Improving Accuracy and Efficiency of Registration by Mutual Information using Sturges' Histogram Rule

Philip A. Legg[a,*], Paul L. Rosin[a,†], David Marshall[a,‡] and James E. Morgan[b,§]

[a]School of Computer Science, Cardiff University. [b]School of Optometry, Cardiff University.

Abstract. Mutual Information is a common technique for image registration in the medical domain, in particular where images of different modalities are to be registered. In this paper, we wish to demonstrate the benefits of applying a common method known in statistics as Sturges' Rule for selecting histogram bin size when computing Entropy as a part of the existing Mutual Information algorithm. Although Sturges' Rule is well known in the field of statistics it has received little attention in the Computer Vision community. By augmenting Mutual Information with Sturges' Rule, we show that this offers an improvement to both the runtime of the algorithm and also the accuracy of the registration. Our results are demonstrated on images of the eye, in particular, Fundus images and SLO (Scanning Laser Ophthalmoscopy) images.

1 Introduction

Mutual Information is a widely used measure for performing image registration in the medical imaging domain, due to its ability to register images of different modalities [1]. Mutual Information relies greatly on a measure known as entropy, which can be thought of as the amount of information an event provides when it occurs [2]. Mutual Information is defined as $I(A, B) = H(A) + H(B) - H(A, B)$, where $H(A)$ is the entropy of the template image, $H(B)$ is the entropy of the section of the reference image at which the template image is currently located and $H(A, B)$ is the joint entropy of the two. We wish to find the registration transformation that maximises $I(A, B)$.

Computation of entropy is based on the probability of the values within the data set, defined as $\sum_{i=0}^{n} -p(i) \log_2 p(i)$ where $p(i)$ is the probability of intensity i occuring within the data set n. One possible approach to finding this probability distribution is by using a histogram. There are alternative methods that exist, such as using a Parzen Window [3], B-splines or k-Nearest Neighbours [4], along with more recent techniques such as that described in [5], although using a histogram tends to be the most popular choice due to its simplicity and computational efficiency.

In the Computer Vision literature relating to Mutual Information, very little mention has been made regarding the selection of histogram bin size. Most papers use a fixed number of bins either equal to or less than the possible data range [6], but this means that no consideration is given to the data being organised. In [7], they state that no method exists for predicting the exact number of bins to use for a histogram, which is clearly not the case, as we shall demonstrate in our work. Even in comprehensive reviews of Mutual Information such as [2], there is no mention of how bin size should be selected and how this could affect the performance of Mutual Information based registration. Nevertheless, bin size is a *crucial* parameter. Excessive quantisation caused by too large a bin size will result in important information being lost. On the other hand, too small a bin size may result in many bins becoming sparsely populated, and consequently making the probability density estimates unreliable.

Sturges' Rule is a well known method used in statistics for histogram binning [8]. Sturges' Rule is one possible technique for determining the size of each group that the data should be separated into, to try give the optimum group size. Another common method in the statistics literature for estimating bin size is Scott's Rule [9]. Scott's Rule is thought to be an improvement over Sturges' Rule as Sturges' Rule can over-smooth a histogram which may be problematic in some applications [10]. We shall consider Scott's Rule within our study and compare how this performs alongside with Sturges' Rule and using 256 bins.

Applying Sturges' Rule can provide a two fold benefit. Reducing the number of bins will reduce computational cost and so improve runtime. It can also be seen as cleaning up an image that may contain irrelevant and distracting detail, consequently improving the accuracy and robustness of the registration process. Sturges' Rule is derived on the assumption of normally distributed data, through the application of a binomial approximation. It is not intended for

*E-mail: P.A.Legg@cs.cardiff.ac.uk
†E-mail: Paul.Rosin@cs.cardiff.ac.uk
‡E-mail: Dave.Marshall@cs.cardiff.ac.uk
§E-mail: morganje3@cardiff.ac.uk

use on data that, for example, features several, well-separated peaks. However, our area of research is targeted towards retinal images, in particular, Fundus images and SLO images [11], where such features of the histogram are not present. In this paper, we shall demonstrate the effects that Sturges' Rule has on these images, and evaluate the performance that this technique offers in Mutual Information based registration in comparison to using 256 histogram bins or using Scott's bin size rule.

2 Method

We will begin by stating the formulae for Sturges' Rule and Scott's Rule. We will then demonstrate the effects that Sturges' Rule can have on an image and its associated histogram. A comparison of number of bins given by Sturges' Rule and Scott's Rule can be found in Section 3.1. We will also discuss Histogram Equalization as a method for bin size selection.

2.1 Sturges' Rule Definition

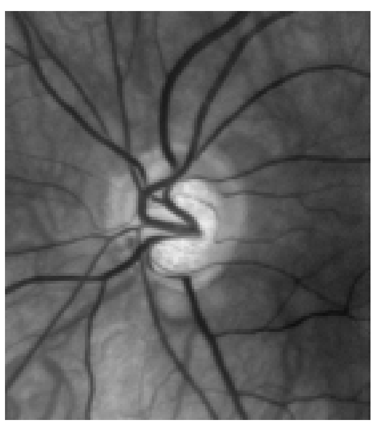
Sturges' Rule is defined as $w = \frac{r}{1+log_2(n)}$ where r is the range of values within the data set, and n is the number of elements within the data set [8]. The result will give the ideal bin width, w, to be used for the histogram (i.e. the range for each group of values). To find the number of bins for an image, we simply use r/w.

2.2 Scott's Rule Definition

Scott's Rule is defined as $w = 3.49\sigma n^{-1/3}$ where σ is the standard deviation of the data set, n is the number of elements within the data set [9], and w is the bin width.

2.3 Sturges' Rule in Practice

In Figure 1, we present a typical 8 bit Fundus photograph of the eye, along with its associated histogram. The effect that Sturges' Rule has had on the histogram is clear to see in Figure 2. Although the histogram contains just 16 bins, the requantized image has retained all the significant detail from the original in Figure 1.



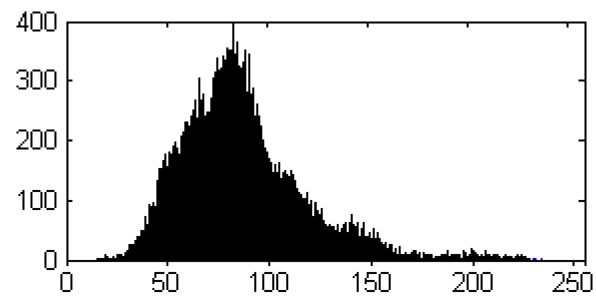


Figure 1. Fundus image with associated histogram

It can be noticed that the new histogram does not capture the exact shape of the original histogram and distinct peaks have been lost due to binning which may lead to important information being lost. Comparing the actual images shown in Figures 1 and 2, we can see the difference that Sturges' Rule has made. Although the images are still very similar in what they represent, areas that were originally smooth have now become solid areas, with more noticable steps between intensity changes. This may be useful in some situations where, like here, the background intensities are unclear and highly varied which may have an adverse effect when it comes to processing the image.

Applying Scott's rule, the Fundus image is represented using 145 bins. Although this has reduced the number of bins slightly, there would be much less of a difference to the original image compared with that of the Sturges' Rule image. The distinct intensity changes seen in the background of the Sturges' Rule image would not be evident on the Scott's Rule image.

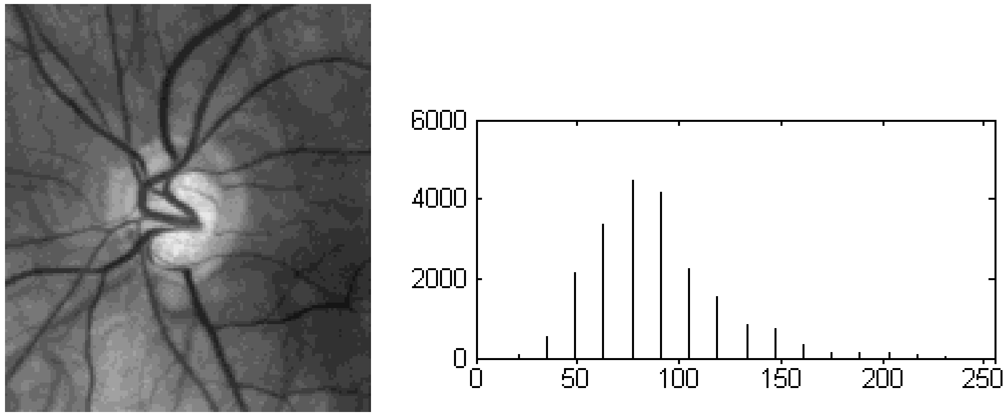


Figure 2. Fundus image with associated histogram, after applying Sturges' Rule

2.4 Histogram Equalization

One other method that is related to bin size estimation is histogram equalization [12]. This normally applies a monotonic remapping of intensity values to make the intensity histogram approximately flat, with the aim of improving visibility of features. As a consequence, adjacent sparsely populated bins are merged, thereby improving probability estimates and reducing the number of bins. The effects of histogram equalization in essence allow for variable bin width within a histogram. We shall investigate the effects of histogram eqaulization on registration in our paper.

3 Testing

Mutual Information can be used to effectively perform image registration on two images of different modalities, by transforming the template image on to the reference image, such that it maximises Mutual Information. Our aim is to successfully register our image data correctly, with a high rate of accuracy that is also time efficient.

For our work, we have two images captured from the eye, a Fundus photograph which will be our reference image, and an SLO (Scanning Laser Ophthalmoscopy) image which is our template image. It can be seen in Figure 3 that the images are of the same source but have different appearances due to the information captured by each camera. For the benefit of this paper, the images have been rotated and scaled appropriately beforehand, as we will just report on the effect of Sturges' Rule on the accuracy and runtime of estimating translation. Similar effects were found for the remaining transformation parameters.

To find where the correct registration occurs, we have adopted two approaches, exhaustive search and hill climbing. Exhaustive search will attempt to match the template image to every possible position on the reference image. This method can cause our search to check areas where we do not wish our images to match at, for instance, in our data we know that the registration will occur around the centre of the reference so searching the edges is not necessary. However, it provides a fair result for the image as a whole that does not rely on having any previous knowledge of the data being registered. Hill climbing is a more common search technique that will start at a given point (in our case, this point will be the centre of the reference image) and try to improve on the existing result by testing local neighbouring positions. This provides a much faster search method, although may not give the true result if caught in a local maximum which differs from the global maximum.

For our testing, we use twenty-six 8 bit greyscale image pairs which we shall perform registration on, using the traditional 256 bin representation, Sturges' Rule, Scott's Rule and Histogram Equalization. The number of bins used for computing the joint entropy is found by the number of bins used in template image $\times$ the number of bins used in reference image segment, since this is computed by means of a 2-D histogram. The dimensions for the images are 153×137 for the reference images, and 50×51 for the template images. In each case, we shall attempt to register the images using the exhaustive search and hill climbing search methods.

3.1 Results

Following our testing of Mutual Information using our four approaches, we wish to quantify the alignment with respect to the ground truth manually determined by a clinician. To do this, we have calculated the error of the translation. The tables below shows the average translation error, the average number of bins used (represented by (A,B) where A is the number of bins used for the reference image and B is the number of bins used for the template image), the average runtime and the number of successful registrations (where a match is found within a 2 pixel radius of the ground truth).

	$SturgesRule$	$256Bins$	$ScottsRule$	$HistogramEqualization$
Average Translation Error	17.15	24.92	18.96	20.46
Average Number of Bins	(14,14)	(256,256)	(134,79)	(60,64)
Average Runtime (secs)	7.38	19.74	13.21	10.54
Successful Registrations	13	2	10	8

Table 1. Results for performing Mutual Information using Exhaustive Search

	$SturgesRule$	$256Bins$	$ScottsRule$	$HistogramEqualization$
Average Translation Error	10.73	14.15	11.34	12.53
Average Number of Bins	(14,14)	(256,256)	(112,78)	(59,64)
Average Runtime (secs)	0.119	0.239	0.172	0.137
Successful Registrations	11	5	8	7

Table 2. Results for performing Mutual Information using Hill Climbing

As can be seen in Table 1, Sturges' Rule gives the lowest translation errors, compared to the other methods used, and manages to successfully register 13 of the 26 test images. Histogram equalization provides a fair improvement over 256 bins by successfully registering 8 of the 26 images. We note that Scott's Rule can give a large difference between the number of bins used in the reference image and the template image. In some cases, Scott's Rule would give an estimated bin size greater than 256. This is not necessarily a problem, but the results of registration are worse than Sturges' Rule. Scott's Rule manages to register 10 of the images and 256 bins only manages to register 2. It is surprising that using 256 bins gives such a low success rate, although in comparison with the other methods presented here, it is likely due to having very sparsely populated bins, especially so in the joint histogram. The joint histogram would have 65536 possible bins, where as there are at most only 2550 points to be binned (the true value would most likely be less as this assumes each pixel and its corresponding pixel in the other image be a unique combination).

Comparing the runtime of performing exhaustive search, using 256 bins takes 19.74 seconds compared to Sturges' Rule which takes 7.38 seconds. As expected, this reduction is due to the fewer number of bins that the entropy formula has to be calculated for. It can be seen that Scott's Rule and Histogram Equalization also offer an improvement to runtime compared to 256 bins.

Table 2 shows that when using Hill Climbing, Sturges' Rule manages to successfully register 11 of the 26 images. Although this is less than the number of registrations by exhaustive search, it can be seen that the translation error has reduced. This anomaly shows that although Hill Climbing may not register the images exactly, the results are much closer to the ground truths than when exhaustive search fails to register. This is because exhaustive search can potentially place the template far away from the desired position due to the nature of the search technique, which is a common occurrence on the failed registrations. Scott's Rule and Histogram Equalization experience a similar situation to Sturges' Rule when using Hill Climbing, which suggests that the algorithms are caught by local maxima within the search space. In comparison, when using 256 bins, we notice an increase in successful registrations. This is due to being caught by local maxima which, in these cases, has been the correct registration. This improvement is purely coincidental and could not be guaranteed when registering other sources. The runtime of Hill Climbing is reduced greatly due to the limited nature of the search, with Sturges' Rule taking 0.119 seconds.

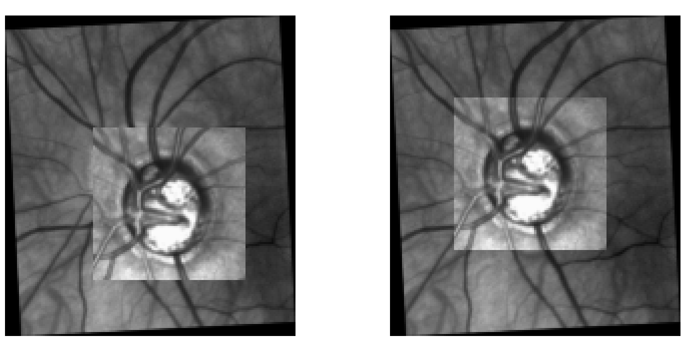


Figure 3. Results of registration using Exhaustive search (256 Bins Vs. Sturges Rule)

Figure 3 shows a comparison between using 256 bins and using Sturges' Rule for selecting the number of bins. It is clear to see that when 256 bins are used, the registration is incorrect. This is likely to be caused by the noisy background that is detracting from the data that we are actually concerned with (the blood vessels and optic disc). When using Sturges' Rule to perform registration, we can see that it is aligned correctly with all corresponding blood vessels matching up. By using Sturges' Rule to reduce the complexity of an image, it is shown that we can achieve better results for registration.

4 Discussion

In this paper we have demonstrated the benefits of using Sturges' Rule in Mutual Information. It is a common method in statistics, but has only ever been briefly mentioned with regards to computing entropy. Our testing shows that it has a large effect as to whether Mutual Information can actually perform the registration correctly, along with improving the runtime of the algorithm.

In the statistics literature, there are concerns that Sturges' Rule can smooth the histogram too much [10], and that Scott's Rule is a better approach to estimate bin size. In the context of image registration for our data we have found this *not* to be the case since Sturges' Rule consistently outperformed Scott's Rule, although both methods are still better than using a traditional 256 bin representation.

Bin size selection is just one aspect that can affect the performance of Mutual Information. We can clearly see from the results that although Sturges' Rule offers an improvement on registration, there is still much scope for developing the algorithm further to give satisfactory results for our data. One drawback of the standard Mutual Information measure is that it is calculated on a pixel by pixel basis, so much spatial information is lost. Existing techniques have attempted to resolve this by performing Mutual Information over the neighbourhood of each individual pixel [13] or by combining the standard Mutual Information measure with local gradient information from the image [14]. We wish to develop on these methods further to allow for successful registration of our image data.

It has been shown that Sturges' Rule offers a simple yet effective way to depict the original image histogram that can be used for determining the entropy result and improve upon the Mutual Information measure, for both accuracy and efficiency.

References

1. P. A. Viola and W. M. Wells III. Alignment by maximization of mutual information . In *ICCV*, pages 16–23, 1995.
2. J. P. W. Pluim, J. B. Antoine Maintz, and M. A. Viergever. Mutual information based registration of medical images: A survey. *IEEE Trans. Med. Imaging*, 22(8):986–1004, 2003.
3. N. Kwak and C. Choi. Input feature selection by mutual information based on Parzen Window. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(12):1667–1671, 2002.
4. N. Nicolaou and S. J. Nasuto. Mutual information for EEG analysis. *Proc. 4th IEEE EMBSS UKRI Postgraduate Conference on Biomedical Engineering and Medical Physics (PGBIOMED'05)*, pages 23–24, 2005.
5. T. Kadir and M. Brady. Estimating statistics in arbitrary regions of interest. *Proceedings of British Machine Vision Conference (BMVC)*, pages 6–7, 2005.
6. Y. Zhu and S. M. Cochoff. Influence of Implementation Parameters on Registration of MR and SPECT Brain Images by Maximization of Mutual Information. *J Nucl Med*, 43(2):160–166, 2002.
7. A. Rajwade, A. Banerjee, and A. Rangarajan. Continuous image representations avoid the histogram binning problem in mutual information based image registration. In *ISBI*, pages 840–843, 2006.
8. H. A. Sturges. The choice of a class interval. *J. American Statistical Association*, pages 65–66, 1926.
9. D. W. Scott. On optimal and data-based histograms. *Biometrika*, 66(3):605–610, 1979.
10. M. P. Wand. Data-based choice of histogram bin width. *The American Statistician*, 51(1):59–60, 1997.
11. P. L. Rosin, D. Marshall, and J. E. Morgan. Multimodal retinal imaging: new strategies for the detection of glaucoma. In *ICIP (3)*, pages 137–140, 2002.
12. M. Sonka, V. Hlavac, and R. Boyle. Image Processing, Analysis, and Machine Vision. *Brooks and Cole Publishing*, 1998.
13. D. B. Russakoff, C. Tomasi, T. Rohlfing, and C. R. Maurer Jr. Image similarity using mutual information of regions. In *ECCV (3)*, pages 596–607, 2004.
14. J. P. W. Pluim, J. B. Antoine Maintz, and M. A. Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE Trans. Med. Imaging*, 19(8):809–814, 2000.