Mutual Information-Based Methods to Improve Local Region-of-Interest Image Registration

K.P. Wilkie and E.R. Vrscay

Department of Applied Mathematics,
University of Waterloo, Waterloo, Ontario, Canada N2L 3G1
{kpwilkie, ervrscay}@uwaterloo.ca

Abstract. Current methods of multimodal image registration usually seek to maximize the similarity measure of mutual information (MI) between two images over their region of overlap. In applications such as planned radiation therapy, a diagnostician is more concerned with registration over specific regions of interest (ROI) than registration of the global image space. Registration of the ROI can be unreliable because the typically small regions have limited statistics and thus poor estimates of entropies. We examine methods to improve ROI-based registration by using information from the global image space.

1 Introduction

We are concerned with multimodal image registration – the geometric alignment of images obtained from different modalities [2]. A typical example occurs in the planning of cancer treatment by radiation therapy, where it is common to use data from both computer tomography (CT) and positron emission tomography (PET), to obtain anatomical and functional information, respectively.

Most image registration procedures compare images directly, with specific attention to landmarks, surfaces, or pixel/voxel intensity values. Limitations of these procedures have led to the use of information theoretic methods to characterize image alignment in terms of pixel correlations. Mutual information (MI), introduced in [5], has become a standard measure of image registration – see [3] for further developments.

Registration involves finding a spatial transformation to align a study image Y to a target image X. In simple cases, affine transformations suffice, but, non-linear transformations, based on elasticity for example, are required to account for deformations caused by patient positioning and internal organ movement.

In cancer treatment, it is common for diagnosticians to be most concerned with a particular region of the image, e.g., a cancerous lesion. Alignment is more important over the regions of interest (ROIs) than over the global images, and diagnosticians commonly improve ROI alignment by perturbing the global registration result. Directly registering ROIs by estimating MI is prone to error: The limited statistics provided by small regions are highly sensitive to noise and to the region of overlap.

M. Kamel and A. Campilho (Eds.): ICIAR 2005, LNCS 3656, pp. 63–72, 2005.

The region of overlap is the common area contained within both the target and transformed study images. It is a function of the transformation, and changes during registration. Pixels contained in the region of overlap determine the overlap statistics used to estimate MI. Overlap statistics can cause artificial increases in MI as images move out of alignment. To counteract this, Studholme *et al.* [4] proposed normalized mutual information (NMI) which is invariant of overlap statistics. Unfortunately, as we will show in Section 2.1, the limited statistics of small regions inhibit the effectiveness of NMI.

To improve ROI registration, we examine similarity measures that combine global image information with ROI information. Our methods employ convex combinations of either the MI of the images and ROIs, or the distributions of the images and ROIs. The extreme limits of such convex combinations correspond to registering the ROIs or registering the images. Our methods seem to improve the registration of ROIs when compared to global image registration.

2 Mathematical Preliminaries

We consider an image X as an array of pixels with greyscale values x_k, where x_k are random variables that assume the discrete greyscale values $g_1, g_2, \cdots, g_N$. Our images are rescaled to 8 bits/pixel, so that $N = 2^8 = 256$ and $g_k = k - 1$. Associated with X is the greyscale probability distribution $\mathbf{p} = (p_1, p_2, \cdots, p_N)$. Here p_k is the frequency of occurence of greyscale value g_k, normalized so that $\sum p_k = 1$. The *entropy* of X is defined as

$$H(X) = H(\mathbf{p}) = -\sum_{k=0}^{N-1} p_k \log p_k, \tag{1}$$

where log denotes $\log_2$, and H is measured in "bits." Entropy is a convex down function: For two probability distributions $\mathbf{p}$ and $\mathbf{q}$, and for $c \in [0, 1]$, entropy satisfies.

$$H(c\mathbf{p} + (1-c)\mathbf{q}) \geq cH(\mathbf{p}) + (1-c)H(\mathbf{q}). \tag{2}$$

Consider two images, X and Y, with respective distributions $\mathbf{p}$ and $\mathbf{q}$. The *relative entropy* between $\mathbf{p}$ and $\mathbf{q}$ is defined as [1]

$$D(\mathbf{p}\|\mathbf{q}) = \sum_{k=1}^{N} p_k \log \frac{p_k}{q_k}, \tag{3}$$

where by convention $0 \log \frac{0}{q} = 0$ and $p \log \frac{p}{0} = \infty$. Relative entropy is non-negative, and zero if and only if $\mathbf{p} = \mathbf{q}$, but it is not a metric since it is not symmetric and it does not satisfy the triangle inequality.

Now let $\mathbf{r}$ denote the joint distribution associated with images X and Y: for $1 \leq i, j \leq N$, r_{ij} is the probability of finding the greyscale values (g_i, g_j) in corresponding pixel pairs. It follows that $\mathbf{p}$ and $\mathbf{q}$ are marginals of $\mathbf{r}$, i.e.

$$\sum_{i=1}^{N} r_{ij} = q_j, \quad \sum_{j=1}^{N} r_{ij} = p_i. \tag{4}$$

If the pixels of X and Y are independent, then $r_{ij} = p_i q_j$ and the joint distribution is the product distribution $\mathbf{d}$, where $d_{ij} = p_i q_j$.

Mutual information, $I(X, Y)$, is defined as the relative entropy between the joint distribution $\mathbf{r}$ and the product distribution $\mathbf{d}$, i.e.

$$I(X, Y) = \sum_{i,j} r_{ij} \log \frac{r_{ij}}{p_i q_j}. \tag{5}$$

It is a measure of the distance from the joint distribution to the product distribution. The higher $I(X, Y)$ is, the more correlated are corresponding pixel pairs. Hence, mutual information is maximized when the two images are most correlated and the joint distribution is "furthest" from the product distribution. As pointed out in [1], $I(X, Y)$ is a measure of the amount of information that one random variable contains about another random variable. If the random variables x_i and y_i are independent for all i, then $I(X, Y) = 0$.

The *joint entropy* associated with the pair of images X and Y is

$$H(X, Y) = H(\mathbf{r}) = -\sum_{i,j} r_{ij} \log r_{ij}, \tag{6}$$

i.e. the entropy of the joint distribution. ¿From Equation (4), $H(X, X) = H(X)$. Mutual information can thus be expressed in terms of entropy,

$$I(X, Y) = H(X) + H(Y) - H(X, Y). \tag{7}$$

Note that $I(X, X) = H(X)$. The relationship between $H(X)$, $H(Y)$, $H(X, Y)$ and $I(X, Y)$ is expressed in a Venn diagram, as shown in Figure 2.2 of [1]. We will compare this to NMI, defined in [4] by

$$\hat{I}(X, Y) = \frac{H(X) + H(Y)}{H(X, Y)}. \tag{8}$$

2.1 Some Simple Examples Tailored to the Registration Problem

In Figure 1 we plot $I(X, Y)$ vs. n, where the target X is the 512×512 pixel *Lena* image, and Y is the same image shifted horizontally by n pixels. In the left plot, Y is periodic; the image wraps around the computation window which is fixed to the size of the target image. This case, though not realistic, is used to demostrate the effects of overlap statistics. In the right plot, X and Y are cropped to the region of overlap. This region decreases as $|n|$ increases, which adversely affects the statistics: As $|n|$ approaches the target image boundary, $I(X, Y)$ falsely indicates registration. In each case, $I(X, Y)$ shows a strong peak at registration, i.e. at $n = 0$. Note, noise greatly reduces the amplitude of these peaks.

To examine effects of limited statistics, we use two transaxial magnetic resonance (MR) brain images, the target is proton density (PD) weighted and the study is T2 relaxation time (T2) weighted; see Figure 2. Figure 3 shows registration curves for horiziontal shifts calculated for three ROIs of various size. The

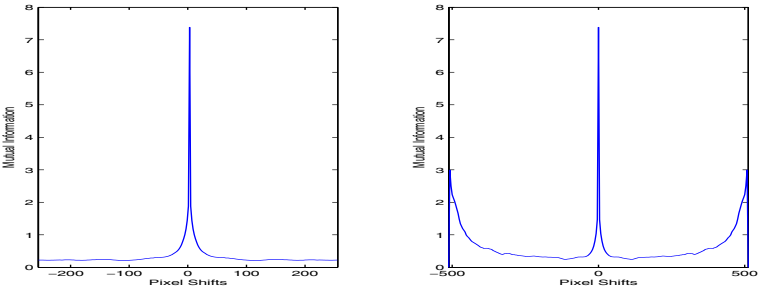

Fig. 1. MI for *Lena* and a horizontally shifted *Lena* using a fixed computation window with periodic boundary conditions (*left*) and the region of overlap (*right*)

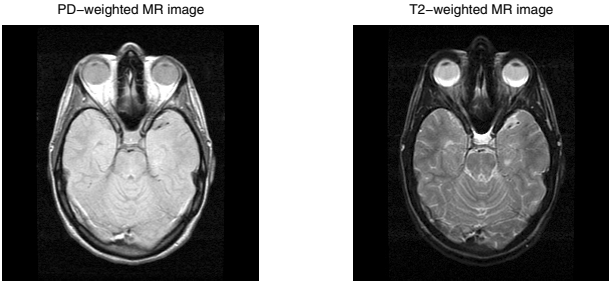

Fig. 2. PD-weighted (*left*) and T2-weighted (*right*) MR brain images

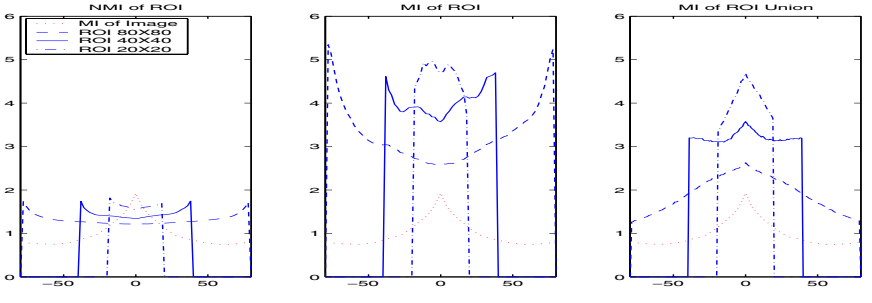

Fig. 3. Registration curves PD- and T2-weighted MR images and three ROI using NMI (*left*), MI (*middle*), and MI computed with ROI union statistics (*right*)

left and middle plots display NMI and MI respectively, both are computed with ROI overlap statistics (contributing pixels lie in the intersection of the ROIs). The right plot displays MI computed with ROI union statistics (contributing pixels lie in the union of the ROIs). Note the left and middle registration curves are distorted by the limited statistics of the ROI overlap: In both cases, no peaks occur at registration. The use of ROI union statistics, however, produce registration curves with clearly defined peaks.

These results suggest important consequences for medical image registration. For example, in prostate-focused registration, a typical 256×256 pixel image displays the prostate as a circle of radius 23 pixel-widths, an area of 1661 pixels. If the 40×40 pixel ROI from Figure 3 approximates the prostate, then even by quadrupling the number of pixels used to estimate MI, using the 80×80 pixel ROI, a meaningless registration curve will still be obtained.

3 Proposed Methods to Improve ROI Registration

As mentioned in Section 1, diagnosticians may be more concerned that images are most accurately aligned over a region of interest, as opposed to the global image. Small ROIs possess insufficient statistics to accurately estimate MI, but still contain useful information. We present two methods to perform ROI registration by "blending" the information of the ROI with that of the global image.

3.1 Method 1: Weighted Mutual Information

A crude way of performing such blending is to construct a convex combination of the MI for two separate problems, namely, (i) the registration of the ROIs and (ii) the registration of the global images. Thus we define

$$\begin{aligned} J(X, Y; c) &= (1-c)I(X_{ROI}, Y_{ROI}) + cI(X, Y), \\ &= (1-c)I(\mathbf{p}_{ROI}, \mathbf{q}_{ROI}) + cI(\mathbf{p}, \mathbf{q}), \quad 0 \le c \le 1. \end{aligned} \tag{9}$$

Here, X_{ROI} and Y_{ROI} denote the ROIs of images X and Y, with distributions $\mathbf{p}_{ROI}$ and $\mathbf{q}_{ROI}$, respectively. The case $c = 0$ corresponds to registration of the ROIs and $c = 1$ corresponds to registration of the global images. $J(X, Y; c)$, which we call *weighted mutual information* (WMI), performs a linear interpolation between these two cases. Note that WMI is *not* a mutual information function in the strict sense.

To facilitate optimization, a desirable feature of such blending would be to enhance the peak of the similarity measure at registration. This, however, may not be achievable: Since the distributions in Equation (9) do not depend on c, it follows that

$$\frac{\partial J(X, Y; 0)}{\partial c} = I(\mathbf{p}, \mathbf{q}) - I(\mathbf{p}_{ROI}, \mathbf{q}_{ROI}). \tag{10}$$

Hence, it is not guaranteed *a priori* that this derivative is positive. Depending upon the nature of the ROIs, it is possible that the RHS of Equation (10) is negative, implying an initial *decrease* in WMI by the inclusion of global image statistics. However, it is not only the *amplitude* of the peaks which is of concern but also the *location* of the peaks, i.e., the point of registration.

This method can be altered to use the normalized mutual information from each of the ROI and image, instead of the mutual information. We call this function WNMI, and it is computed in an analagous way to WMI, see Equation (9).

3.2 Method 2: Mutual Information of Weighted Distributions

We now propose a method that employs weighted probability distributions – a kind of statistical blending of an image with its ROI. For images X and Y, we define the following weighted distributions (N-vectors), for $0 \leq c \leq 1$:

$$\begin{aligned} \mathbf{p}_c &= (1-c)\mathbf{p}_{ROI} + c\mathbf{p} \\ \mathbf{q}_c &= (1-c)\mathbf{q}_{ROI} + c\mathbf{q}, \end{aligned} \tag{11}$$

and the following weighted joint distribution ($N \times N$ matrix),

$$\mathbf{r}_c = (1-c)\mathbf{r}_{ROI} + c\mathbf{r}. \tag{12}$$

By construction, the above weighted distributions are probability distributions that linearly interpolate between the ROI and global image statistics.

To construct a similarity measure, we use these weighted distributions in the definition of mutual information. From Equation (7), we define

$$K(X,Y;c) = H(\mathbf{p}_c) + H(\mathbf{q}_c) - H(\mathbf{r}_c). \tag{13}$$

$K(X,Y,c)$ is the *mutual information of weighted distributions* (MIWD) derived from the global images and their regions of interest. Note the two limiting cases,

1. $K(X,Y;0) = I(X_{ROI}, Y_{ROI})$, the MI of the ROIs of images X and Y.
2. $K(X,Y;1) = I(X,Y)$, the MI of images X and Y.

Using the convexity of entropy, Equation (2), and the fact that $\log(x)$ is an increasing function of x, one can derive the following inequality from Equation (13) to relate WMI and MIWD (see Appendix for details):

$$K(X,Y;c) \geq J(X,Y;c) - E(c), \tag{14}$$

where

$$E(c) = -(1-c)\log(1-c) - c\log c. \tag{15}$$

Note that $E(c)$ is the entropy of a binary random variable with distribution $(c, 1-c)$, so $E(0) = E(1) = 0$ and $E(1/2) = 1$.

In Figure 4 we see the relationship between WMI and MIWD for our PD- and T2-weighted MR images at registration, using two different ROIs for weighting: a brain matter ROI and an eye socket ROI. Notice that as the weighting parameter c changes, moving from the ROI to the global image, WMI is a linear interpolation of the two values, whereas the behaviour of MIWD is dependent on the statistics of the ROIs and images. (From the plot, we see that the MIWD function is not convex.) The value of c which provides the right amount of mixed statistics to most accurately register regions of interest is difficult to determine, and the subject of future investigations. In the discussions to follow, we have used $c = \frac{1}{2}$, corresponding to an equal weighting of both the image and the ROI statistics.

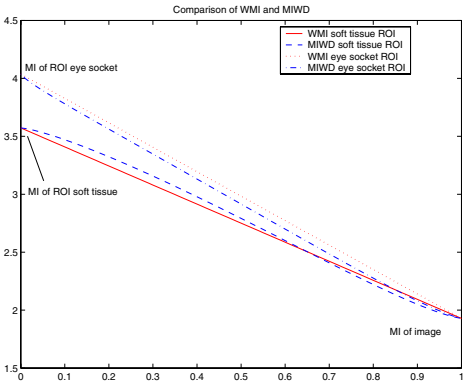

Fig. 4. WMI and MIWD *vs.* the weighting parameter c, for PD- and T2-weighted MR brain images at registration using a flat ROI (*dashed*) and an active ROI (*dotted*)

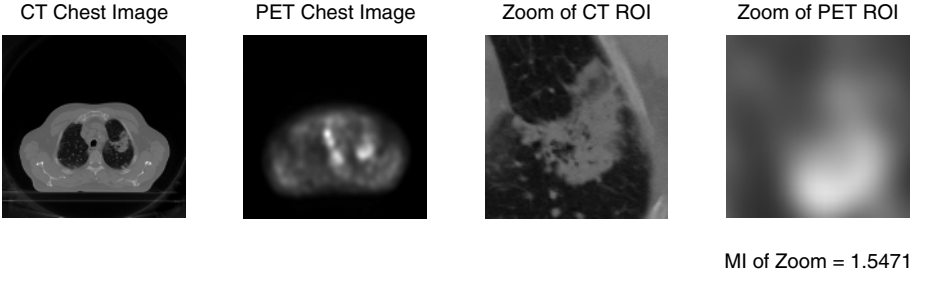

Fig. 5. Original CT and PET transaxial chest images (*left*). Zooms of the regions of interest in CT image space (*right*).

4 Results of an ROI-Based Registration Experiment

We have constructed a simple experiment using PET and CT transaxial chest images that were manually aligned by a rigid-body transformation based on fiducial point matching. Our goal is to improve the alignment of the unusually bright area in the PET image with the visible tumor in the CT image. The ROIs were independently defined for each image as approximately 3700 pixel polygons. The images and a zoom of the ROIs (as defined by the target CT image) are shown in Figure 5. Notice the poor alignment of the tumor regions.

For simplicity, the registration transformation was limited to vertical translations. The resulting curves are shown in Figure 6. Negative shifts correspond to upward translations, and positive to downward translations, of the study PET image with respect to the target CT image. The curves for WNMI, WMI, and MIWD were computed using two rules:

1. If the ROIs do not overlap, use the value of the global MI or NMI (set $c = 1$).
2. If the ROIs overlap, weight using ROI union statistics.

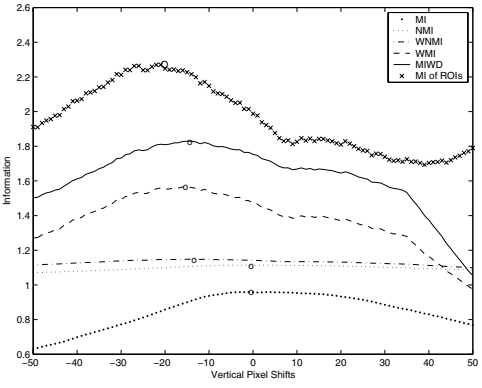

Fig. 6. Registration curves for vertical translations of the PET image with respect to the CT image, using ROI union statistics when needed. The curves represent MI (*dots*), NMI (*dotted*), WNMI (*dash-dot*), WMI (*dashed*), MIWD (*solid*), and MI of the ROIs (*xxx*). The maximum of each curve is indicated by a circle.

In order to maintain continuity of the similarity measure, for optimization purposes, the value of c is made to decrease continuously from $c = 1$, at no ROI overlap, to the desired weighting factor, here $c = \frac{1}{2}$, at and above some percentage of ROI overlap. One can think of this as slowly zooming in the statistics from the global images to the local ROIs.

In Figure 6, each curve's maximum determines the vertical shift required to register the images (and ROIs) according to the similarity measure defining the curve. The registration transformations determined by MI and NMI are small translations, about 1 pixel, vertically up from the original position. The transformation determined by registering only the ROIs, using ROI union statistics, is a much larger upward translation, about 20 pixels. However, the registration transformations determined by WNMI, WMI, and MIWD, lie in between these two extremes with vertical translations around 13 pixels. Shifts of 13 pixels – roughly 11 mm – are acceptable and may be compensating for internal organ deformations caused by different positioning of the patient in PET and CT.

The registration curve for MI of ROIs, in Figure 6, demonstrates that it is a poor similarity measure. The curve is rough with multiple local maxima which inhibit optimization. Furthermore, it may not be desirable to discard the global image information even when performing ROI-based image registration.

Figure 7 shows the results of registration by each similarity measure over the zoomed region. Below each image is the mutual information calculated over the zoomed region. While MI is affected by the limited statistics of the zoom region (90×90 pixels), we use it as a quantitative measure of alignment for lack of something more robust. MI is highest for registration by MI of the ROIs, and second highest for registration using mixed statistics. Visually, the ROIs in the bottom row (registration by WNMI, WMI, and MIWD) are more accurately registered than those in the top row (registration by MI, NMI, and MI of ROIs).

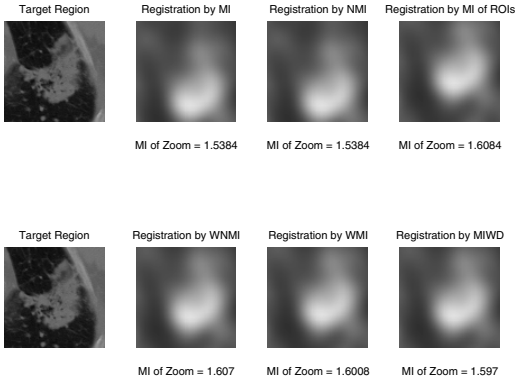

Fig. 7. Zoom of ROI as defined in the CT image (*top row left*). Results of registration using MI, NMI, and MI of ROIs (*top row left to right*), and using WNMI, WMI, and MIWD (*bottom row left to right*).

5 Summary and Concluding Remarks

In this paper, we have proposed two methods of blending image statistics with ROI statistics to improve the registration of small regions. The first method, WMI, is a simple convex combination of the mutual information of the global images with that of the ROIs. The second method, MIWD, takes the mutual information of the convex combinations of the marginal and joint probability distributions from both the global images and ROIs.

We also explored the use of ROI union statistics as opposed to ROI overlap statistics when dealing with small regions. Using the ROI union region is a reasonable way to avoid the problems associated with using the overlap region, especially for regions that are small to begin with. Taking the union region of global images will reduce the effect of overlap statistics, but may have the possibly negative effect of causing a registration of the background instead of the foreground objects. This is not a problem for ROI-based registration, especially in medical images, where the ROIs typically lie inside the foreground object.

The experiments presented above suggest the best registration result for small ROIs may be obtained by maximizing WNMI, WMI, or MIWD using ROI union statistics. We are still investigating the robustness of these similarity measures with respect to the weighting parameter c as well as the content of the images and regions of interest.

In applications, the ROI can be any interesting area (or areas) of the image, for example significant edges. Our methods are easily modified for these cases. One of the original motivations of this study was the idea of an "activity-based" registration procedure, where the ROIs of an image are decided by some criteria, e.g., variance or local entropy.

Finally, there may well be other mathematical ways of combining region statistics, a subject that we are currently exploring.

Acknowledgements

We thank Dr. R. Barnett, Medical Physics Division, Grand River Regional Cancer Center, Kitchener, Ontario, for suggesting this problem, for helpful conversations and for supplying the CT and PET data. We also thank Prof. J. Orchard, School of Computer Science, University of Waterloo, for the two MR brain images. This research has been supported in part by the Natural Sciences and Engineering Research Council of Canada in the form of a Postgraduate Scholarship (KPW) and Individual Research Grant (ERV).

References

1. T.M. Cover and J.A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
2. J.V. Hajnal, D.L.G. Hill and D.J. Hawkes, Editors, *Medical Image Registration* (CRC Press, Boca Raton, 2001).
3. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal and P. Suetens, Multimodality image registration by maximization of mutual information, IEEE Trans. Image Proc. **16**, 187-198 (1997).
4. C. Studholme, D.L.G. Hill, D.J. Hawkes, An overlap invariant entropy measure of 3D medical image alignment, Pattern Recognition. **32**,71-86 (1999).
5. P. Viola and W.M. Wells, III, Alignment by maximization of mutual information, Proc. 5th Int. Conf. Computer Vision, 16-23 (1995).

Appendix

In this Appendix, we derive the inequality in Eq. (14). From the convexity property of entropy,

$$\begin{aligned} H(\mathbf{p}_c) &\geq (1-c)H(\mathbf{p}_{ROI}) + cH(\mathbf{p}) \\ H(\mathbf{q}_c) &\geq (1-c)H(\mathbf{q}_{ROI}) + cH(\mathbf{q}). \end{aligned} \tag{16}$$

As for the final term in (14) (we use superscripts for c and ROI to minimize notational complexity)

$$\begin{aligned} -H(\mathbf{r}_c) &= \sum r_{ij}^c \log r_{ij}^c \\ &= \sum \left[(1-c)r_{ij}^{ROI} + cr_{ij}\right] \log \left[(1-c)r_{ij}^{ROI} + cr_{ij}\right] \\ &= (1-c)\sum r_{ij}^{ROI} \log \left[(1-c)r_{ij}^{ROI} + cr_{ij}\right] + \\ &\qquad + c\sum r_{ij} \log \left[(1-c)r_{ij}^{ROI} + cr_{ij}\right] \\ &\geq (1-c)\sum r_{ij}^{ROI} \log \left[(1-c)r_{ij}^{ROI}\right] + c\sum r_{ij} \log \left[cr_{ij}\right] \\ &= (1-c)\sum r_{ij}^{ROI} \log r_{ij}^{ROI} + c\sum r_{ij} \log r_{ij} + (1-c)\log(1-c) + c\log c \\ &= -(1-c)H(\mathbf{r}_{ROI}) - cH(\mathbf{r}) + (1-c)\log(1-c) + c\log c. \end{aligned}$$

We have used the fact that the elements of $\mathbf{r}$ and $\mathbf{r}_{ROI}$ sum to unity. Combining these results yields the desired inequality.