

# Robust Inter-Slice Intensity Normalization using Histogram Scale-Space Analysis.

J. Dauguet<sup>1,2,3</sup>, J.-F. Mangin<sup>1</sup>, T. Delzescaux<sup>1</sup>, and V. Frouin<sup>1</sup>

<sup>1</sup> Service Hospitalier Frédéric Joliot, CEA, Orsay, France

<sup>2</sup> INRIA, Epidaure Project, Sophia Antipolis, France

<sup>3</sup> Ecole Centrale Paris, Laboratoire de Mathématiques Appliquées aux Systèmes, Châtenay-Malabry, France

**Abstract.** This paper presents a robust method to correct for intensity differences across a series of aligned stained histological slices. The method is made up of two steps. First, for each slice, a scale-space analysis of the histogram provides a set of alternative interpretations in terms of tissue classes. Each of these interpretations can lead to a different classification of the related slice. A simple heuristic selects for each slice the most plausible interpretation. Then, an iterative procedure refines the interpretation selections across the series in order to maximize a score measuring the spatial consistency of the classifications across contiguous slices. Results are presented for a series of 121 baboon slices.

## 1 Introduction

Histological data often present large and discontinuous changes of intensities between slices. The sectioning process, indeed, does not assure a perfect constant thickness for all the slices. The staining intensity is sensitive to these variations of thickness, and is also sensitive to the concentration of the solution and the incubation time. Finally, the acquisition of the slices on the scanner can accentuate these differences.

The design of postprocessing procedures overcoming these intensity inhomogeneities is complex for several reasons:

- the number of homogenous tissues and their relative sizes vary from one slice to another;
- the intensity distributions may present high discontinuities from one block of contiguous slices to another : actually, the brain after extration is usually divided into small blocks mainly for a more homogeneous fixation;
- the relative contrasts between the intensity ranges corresponding to the different tissue classes can be radically different from one slice to another;
- large spatial inhomogeneities can occur inside some slices, which disturbs the classification of the different types of tissues, especially in case of low contrast.

A previous attempt to overcome this kind of problems led to match the intensity distribution of contiguous slices using a strategy based on a 1D linear registration of histograms [3]. This approach was successful for a dataset including two tissue classes only and presenting a good contrast. However, we observed some failures of the same approach for datasets corrupted by the various artifacts mentioned above. Some of these

failures resulted from local maxima of the similarity measure used to register the histograms. In the worst cases, the global maximum of this similarity measure did not correspond to the targeted registration.

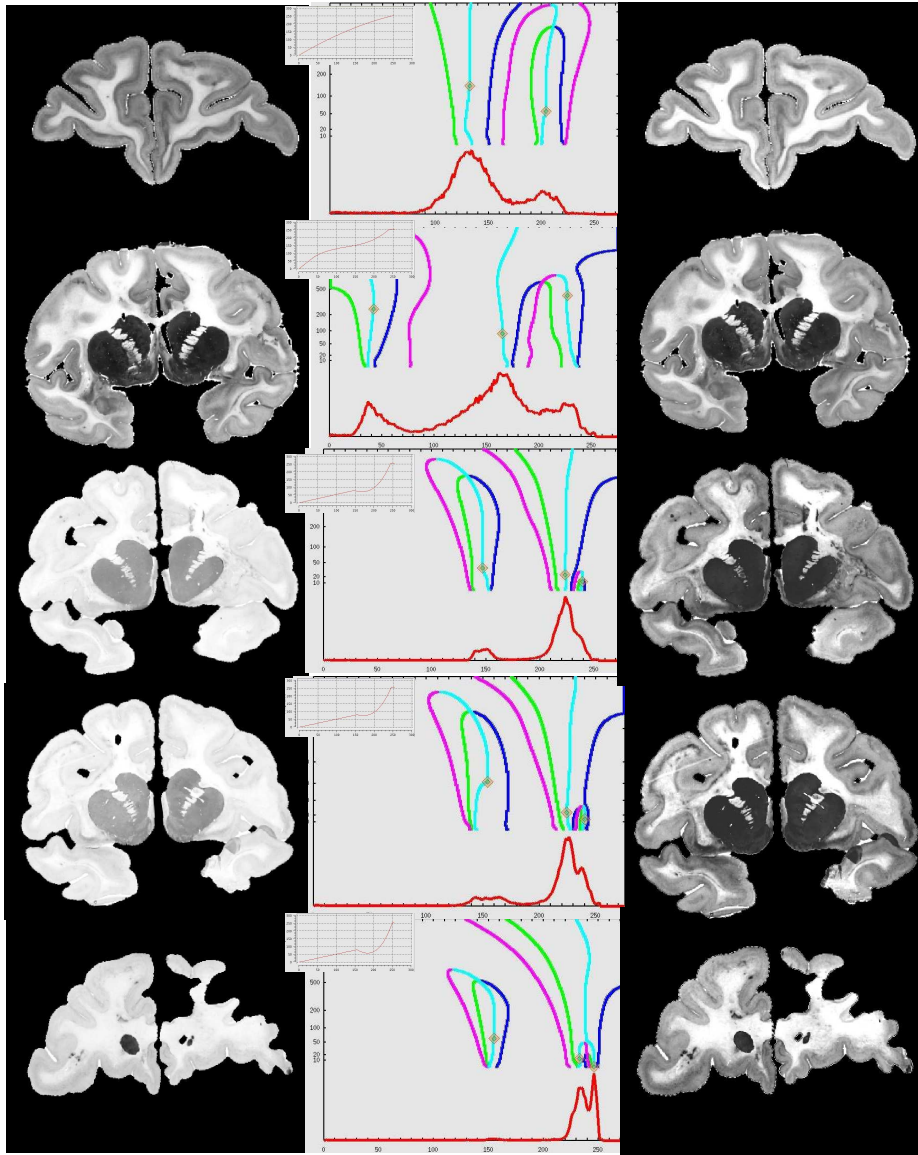
We present in this paper a different strategy which aims at increasing the robustness of the intensity restoration process. Some of the difficulties of the linear-registration based approach could be overcome using some non linear normalization [6], in order to get the possibility to deal with more than two classes. Such an approach, however, may still be sensitive to minimization problems. Therefore, rather than performing a blind registration between histograms, our new strategy detects the ranges of intensities corresponding to the different tissues before matching the histograms. For the sake of robustness, an iterative algorithm updates these detections until achieving consistent tissue classifications across contiguous slices. This new strategy is applied on a set of 121 realigned histological slices of a baboon brain acquired with a resolution of 0.16mm in the sectioning incidence (coronal) and an inter-slice distance of 0.72mm. The slices have been stained with a general marker which created a contrast between gray and white matter (Nissl staining), and a more specific marker of the basal ganglia and thalamus (Acetylcholinesterase histochemistry) which produces a very intense staining. The set has been realigned following a method described in [1].

## 2 Method

The method described in this paper is made up of two steps. First, for each slice, a multi-scale histogram analysis is performed. This analysis provides several possible interpretations of the histogram shape in terms of underlying tissue classes. Each interpretation can be used to compute a classification of the corresponding slice. A simple heuristic selects for each slice the most likely interpretation according to *a priori* knowledge about the acquisition process. The second step leads to question the spatial consistency of the resulting classifications across contiguous slices in order to detect some failures of this heuristic. This second step is based on an iterative process. For each slice, a simple score is used to rank the set of histogram interpretations according to consistency with the interpretations selected for the neighboring slices. If the current selected interpretation is not the best one, an update is applied. Several loops on the set of slices are performed until convergence of the process.

### 2.1 Histogram scale-space analysis

Because of the various artifacts mentioned in the introduction, few *a priori* knowledge on the shapes of the histograms can be used to analyse them automatically. The main one lies in the relative positions of the tissues: the marker of the basal ganglia is darker than grey matter, which is darker than white matter. Another interesting property is true for most of the slices: the contrast between basal ganglia and grey matter is higher than the contrast between grey and white matters. Otherwise, the distributions of intensities of each of the tissues vary largely across slices. In the worst cases, one given tissue can correspond to several histogram modes (several maxima), induced by some variability in the staining process or by intra-slice spatial inhomogeneities. The fact that one class of tissue can be represented by several neighboring histogram modes leads to develop a multiscale strategy. Smoothing the histogram, indeed, is supposed to mix together the modes corresponding to the same tissue before merging the different classes.



**Fig. 1.** Some examples of the behaviour of the heuristics analysing successfully most of the slices of the dataset. **left:** raw slice, **middle:** The scale-space trajectories of the extrema used to interpret the histogram. Cyan (respectively violet) color corresponds to the second derivative minima (resp. maxima). Green and dark blue colors correspond to first derivative extrema. Diamond shapes mark the drift velocity minima corresponding to the scale of the detected modes. The polynomial normalizing the slice intensity is superimposed. **right:** normalized slice.

Linear scale-space analysis is an appealing multiscale approach when dealing with 1D signals because of the simplicity of the extrema behavior induced by the causality property [7, 2]. This scale-space can be computed by applying the heat equation to the signal, which corresponds to smoothing it with a Gaussian kernel with increasing width. The extrema of the initial signal can be tracked throughout the scales until their extinction, which occurs when one maximum and one minimum get in touch with each other at a bifurcation point. For a 1D signal, the number of maxima always decreases with scale until reaching a stage where only one maximum remains. Since extrema of the histogram and of its first derivatives often have direct semantic interpretations, they allow the analysis of the histogram shape relative to the underlying tissue classes.

In simple cases, smoothing the histogram until only three maxima survive is sufficient to detect the modes corresponding to the three classes of tissues. The general situation, however, requires a better heuristics. For instance, the slices which do not cross the basal ganglia lead to only two classes of tissues, which has to be inferred. More difficult cases occur when the contrast between grey and white matter is very low. In such cases the two related histogram modes merge very early in the scale-space, while several maxima induced by the basal ganglia may still exist. In the worst cases, this contrast is so low that the part of the histogram corresponding to grey and white matter includes only one maximum (see Fig. 1). It has been shown elsewhere, however, that such situations can be overcome through the use of the second derivative minima [5]. This larger set of extrema, which is related to the histogram curvature, embeds more information about the various classes of tissues making up the slice.

Hence, each second derivative minimum is supposed to mark a mode of the histogram. Moreover, the location of this minimum along the intensity axis provides an estimation of the average intensity of this mode. This location, however, varies with the scale. The mode represented by this minimum, indeed, varies also with scale. Intuitively, a minimum may stand for an artifactual mode at low scales, and represent a tissue of interest at higher scales after several artifactual modes have been merged together by the smoothing. During this merging process, the location of the minimum along the intensity axis varies rapidly from the average intensity of the artifactual mode to the average intensity of the tissue of interest. In fact, most of the extrema trajectories in scale-space alternate such periods of high horizontal drift velocity with periods of stability which could be related to scale ranges where they are caught by some underlying mode. Therefore, our method associates a different mode to each local minimum of the horizontal drift velocity along the trajectory of second derivative minima [5].

In order to sort these modes in decreasing order of interest, two features are used: the extinction scale of the second derivative minimum trajectory they belong to, and an evaluation of the number of slice's pixels they stand for. This evaluation is the integral of the histogram in the range defined by the locations at the mode's scale of the second derivative minimum and the closest first derivative minimum [5]. First derivative minima correspond in fact to the largest slope points of the histogram's hills. Hence, this integral corresponds approximately to half the volume of the mode in terms of slice's pixels. Several interpretations of the histogram shape can be derived from the modes at the top of the hierarchy. The interpretation systematically selected as the most plausible one results from the following heuristics:

1. Compute the scale-space until only one second derivative minimum survives;
2. Detect the highest drift velocity minimum along the trajectory of this minimum, which stands for a mode GW representing the sum of grey and white matter;

3. Detect the second derivative minimum trajectory with the highest extinction scale located on the right of the first derivative minimum associated to the left slope of GW;
4. Detect the next drift velocity minimum along the first trajectory and the highest one along the second trajectory. The minimum on the left stands for a mode G representing grey matter, the one on the right stands for a mode W representing white matter;
5. Detect the second derivative minimum with the highest extinction scale located on the left of the intensity range explored previously. The highest mode M is associated to the marker of the basal ganglia if its volume is more than 100 pixels.

The Figure 1 gives a few examples of the behaviour of this heuristics. For each slice, alternative interpretations are also derived from the scale-space of the histogram, to deal with potential failures of this heuristics. They correspond for instance to associating the modes G and W respectively to the leftmost and to the rightmost two biggest modes. Here is the list of the different alternative interpretations taken into account for the results presented in this paper:

1. The biggest mode can stand for either “G” or “W” modes (extreme or pathological slices);
2. The two biggest modes stand for “G and W modes”;
3. The three biggest modes stand for “M, G and W modes” or “M, G and X modes”, where the mode X does not lead to a class of tissue;
4. The four biggest modes stand for “M, X, G and W modes”.

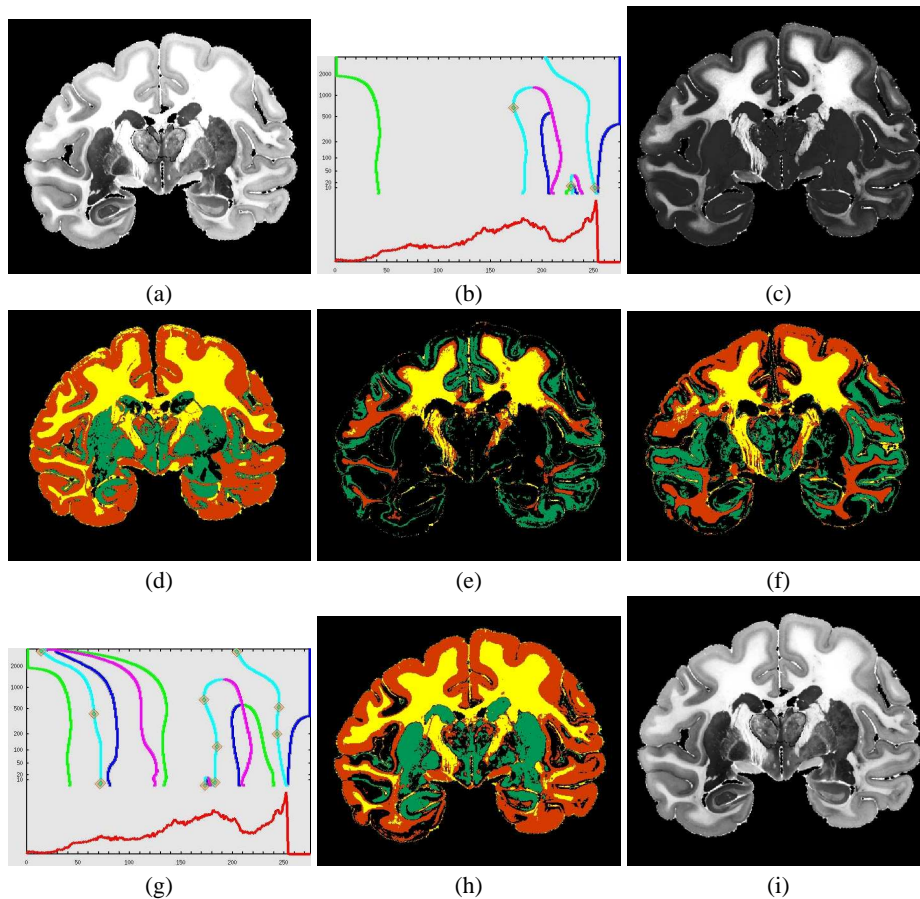
It should be noted that this list of interpretations does not explore all the combinatorial possibilities and results from a tuning to some events found in our data.

Each interpretation can be simply converted into a classification of the respective slice: the histogram is split into classes by thresholds defined as the middle points between the estimated mode’s averages. These classifications may include one, two or three classes.

## 2.2 Classification spatial consistency

The second step of the process detects the failure of the main heuristics, using for each slice a score based on the consistency between the slice’s classification and the two classifications of the contiguous slices. This score is based on the number of voxels which belong to the same class in two contiguous slices. For each slice, the score obtained by the heuristics interpretation is compared to the scores obtained by the other interpretations. The interpretation with the best score is selected (see Fig; 2). This process is iterated until convergence, which occurred in two or three iterations in our experiments, thanks to the robust behaviour of the heuristics.

During a final stage, we correct the intensity of each slice by estimating the best polynomial matching the detected classes of tissues with template values selected by the user. The degree of this polynomial (1, 2, or 3) depends on the number of tissues found in the slice. Three-degree polynomials are replaced by a linear interpolation on the  $[0, m]$  segment, where  $m$  stands for the position of the marker maximum, to avoid polynomial “rebunds” problems.



**Fig. 2.** An histological stained slice (a), its heuristic histogram analysis (b), the corresponding normalisation. The classification of the previous slice (d), of the current slice following the heuristics (e) and of the next slice (f). Alternative histogram analysis with 4 biggest mode detected (M,X,G,W) (g), the best classification according to the consistency with the neighboring slices (h) and the final corrected slice (i).

### 3 Results and Discussion

We present in figure 3 a slice orthogonal to the sectioning incidence, first before any processing, second after intensity normalization using only the heuristics, and finally after the whole process. The iterative step has converged after two iterations. The final histogram interpretation selected by the whole process stems from the heuristics for about 85% of the slices. For the final intensity normalization, we set the average intensity of the white matter to 230, the intensity of the grey matter to 160 and finally 80 for the marker. As far as computing time is concerned, a complete normalization of the whole volume  $500 \times 440 \times 121$  with 2 iterations on a Pentium IV 1.8GHz takes about 10

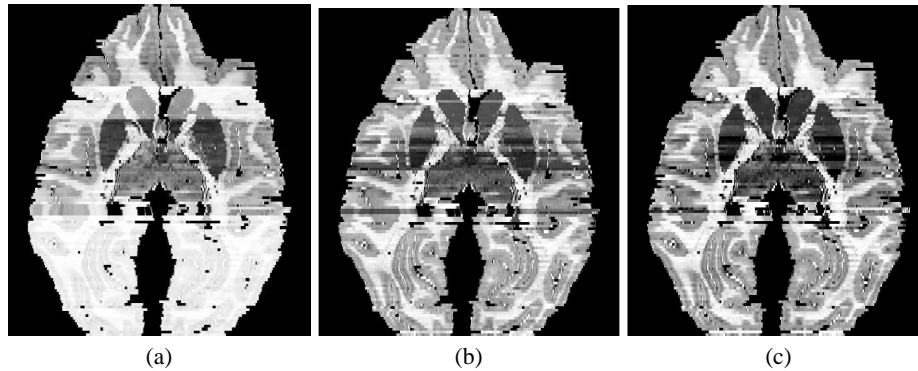
minutes. The result is globally satisfying. In fact, the remaining problems (imperfections in the marker regions) stem mainly from intra-slice spatial inhomogeneities that were not taken into account by the process described in this paper, or from pathological sections.

The Figure 1 provides an insight into the variability of the histogram shapes to be analysed. The first and the second rows show simple histograms with two or three modes. The third row shows an example of low contrast where the white matter mode does not lead to any maximum in the histogram but is recovered via the curvature-related second derivative minimum. The fourth row shows a successful estimation of the average intensity of the marker at high scale while this class of tissues is split into a lot of modes at lower scales. The fifth row shows the detection of the marker mode in a limit case where it is made up of a few hundred pixels, namely for a slice located at the beginning of basal ganglia.

The figure 2 describes the behaviour of the second step of the method for a slice leading to a failure of the heuristics. The competition between the alternative interpretations of the histogram leads to the selection of the (M,X,G,W) hypothesis. In fact none of the interpretations is fully satisfying in terms of classification, because of a large spatial inhomogeneity inside the slice. Therefore, our algorithm can only select the interpretation leading to the best 3D consistency.

This kind of situations calls for applying some intra-slice bias correction procedure somewhere into the restoration process. The very low grey/white contrast of some of our slices, however, prevented us to perform such a correction systematically before intensity normalization. Indeed, the risk of mixing definitively the range of intensities corresponding to grey and white matter is too high. To prevent such problems to occur, we plan to address the bias correction process through a two stage procedure performed on a slice by slice basis. An initial correction will be performed before intensity normalization using a highly regularized correction field preventing any mixing between grey and white matter. A final correction will be performed with more degrees of freedom for the field after interpretation of the histogram. This interpretation, indeed, yields an estimation of the contrast between grey and white matter, which can be used to perform a slice by slice tuning of the bias field regularization. First experiments using an entropy minimization approach are promising but beyond the scope of this paper [4].

While our method includes some *ad hoc* tuning to some of our dataset features (the heuristics and the set of alternative interpretations), the strategy is generic and could be adapted easily to other kinds of dataset. Thanks to the iterative correction process, indeed, no strong robustness is required for the initial heuristics. The main requirement to assure a good behaviour of the whole process is that the set of alternative histogram interpretations is rich enough to cover all the possible cases. In our opinion, with a reasonable intra-slice bias, the hypothesis that each tissue class leads to one of the biggest scale-space mode should give the possibility to build a small successful set of alternative interpretations for various kinds of data. Compared to a blind registration procedure prone to local maxima of the similarity measure [3], our point of view assumes that each of these potential local maxima corresponds to a different matching of the histograms modes. Hence our method selects first the most frequent matching with a template histogram shape provided by the heuristics and iteratively propagates this choice throughout the series of slices in order to achieve 3D spatial consistency of the resulting classification.



**Fig. 3.** An axial view of the histological volume : before intensity correction (a), after the heuristic strategy (b) and after the spatial consistency checking (c).

## 4 Conclusion

We have presented in this paper a generic strategy to perform robust histogram normalization throughout a series of registered images. This method can deal with variations of the number of tissue classes throughout the series and provides a 3D consistent classification of the data. While the approach described in the paper requires an heuristics providing an initial correct classification of most of the slices, the framework may be extended to complex situations where no such heuristics is available. The iterative step indeed could consists in a global stochastic minimization of the sum of scores evaluating spatial consistency of the classification of contiguous slices.

## References

1. T. Delzescaux, J. Dauguet, F. Condé, R. Maroy, and V. Frouin. Using 3D non rigid FFD-based method to register post mortem 3D histological data and in vivo MRI of a baboon brain. In *MICCAI, Montréal, LNCS 2879, Springer Verlag*, pages 965–966, 2003.
2. J.J. Koenderink. The structure of images. *Biol. Cybernetics*, 50:363–370, 1984.
3. Grégoire Malandain and Eric Bardinet. Intensity compensation within series of images. In *MICCAI*, volume 2879 of *LNCS*, pages 41–49, Montréal, Canada, 2003. Springer Verlag.
4. J.-F. Mangin. Entropy minimization for automatic correction of intensity nonuniformity. In *IEEE Work. MMBIA*, pages 162–169, Hilton Head Island, South Carolina, 2000. IEEE Press.
5. J.-F. Mangin, O. Coulon, and V. Frouin. Robust brain segmentation using histogram scale-space analysis and mathematical morphology. In *Proc. 1st MICCAI, LNCS-1496*, pages 1230–1241, MIT, Boston, Oct. 1998. Springer Verlag.
6. S. Prima, N. Ayache, A. Janke, S. Francis, D. Arnold, and L. Collins. Statistical Analysis of Longitudinal MRI Data: Application for detection of Disease Activity in MS. In *MICCAI*, pages 363–371, 2002.
7. A.P. Witkin. Scale-space filtering. In *In International Joint Conference on Artificial Intelligence*, pages 1019–1023, 1983.