) Generation

Each pixel is represented as a feature vector. These form the feature vector set used to derive content classes based on co-occurrence statistics. To tackle the various stains and tissue types used in our study, we include a step to automatically select appearance features for each type of image from a feature pool consisting of color, intensity and texture information. This feature selection process is performed only once for each type of image pair.

The feature pool is as follows:

1) Each RGB color channel of the original image and Gaussian smoothed images. The Gaussian kernels are: 7×7 ($\sigma = 3$), 15×15($\sigma = 5$).
2) A texture descriptor based on grayscale images defined in (1) and (2). The kernel sizes ($w \times h$) used are: 15×15, 30×30. The grayscale image is obtained by averaging RGB color channels.

The texture representation (2) is a one-dimensional feature that is the root squared sum of the difference operators (1) in directions (1,0), (0,1), (1,1)

$$I_{(w,h,dx,dy)}(x,y) = \frac{1}{(w\times h)} \sum_{\Delta y=-w/2}^{w/2} \sum_{\Delta x=-h/2}^{h/2} \begin{matrix} I(x+\Delta x, y+\Delta y) \\ -I(x+\Delta x+dx, y+\Delta y+\mathrm{d}y) \end{matrix} \tag{1}$$

$$I^*_{(w,h)}(x,y) = \sqrt{I^2_{(w,h,1,0)}(x,y) + I^2_{(w,h,0,1)}(x,y) + I^2_{(w,h,1,1)}(x,y)}. \tag{2}$$

It should be noted the features listed in the pool could be potentially replaced by other more complicated features if these prove insufficient for a particular image type. However, these features have proved sufficient for the images studied in this paper. The rationale for using these simple appearance features in our study is based on two criteria: 1) low computational cost and 2) we are interested in co-occurrence statistics common between the image pairs, rather than trying to extract specified texture information from a single image to identify a particular tissue type, as proposed in [19], which differentiates the cancerous glands from normal glands by knowing that normal glands tend to be strongly oriented in a particular direction and cancerous glands not.

From the feature pool, the best set of features for each type of image pair is selected by maximizing (3) (Section III-C) over all feature subsets. A greedy search method (as exhaustive search is prohibitively expensive) is used to explore subsets of feature combinations. In Table III, we list the selected features.

From this point forward, we assume the appearance features have been determined for the image pair I_R, I_F by this method. For every pixel (x, y) in I_R, a feature vector $\boldsymbol{u}_R(x, y)$ is computed, which forms the feature vector set U_R. Similarly, the feature vector set U_F is computed for I_F.

B. Clustering Appearance Feature Vector Sets

A clustering algorithm is applied to the appearance feature vector sets U_R, U_F to obtain N_R, N_F clusters and the corresponding cluster-label images I_R^C, I_F^C, respectively. We propose a principal eigenvector binary tree clustering algorithm that creates a binary tree structure for a feature vector set and uses the leaf associated index as cluster labels. Take the feature vector set U_R for instance, the root of the tree consists of a single cluster containing all elements of U_R. Principle component analysis (PCA) is applied at each level of the tree, and the elements are partitioned into two children by thresholding their projections onto the first principle component at zero (the mean). This process results in N_R clusters. Each cluster corresponds to a leaf of the tree. The same clustering algorithm is applied to the feature vector set U_F to get N_F clusters.

It should be noted that the main objective of the proposed clustering method is to guarantee the maximal separation of pixels belonging to different content categories even under similar appearance. For example, the intervertebral image stained with Alcian blue, etc., in Fig. 4 has only a very subtle difference between the light blue tissue and the white background. Another aim of developing this clustering method is computational efficiency (provided by the tree structure).

Denoting the labels of the N_R clusters as $\{c_{Ri}\}$ and the labels of the N_F clusters as $\{c_{Fj}\}$, we obtain a cluster-label image I_R^C corresponding to the reference image I_R, where $I_R^C(x,y) = C_R(\boldsymbol{u}_R(x,y)) = c_{Ri}$. $C_R(.)$ is the mapping function output from the principal eigenvector binary tree clustering algorithm for image I_R. Similarly, a cluster-label image I_F^C corresponding to the floating image I_F is obtained, where $I_F^C(x,y) = C_F(\boldsymbol{u}_F(x,y)) = c_{Fj}$.

C. Partitioning of Clusters into Content Classes

The objective of this step is to partition the obtained N_R, N_F clusters from the two images into N_T common groups (N_T "content classes").[2] This is in keeping with our aim to increase the similarity of the representation of the image pair. Denoting the labels of the N_T classes as $\{t_i\}$. Let $\phi^{(n)}$ be a hypothesized partition scheme of cluster-label image I_R^C, which gives content-label image $I_R^T(x,y) = \phi^{(n)}(I_R^C(x,y)) = t_i$; and $\psi^{(m)}$ a hypothesized partition scheme of I_F^C, which gives content-label

[2]In our implementation, the default number of content classes is three. If having $N_T > 3$, some content classes may contain very few pixels and their corresponding channel of the probability image are treated as background and removed in registration.

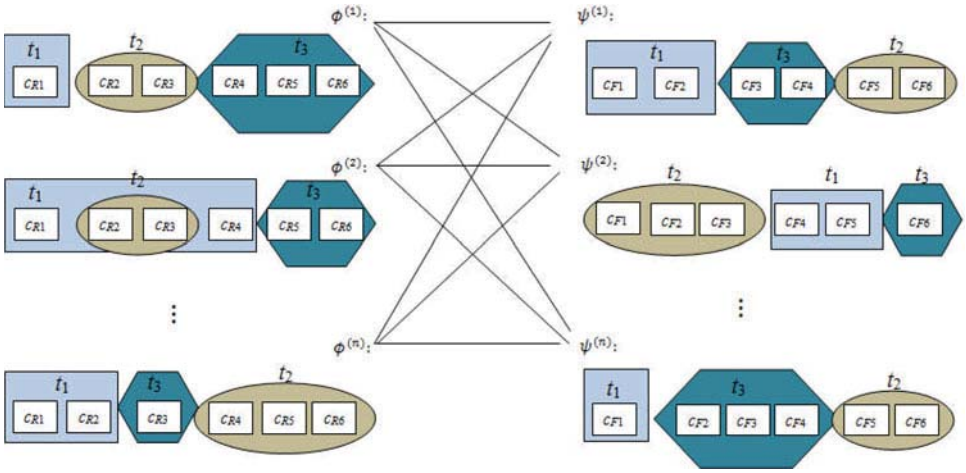

Fig. 5. Process of partitioning clusters: partitioning each set of six clusters into three content classes (t_1, t_2, t_3). Each line between $\phi^{(n)}$ and $\psi^{(m)}$ represents a hypothesized partition scheme, i.e., a combination for which mutual information is calculated.

image $I_F^T(x,y) = \psi^{(m)}(I_F^C(x,y)) = t_j$. The "optimal" content classes are determined by the (nth, mth) hypothesized partitions that maximize mutual information between $\phi^{(n)}(I_R^C(x,y))$ and $\psi^{(m)}(I_F^C(x,y))$, as described by (3). Fig. 5 is a simplified example demonstrating hypothesized schemes of partitioning six clusters into three groups (aka. content classes t_1, t_2, t_3)

$$n, m^* = \arg\max_{n,m}(\mathbf{MI}(\phi^{(n)}(I_R^C); \psi^{(m)}(I_F^C))) \tag{3}$$

As the complexity of an exhaustive search of N_T classes partitions is $O(N_T^{\min(N_R,N_F)})$, which is prohibitively computationally expensive, we approximate the N_T class problem using the principle of alpha-expansion [20] and conduct exhaustive searches of binary assignments for each class. In this search we consider the binary problems of assigning a cluster to a particular class, or it retaining its original assignment. Search is initialized by assigning all clusters of the reference image to a single class, and iteratively repeating the binary searches until convergence. The complexity is therefore reduced to $O\left(N_T * 2^{\min(N_R,N_F)}\right)$.

D. Iteratively Refining Content Classes by Introducing Spatial Features

In this step, we iteratively refine the derived content classes based on spatial features. Due to the fact that the appearance features are sometimes not sufficient to distinguish pixels of different content categories, those pixels could be wrongly clustered into one cluster. This consequently affects the result of the partition, as exemplified in Fig. 6. Comparing the cluster-label images I_R^C and I_F^C, the pixels belonging to two clusters c_{F1} and c_{F3} in I_F^C are clustered into a single cluster c_{R1} in I_R^C, which results in the number of pixels belonging to the emergent content class t_1 in I_R^T and I_F^T being significantly different (enclosed by blue curves). We overcome this by introducing spatial information to enforce the separation of wrongly clustered pixels, e.g., dividing the cluster c_{R1} in I_R^C into several smaller, spatially separate clusters, as shown in Fig. 6 (steps b and c).

We define "conflicting pixels" as pixels labeled as t_i in I_R^T but labeled as t_j $(i \neq j)$ in I_F^T, at the corresponding locations.

The cluster containing the largest number of conflicting pixels in the cluster-label image I_R^C or I_F^C is identified, as described in (4). To distinguish it from other cluster labels, we name it as c^*. For clarity, in the following discussions we use I_R^C as an example

$$c^* = \arg\max_{\{c_{Ri}\}} \left(\sum_{x,y \in I_R^C} \begin{cases} 1, & \text{if } I_R^C(x,y) = c_{Ri} \text{ and } I_R^T(x,y) \neq I_F^T(x,y) \\ 0, & \text{otherwise} \end{cases} \right). \tag{4}$$

Next, for every pixel $I_R^C(x,y) = c^*$, a spatial feature vector $s_{Ri} = [x, y]$ is created, which constitutes a spatial feature vector set S_R. By applying the principal eigenvector - binary tree clustering method (discussed in Section III-B) to S_R, the cluster c^* is split into a set of smaller clusters $\{r_i\}$. Consequently, the cluster-label image I_R^C is updated by replacing label c^* with $\{r_i\}$. Fig. 6(c) demonstrates a synthetic example in which the cluster c_{R1} in I_R^C is split into two new clusters c_{R1-1} and c_{R1-2}. The pixels previously labeled as c_{R1} are replaced by the new labels.

Using the renewed cluster-label image I_R^C, as shown in Fig. 6(d), the content classes are recalculated using (3) (Section III-C). Consequently, we have the refined content-label images I_R^T, I_F^T, as shown in Fig. 6(e). This refinement process is repeated until the number of "conflicting pixels" in any cluster is about 1% of the total number of pixels (typically ten iterations). This threshold is based on the fact that we expect the following errors: 1) the anatomical structures on adjacent tissue sections are similar but not identical and 2) misalignment exists between adjacent tissue sections at level X_i. The resultant partition scheme after the refinement is denoted as ϕ, ψ for I_R^C, I_F^C, respectively. A stopping criterion is necessary as otherwise the eventual result would be two identical content-class images.

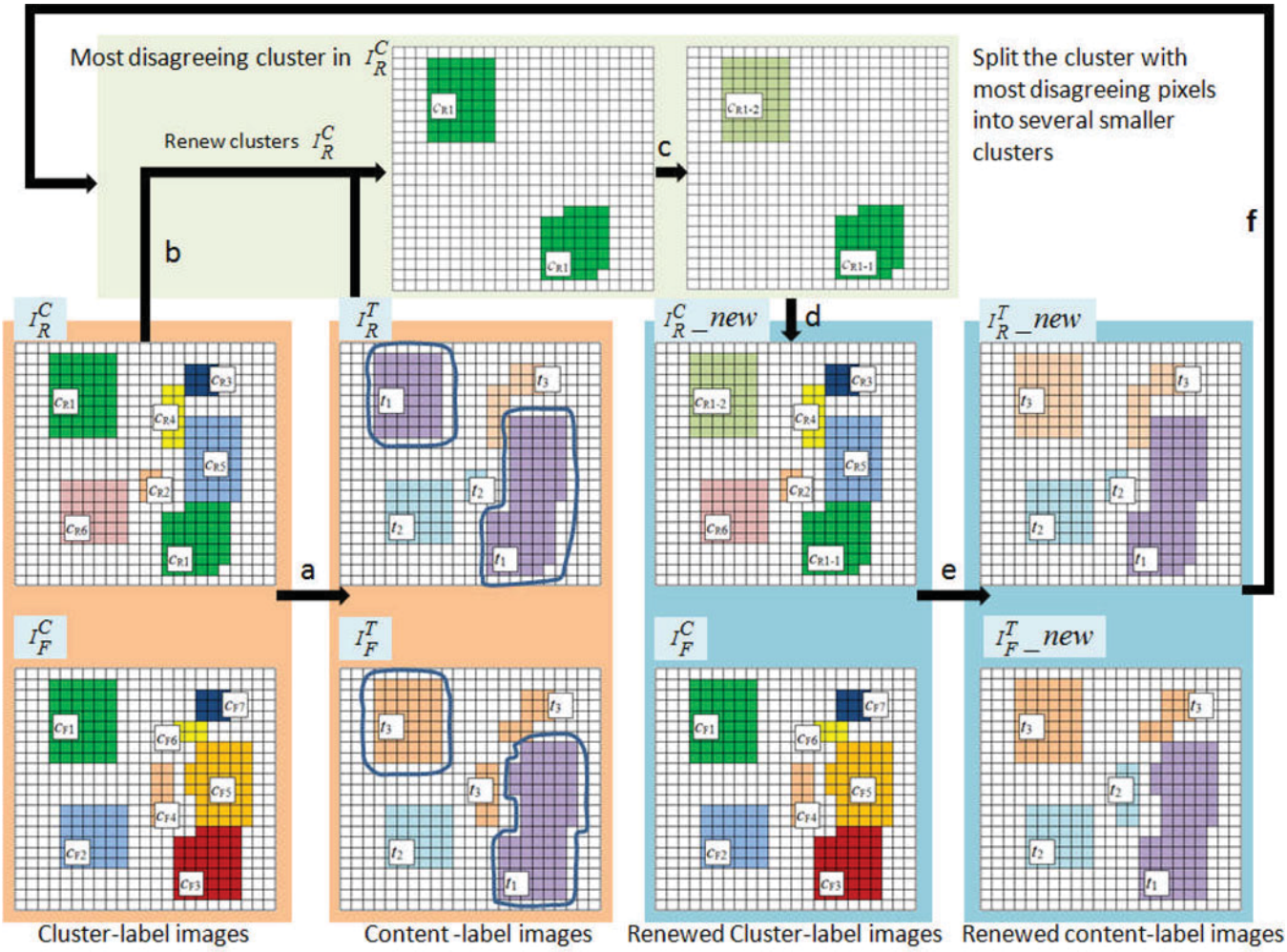

Fig. 6. Synthetic example of partitioning clusters into content classes (see Section III-C) and the following refinement process (see Section III-D). (a) Finding the best partition scheme $\phi^{(n)}, \psi^{(m)}$ that maps cluster-label images I_R^C, I_F^C to content-label images I_R^T, I_F^T by maximizing mutual information, (b) identifying the cluster containing the largest number of conflicting pixels, (c) dividing the identified cluster from (b) into several smaller clusters using spatial information, (d) renewing the cluster-label image I_R^C with the results of (c), and (e) finding the new optimal partition scheme $\phi^{(n\prime)}, \psi^{(m\prime)}$ and renewing content-label images I_R^T, I_F^T.

E. Generation of Probability Images

We convert the refined classification result into probability images that provide a continuous image representation that can be used within our image registration framework. First, we calculate the normalized joint probability density matrix $P\left(I_R^C = a, I_F^C = b\right), a \in \{c_{Ri}\}, b \in \{c_{Fj}\}$. For each emergent content class t_m, we generate a pair of probability images I_{R,t_m}^P, I_{F,t_m}^P for the input image pair, described as follows:

$$I_{R,t_m}^P(x,y) = \sum_{j=1}^{N_F} \begin{cases} P\left(I_R^C, I_F^C\right), & \text{if } \psi\left(I_F^C\right) = t_m \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

$$I_{F,t_m}^P(x,y) = \sum_{i=1}^{N_R} \begin{cases} P\left(I_R^C, I_F^C\right), & \text{if } \phi\left(I_R^C\right) = t_m \\ 0, & \text{otherwise} \end{cases}. \quad (6)$$

Taking Fig. 6 as an example, pixels in the content-label image I_R^T which identified as the t_3 class are assigned the probability value calculated from (5); pixels that do not belong to t_3 classes, are assigned zero probability.

Fig. 2(c)–(e) shows an example of a set of calculated multichannel probability images. The image pair (taken from the liver#1 specimen) is stained by H&E and Sirius red stains, respectively.

IV. Evaluation

A. Materials

77 pairs of tissue sections were taken from three human liver tissue specimens showing cirrhosis and one ovine intervertebral disc specimen (which had been decalcified prior to sectioning). The liver#1 specimen had four artificial vertical holes drilled before it was sectioned. Using standard histological techniques, all specimens were formalin fixed, paraffin embedded, and cut in $\sim$5 μm sections using a microtome. In our study, adjacent sections were stained pairwise, e.g., one of each pair with H&E and other with CK7. In total, six types of histochemical stains were used. The images of these sections were digitized using an Aperio XT scanner at 20$\times$ or 40$\times$ magnification giving a

TABLE II
SPECIMENS

Data group	Specimen	Stain #1	Stain #2	Number of slide (pairs)	Magnification	Image size (pixels)	Full resolution (μm/pixel)
1	Liver #1	Haematoxylin and Eosin (H&E)	Sirius Red	6	20×	~23K×23K	0.504
2	Liver #2	Haematoxylin and Eosin (H&E)	Sirius Red	5	20×	~43K×40K	0.504
3	Liver #3	Haematoxylin and Eosin (H&E)	DAB	6	40×	~45K×27K	0.251
4	Intervertebral disc	Alcian Blue	Elastic Picro Sirius Red (ERSR)	5	40×	~95K×70K	0.246
5	Intervertebral disc	Alcian Blue	FAST	3	40×	95K×70K	0.246
6	Intervertebral disc	FAST	Elastic Picro Sirius Red (EPSR)	3	40×	95K×70K	0.246
7	Liver #1	Haematoxylin and Eosin (H&E)	Sirius Red	49	20×	~23K×23K	0.504

final image resolution of 0.46 or 0.23 μm per pixel. A summary of the datasets is given in Table II. Figs. 7 and Fig. 8 show the examples of the image pairs and their manual annotations. Local NHS research ethical approval was obtained for use of the human tissue. The ovine disc (vertebra specimen) was harvested from a mature (7–8 years old) spine released by the Official Veterinarian of the local abattoir, subsequent to the approval of the application made to the area Veterinary Manager at the Food Standards Agency.

Among 77 image pairs, we include some images with artefacts that are often introduced during the preparation of histopathology images, such as:

1) *Luminance gradient*: Sections mounted close to the edge of the glass slide produce images with significant luminance gradients (e.g., in Fig. 7, liver#3 section stained with CK7).
2) *Nontissue noise*: Dust and air bubbles in the slide (see Fig. 8).
3) *Staining variations*: Differences in section thickness, staining duration, and stain concentration result in color variations (e.g., in Fig. 7, liver#3 section stained with H&E).

B. Evaluation Design

The goal of developing our 2-D content classification method is to improve registration accuracy by using probability images that enhance the structural similarity between image pairs. Therefore, to evaluate the content classification method, we use registration performance as a criterion, which is measured by the Hausdorff distances (μm) between manually annotated corresponding regions (boundaries) on image pairs after registration. The corresponding regions were annotated by the first author and reviewed for approval by a consultant pathologist using the Aperio ImageScope slide-viewing software on the original virtual slide image.

Multiresolution registration (see Section II) is applied with 3 or 4 levels such that image size at the lowest resolution is closest to 1 K$\times$1 K and the "final" level of the registration resolution is closest to 4 μm/pixel. The resulting transformations are then applied to the full-resolution images (and thus to those manual annotations on the images) to calculate the Hausdorff distances (μm) between the corresponding annotations for evaluation.

The size of blocks for matching is a constant 256$\times$256 (pixels) for each scale and blocks are 50% overlapping. This gave the best results in our previous study [9].

By covering a wide range of stain types and two very different types of tissues, our experiments try to reflect the fact that our content classification method makes no assumptions about either anatomic structures or stain types. Moreover, there is no other preprocessing required except conducting automatic appearance features selection once for each pairwise image type. Our evaluation consists of four experiments:

1) The application of automatic appearance feature selection, given in Section IV-C.
2) In Section IV-D, we compare registration performance between using content classification (CC) and not using content classification (NC), i.e., intensity based registration. The data groups 1–6 (28 image pairs in total, each having 20–30 manual annotations, as shown Fig. 7) are used in this evaluation.
3) In Section IV-E, we compare our content classification based registration against the state-of-the-art mutual information based registration [7]. The data groups 1–6 are used in this evaluation.
4) In Section IV-F, we evaluate the influence of sectioning distance between the stained sections on our classification method. Dataset 7 (49 image pairs, with manual annotations of the four artificial holes, as shown Fig. 8) is used in this evaluation.

C. Experiment 1: Feature Selection

As discussed in Section III-A, for each image pair type at each resolution, such as H&E/Sirius Red, H&E/DAB, etc., the best set of features is selected from the feature pool by maximizing (3) over all feature subsets (via greedy search). Feature selection is performed on subimages of size 1024$\times$1024 pixels, the same as the classification process. In our experiments, we tested those subimages located at both the boundary and the center of

Fig. 7. Examples of image pairs and their manual annotations. From top to the bottom, they are sections taken from liver#1 (H&E/Sirius Red), liver#2 (H&E/Sirius Red), liver#3 (CK7/H&E), intervertebra disc (EPSR/Alcian blue), intervertebra disc (Alcian blue/FAST), intervertebra disc (FAST/EPSR). Three liver specimens are viewed at 0.5× magnification. Sections of vertebra discs are viewed at 0.6× magnification. Manual annotations used in the evaluation are superimposed on the images.

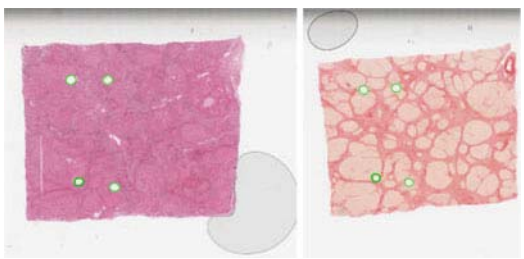

Fig. 8. Data group 7, four artificial vertical holes were manually annotated (in green).

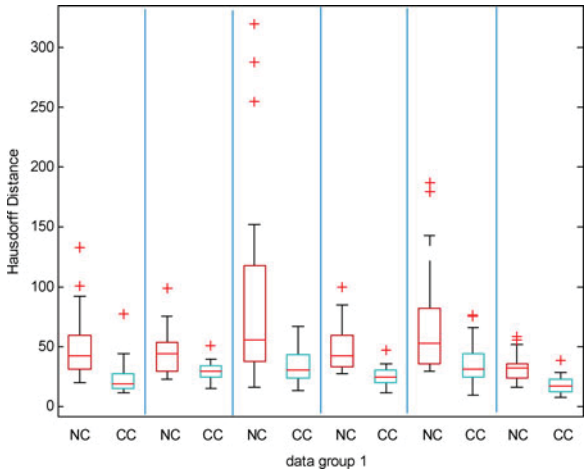

Fig. 9. Boxplot of the registration accuracy of consecutive tissue sections stained by H&E/Sirius Red (data group 1). In total, six image pairs of the liver#1 specimen. The registration performance is compared between registrations using content classification (CC) and not using content classification (NC), i.e., intensity-based registration.

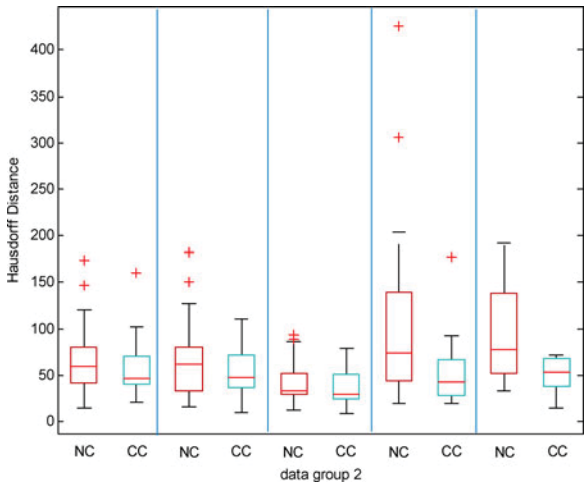

Fig. 10. Boxplot of the registration accuracy of consecutive tissue sections stained by H&E/Sirius Red (data group 2). In total, five image pairs of the liver#2 specimen.

the images with little difference in the final registration performance. Generally, the features selected from the center are more representative, and are adopted in the following experiments.

Taking H&E/Sirius Red as an example, we randomly choose one image pair from the data group 1. Applying the feature selection process on its subimages, we then get a set of features for the image stained by H&E and another set of features for the image stained by Sirius Red (see Table III). The obtained features are used for the content classification for all image pairs of the group 1. To test the robustness of the selected features to staining variations, we applied the same feature sets to images of data group 2, which too are stained by H&E/Sirius Red but present significant color variations from those image pairs in data group 1 (see Fig. 7). The registration results of data groups 1–2 can be found in Figs. 9 and 10, respectively.

TABLE III
SELECTED FEATURES FOR THE FINAL REGISTRATION LEVEL

		Guassian kernel 7×7 (σ=3)			Guassian kernel 15×15 (σ=5)			Texture descriptor	
		B	G	R	B	G	R	w×h 15×15	w×h 30×30
Stain #1	H&E		√	√				√	
Stain #2	Sirius Red		√	√			√	√	
Stain #1	H&E			√					√
Stain #2	DAB	√						√	√
Stain #1	Alcian Blue	√	√					√	√
Stain #2	EPSR	√	√	√	√			√	
Stain #1	Alcian Blue	√	√					√	√
Stain #2	FAST	√	√	√	√			√	√
Stain #1	FAST				√	√		√	√
Stain #2	EPSR	√						√	√

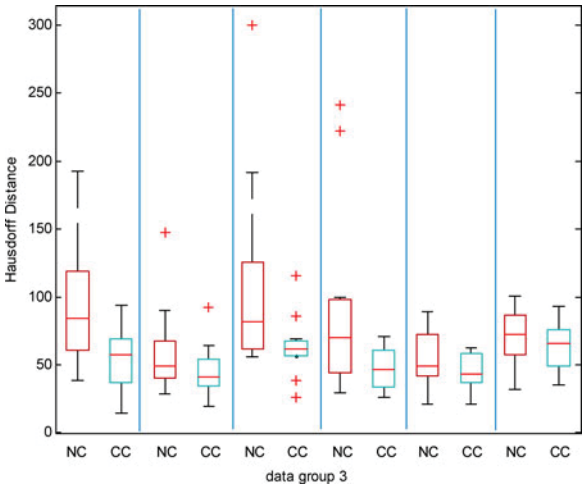

Fig. 11. Boxplot of the registration accuracy of consecutive tissue sections stained by H&E/CK7 (data group 3). In total, seven image pairs of the liver#3 specimen.

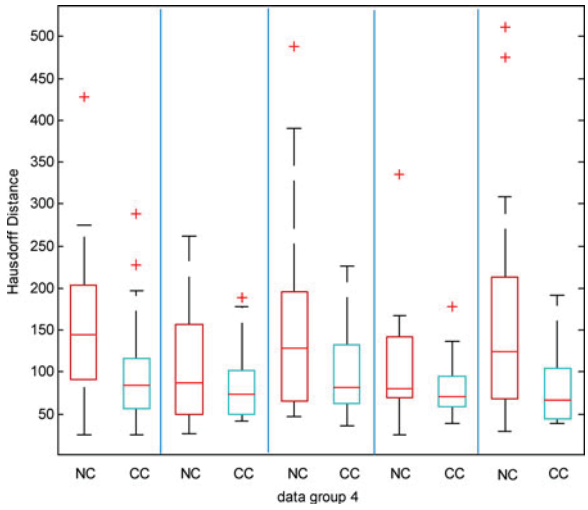

Fig. 12. Boxplot of the registration accuracy of consecutive tissue sections stained by Alcian blue/EPSR (data group 4). In total, five image pairs of the intervertebral disc specimens.

The feature selection process was also applied on other data groups covering various stains: H&E/DAB, Alcian blue /EPSR, Alcian blue /FAST and FAST/EPSR. The selected features sets of the "final" level of registration resolution are exemplified in Table III.

D. Experiment 2: Registration Performance

As mentioned in Section IV-B, we use registration performance as a criterion to evaluate the content classification method. We compared the registration accuracy between using no content classification (NC), and the registration process using the content classification (CC). We report registration results for 28 histology image pairs (data groups 1–6), evaluated by measuring the Hausdorff distances (μm) between the corresponding annotations (transformed by the registration result at full resolution X_0). There are 20–30 corresponding regions (see Fig. 7) manually annotated on each image.

Figs. 9–14 show that there is improvement in registration accuracy using the content classification (CC), when compared to the registration process without using content classification (NC). In Fig. 15 visualizes the boundary smoothness of an anatomical structure (blood vessel), and illustrates that CC based registration can achieve better accuracy than NC based registration. A paired t-test was performed and a statistically significant improvement in error was detected for CC over NC for all data sets, as listed in Table IV.

E. Experiment 3: Content Classification Based Registration Versus Mutual Information-Based Registration

We compared our content classification based registration against to the state-of-the-art mutual information based registration [7]. We replicate the implementation as described in [7], except:

1) Fifty bins are used in Mattes mutual information [21], instead of 32 bins. We have tested both 50 bins and 32 bins on the 28 datasets. The performance using 50 bins is better than, or similar to, that using 32 bins.
2) Instead of using 4096 random samples, we define the number of samples as a fraction (20%) of the total number of pixels in the image, which gives good results for all 28 datasets.

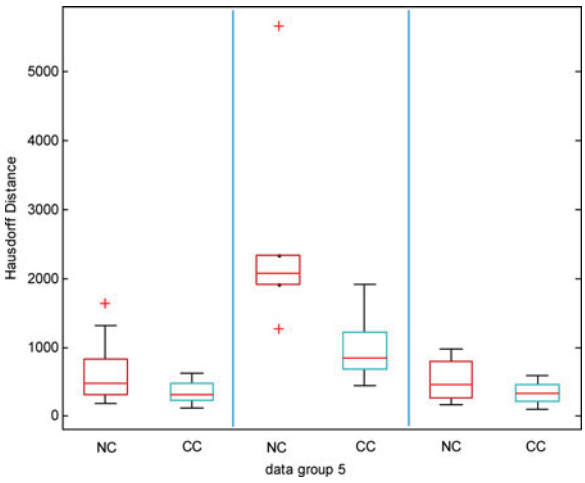

Fig. 13. Boxplot of the registration accuracy of consecutive tissue sections stained by Alcian blue/FAST (data group 5). In total, three image pairs of the intervertebral disc specimens.

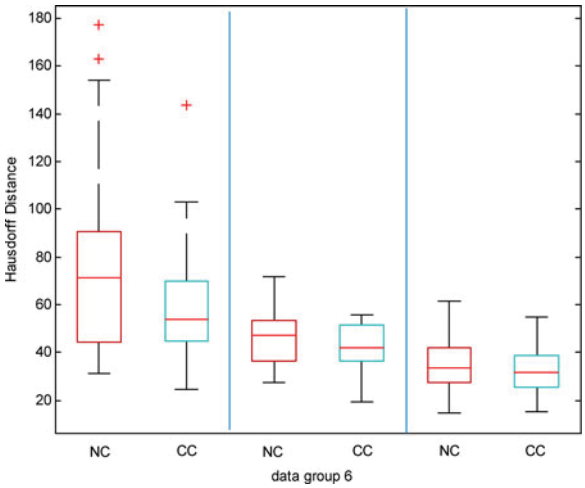

Fig. 14. Boxplot of the registration accuracy of consecutive tissue sections stained by EPSR/FAST (data group 6). In total, three image pairs of the intervertebral disc specimens.

Performance of mutual information based registration is evaluated on data groups 1–6 (28 image pairs), with the same multiresolution levels applied as in Experiment 2. Comparative results are summarized (groupwise) in Fig. 16. These show that our method slightly outperforms MI-based registration.

Computational efficiency is also an important factor. We compare the two methods by execution time, at the specified image sizes and the number of multiresolution levels as listed in Table V. Results are summarized (groupwise) in Fig. 17. These show that MI based registration generally takes a significantly longer time than our method and has larger variance between the different image pairs from the same group. This can be explained by the iterative nature of the MI algorithm, which has a stopping criteria based on convergence. All computations were carried out on a desktop computer with the following specifications: Intel $i7$ dual core $\times 3.07$ GHz, 8 GB RAM, Windows 7. It should be noted that our method is not yet optimized for parallel computation.

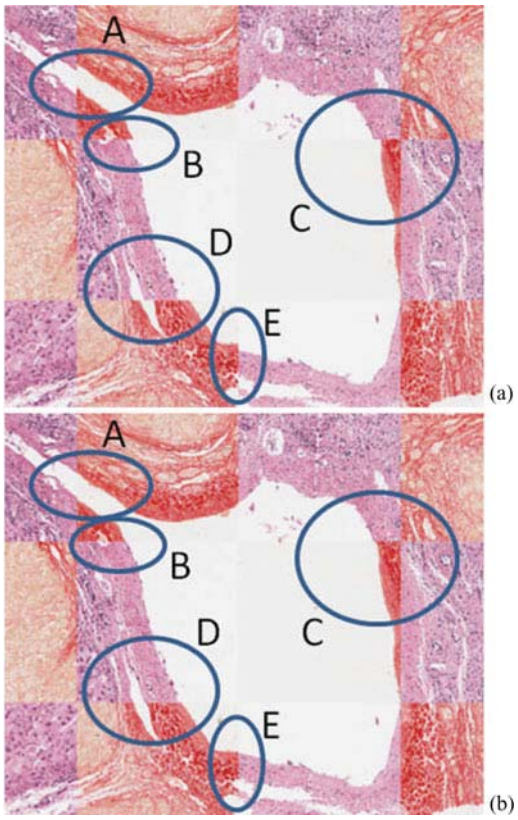

Fig. 15. Checkerboard visualization (at magnification $10\times$) of the registration accuracy (liver images stained by H&E and Sirius Red). (a) Results of registration done at the magnification of $2.5\times$, without applying classification (NC). (b) Results of applying content classification.

TABLE IV
PAIRED t-TEST

Data group		Mean	Std.	Annotation region pairs	P(T<=t) two-tail
1	(*NC*)	60.95	51.58	89	3.89E-08
	(*CC*)	28.16	12.01	89	
2	(*NC*)	71.35	63.14	84	0.0066
	(*CC*)	57.46	48.04	84	
3	(*NC*)	80.45	51	72	2.61E-07
	(*CC*)	53.31	20.26	72	
4	(*NC*)	255.29	255.57	39	0.06E-05
	(*CC*)	127.17	99.16	39	
5	(*NC*)	182.29	141.78	86	9.72E-09
	(*CC*)	121.50	95.14	86	
6	(*NC*)	52.73	34.47	55	0.0011
	(*CC*)	44.96	22.83	55	

TABLE V
CONDITIONS UNDER WHICH THE EXECUTION TIME IS MEASURED

Data group	Specimen	Image size at final registration level (pixel)	levels
1	Liver#1	4480×3992	4
2	Liver#2	6808×6232	4
3	Liver#3	6660×5188	3
4	Intervertebral disc (1)	6748×5212	3
5	Intervertebral disc (2)	6748×5212	3
6	Intervertebral disc (3)	6860×5228	3

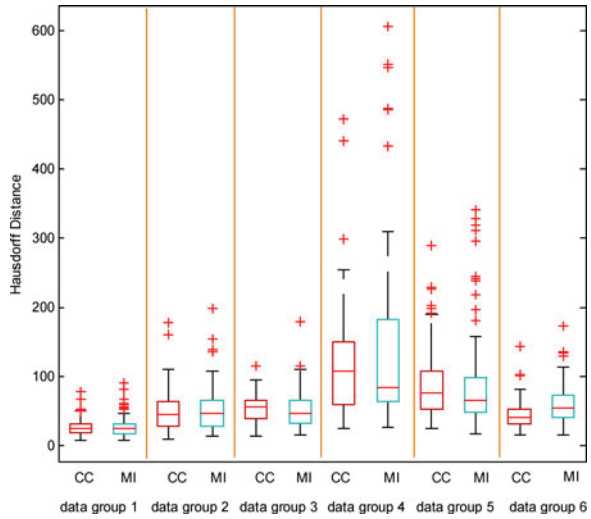

Fig. 16. Comparison of registration performances between content-classification-based method (CC) and a state-of-the-art MI-based method [7].

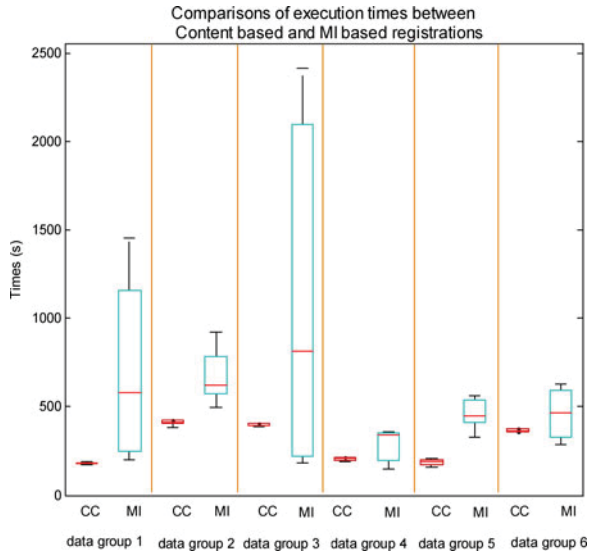

Fig. 17. Comparison of execution times between MI-based registration and content classification based registration (CC).

F. Experiment 4: The Influence of Sectioning Distance

The influence of sectioning distance between the differently stained sections on our classification method is of interest since we calculate co-occurrence statistics between image pairs. Large distances between sections (e.g., when the pair is not consecutive serial sections) may lead to noticeable variations in tissue structures between the image pairs, which could affect the correctness of the classification.

It is not ideal to use the real anatomic structures as the ground truth to evaluate the influence of sectioning distance on the registration accuracy, since there are natural variations between anatomical structures on different tissue sections. Data group 7 (see Table II and Fig. 8) was designed for this experiment as four artificial vertical holes were drilled before the specimen was sectioned, and subsequently annotated (see Fig. 8). In total, 49 image pairs were evaluated, with 25 consecutive sections (section distance $\sim$5 μm), and 24 nonconsecutive sections (average section distance $\sim$30 μm). Fig. 18 shows that the distance between sections can affect the registration accuracy that is decreased when the gap between image pairs increases. This experiment indicates the desirability of registering as closely spaced sections as possible.

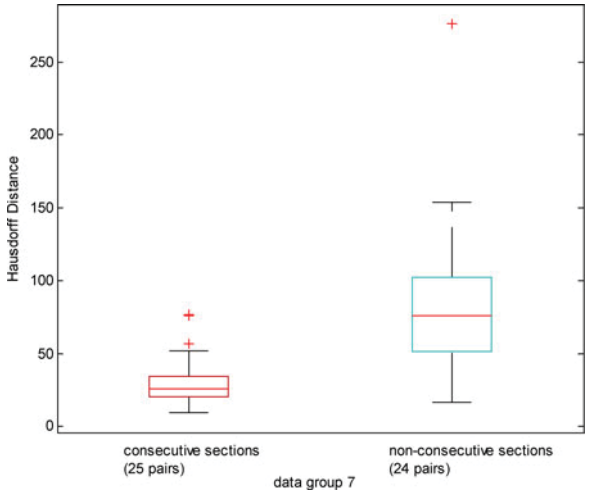

Fig. 18. Comparison between the registration accuracy of the consecutive image pairs (25 pairs, 100 annotations) and that of nonconsecutive image pairs (24 pairs, 96 annotations).

V. Conclusion

This paper is motivated by the application of registering differently stained 2-D histology images. One of the main challenges of such a registration problem is the dissimilarity of structural appearance between image pairs. This paper addresses this challenge by proposing a novel unsupervised content classification method that automatically identifies common content classes from differently stained histology image pairs. The algorithm outputs multichannel probability images, where each channel corresponds to one content class. The advantage of using the probability images for registration is that they enhance the structural similarity between the image pair. Since the content classes are emergent from co-occurrence statistics between image pairs, they do not necessarily map to real anatomical tissue classes, which is also not prerequisite for registration. Comparing to previous work on multistained histology image registration, our content classification based registration method makes no assumption on either the stain types or the existence of particular anatomic structures. This makes it applicable to a wide variety of tissue types, stains, and problem domains with no manual parameter tuning, as demonstrated in our evaluation.

Experiments 2 and 3 proved that our classification based registration method outperformed both intensity based and mutual information based registration methods. Other potential applications of the registration method (in addition to those described in this paper) could be to study co-expression of multiple biomarkers in Tissue microarrays, or identify cancer compartments based on one stain and relate this to structure (cells, vessels, etc.) identified using a separate stain.

REFERENCES

[1] M. B. Klevesath, L. G. Bobrow, S. E. Pinder, and A. D. Purushotham, "The value of immunohistochemistry in sentinel lymph node histopathology in breast cancer," *Brit. J. Cancer*, vol. 92, pp. 2201–2205, 2005.

[2] C. Derham, H. Yow, J. Ingram, J. Fisher, E. Ingham, S. A. Korrosis, and S. Homer-Vanniasinkam, "Tissue engineering small-diameter vascular grafts: Preparation of a biocompatible porcine ureteric scaffold," *Tissue Eng. Part A*, vol. 14, no. 11, pp. 1871–1882, Nov. 2008.

[3] V. Wijayathunga, A. Jones, R. Oakland, N. Furtado, R. Hall, and R. Wilcox, "Development of specimen-specific finite element models of human vertebrae for the analysis of vertebroplasty," in *Proc. Institution Mech. Eng. Part H: J. Eng. Med.*, 2008, vol. 222, pp. 221–228.

[4] A. Can, M. Bello, H. E. Cline, X. Tao, F. Ginty, A. Sood, M. Gerdes, and M. Montalto, "Multi-modal imaging of histological tissue sections," in *Proc. 5th IEEE Int. Symp. Biomed. Imag. Nano Macro*, 2008, pp. 288–291.

[5] L. Cooper, O. Sertel, J. Kong, G. Lozanski, K. Huang, and M. Gurcan, "Feature-based registration of histopathology images with different stains: An application for computerized follicular lymphoma prognosis," *Comput. methods programs biomed.*, vol. 96, no. 3, pp. 182–92, 2009.

[6] L. Cooper, S. Naidu, G. Leone, J. Saltz1, and K. Huang, "Registering high resolution microscopic images with different histochemical stains—A tool for mapping gene expression with cellular structures," in *Proc. Workshop Microscopic Image Anal. Appl. Biomed.*, 2007, pp. 1–7.

[7] D. Mueller, D. Vossen, and B. Hulsken, "Real-time deformable registration of multi-modal whole slides for digital pathology," *Comput. Med. Imag. Graph.*, vol. 35, no. 7–8, pp. 542–556, 2011.

[8] B. Ma, Z. Lin, S. Winkelbach, W. Lindenmaier, and K. Dittmar, "Automatic registration of serial sections of mouse lymph node by using Image-Reg," *Micron*, vol. 39, no. 4, pp. 387–396, 2008.

[9] N. Roberts, D. Magee, Y. Song, K. Brabazon, M. Shires, D. Crellin, N. Orsi, P. Quirke, and D. Treanor, "Toward routine use of 3D histopathology as a research tool," *Amer. J. Pathology*, vol. 180, no. 5, pp. 1835–1842, 2012.

[10] A. Ruiz, M. Ujaldon, L. Cooper, and K. Huang, "Non-rigid registration for large sets of microscopic images on graphics processors," *J. Signal Proc. Syst.*, vol. 55, no. 1, pp. 229–250, 2009.

[11] C. Kuglin and D. Hines, "The phase correlation image alignment method," in *Proc. Int. Conf. Cybern. Soc.*, 1975, pp. 163–165.

[12] D. Casasent and D. Psaltis, "Position, rotation, and scale invariant optical correlation," *Appl. Opt.*, vol. 15, no. 7, pp. 1795–1799, 1976.

[13] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual information matching in multiresolution contexts," *Image Vision Comput.*, vol. 19, no. 1–2, pp. 45–52, Jan. 2001.

[14] F. Maes, D. Vandermeulen, and P. Suetens, "Comparative evaluation of multiresolution optimisation strategies for multimodality image registration by maximisation of mutual information," *Med. Image Anal.*, vol. 3, no. 4, pp. 373–386, 1999.

[15] K. Mosaliganti, T. Pan, R. Sharp, R. Ridgway, S. Iyengar, A. Gulacy, P. Wenzel, A de Bruin, R. Machiraju, K. Huang, G. Leone, and J. Saltz, "Registration and 3D visualization of large microscopy images," *Proc. SPIE*, vol. 6144, pp. 923–934, 2006.

[16] S. Ourselin, A. Roche, S. Prima, and N. Ayache, "Block matching: A general framework to improve robustness of rigid registration of medical images," in *Proc. Med. Image Comput. Comput. Assist. Intervention (MICCAI)*, 2000, vol. 1935, pp. 557–566.

[17] A. Pitiot, E. Bardinet, P. Thompson, and G. Malandain, "Piecewise affine registration of biological images for volume reconstruction," *Med. Image Anal.*, vol. 10, no. 3, pp. 465–483, 2005.

[18] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual information based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.

[19] S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi, "A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1205–1218, May 2012.

[20] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[21] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "Non-rigid multimodality image registration," *Med. Imag. Image Proc.*, pp. 1609–1620, 2001.

[22] Y. Song, D. Treanor, A. Bulpitt, and D. Magee, "3D reconstruction of multiple stained histology images," presented at 15th MICCAI Joint Workshop Histopathol. Image Anal.: Image Comput. Digit. Pathol. (HIMA), Nice, France, 2012.

[23] P. Boutros and A. Okey, "Unsupervised pattern recognition: An introduction to the whys and wherefores of clustering microarray data," *Briefings Bioinform.*, vol. 6, no. 4, pp. 331–343, 2005.

[24] D. Judd, P. Mckinley, and A. K. Jain, "Large-scale parallel data clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 871–876, Aug. 1998.

[25] S. K. Bhatia and J. S. Deogun, "Conceptual clustering in information retrieval," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 28, no. 3, pp. 427–436, Jun. 1998.

[26] H. Frigui and R. Krishnapuram, "A robust competitive clustering algorithm with applications in computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 450–465, 1999.

Authors' photographs and biographies not available at the time of publication.