A Bayesian Approach to Perfusion Imaging Using ASL MRI*

Miguel L. Rodrigues[1,2,**], Patrícia Figueiredo[1,2], and João Miguel R. Sanches[1,2]

[1] Institute for Systems and Robotics
[2] Department of Bioengineering,
Instituto Superior Técnico, Technical University of Lisbon
m.lourenco.rodrigues@ist.utl.pt

Abstract. *Arterial Spin Labeling* (ASL) is a non-invasive technique for generating perfusion images of the brain. Following an alternating labeling/control acquisition sequence, the small magnetization difference between labeled and non-labeled images is usually detected by performing image subtraction. In order to increase the *Signal to Noise Ratio* (SNR) a large number of trials is needed to observe these signal differences. In this work, the magnetization difference estimation problem is formulated in a Bayesian framework, where spatio-temporal priors are used to deal with the ill-posed nature of the estimation task. The *a priori* assumption that no drastic signal variations are expected along the same tissue, except at the organ boundaries, is modeled by Gibbs distribution functions. To evaluate the performance of the proposed algorithm, the results obtained using synthetic data were compared against the two most common subtraction methods usually discribed in the literature. The results are very encouraging. A real data set is used to illustrate the application of the method and the results are consistant with the traditional methods.

Keywords: arterial spin labeling, bayesian approach, perfusion, magnetization, spatio-temporal priors.

1 Introduction

Cerebral Blood Flow (CBF) is a measure of the volume of blood passing through a point in the brain circulation per unit of time and, in a healthy individual, alterations in neural activity lead to changes in local CBF[1]. If the volume of the tissue is taken into account, the *perfusion* can also be measured, as the CBF per unit volume of tissue. In fact, perfusion is the process by which the nutrients in the blood stream are delivered to the tissues through the capillary bed.

Arterial Spin Labeling Magnetic Resonance Imaging (ASL MRI) perfusion image modality is an emerging brain imaging technology since the last two decades. In the ASL technique, the blood passing by a given region (upstream from the

* This work was supported by the FCT project [PEst-OE/EEI/LA0009/2011].
** Corresponding author.

J.M. Sanches, L. Micó, and J.S. Cardoso (Eds.): IbPRIA 2013, LNCS 7887, pp. 691–698, 2013.

region of interest) is labeled with an inversion pulse, and after a certain time interval (TI), when the blood reaches the region of interest, an image is acquired. If a control image, where no labeling was performed, is subtracted from the labeled acquisition, the resulting image has a small magnetization difference caused by the exchange of labeled water molecules from the blood to the surrounding tissue, which can be used to quantify CBF[2]. Since the magnetization difference is a very small fraction of the tissue signal (approximately 1-2%), a set of label-control acquisitions has to be performed to increase SNR.

In this work, pair-wise subtraction and surround subtraction[3] will be used to assess and compare the results obtained by the algorithm.

2 Problem Formulation

In this paper, an alternative method to the traditional image subtraction strategies is proposed.

Let $Y(t)$ be a sequence of L Pulsed ASL (PASL) images with $N \times M$ pixels. The observation model adopted in this paper is

$$Y(t) = F + D(t) + v(t)\Delta M + \mathbf{\Gamma}(t) \tag{1}$$

where $t \in \{1, 2, ..., L\}$ is an image index, $Y(t)$ is the t^{th} noisy image within the sequence, F is a time invariant $N \times M$ image describing the static magnetization of the tissues, $D(t)$ is a slow variant image describing the baseline fluctuations of the signal along time, called *Drift*, and ΔM is the magnetization variation in the tissues caused by the alternate inversion process occurring at each *inversion time* (TI), as described by the general kinetic model proposed by Buxton and colleagues[4], illustrated in Figure 1. $\mathbf{v}(t)$ is a binary signal indicating the labeled periods related with the inversion process. The image $\mathbf{\Gamma}(t) = \{\eta_{i,j}(t)\}$ is assumed

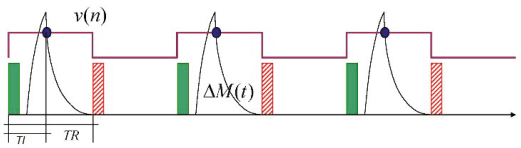

Fig. 1. Schematics of PASL technique. The rectangular signal (purple) represents $v(t)$, and its value is equal to one for labeling acquisition and zero for a non-labeling acquisition. TI is the instant where the acquisition is made, and TR is the time period before another sequence of labeling or non labeling can occur.

to be *Additive White Gaussian Noise* (AWGN) where $\eta_{ij}(t) \sim \mathcal{N}\left(0, \sigma_y^2\right)$ are stationary independent and identically distributed (iid) random variables with Gaussian distribution.

In order to simplify the problem formulation and algorithm design let us consider the following alternative formulation for the observation model described in equation 1,

$$\mathbf{Y} = \mathbf{f}\mathbf{u}^T + \mathbf{D} + \mathbf{\Delta M}\mathbf{v}^T + \mathbf{N} \tag{2}$$

where the t^{th} columns of $\mathbf{Y} = \{y_i(t)\}$, $\mathbf{D} = \{d_i(t)\}$ and $\mathbf{N} = \{\eta_i(t)\}$ are vectorized versions of the corresponding images, arranged by *lexicographic* order of their pixel index [5]. In practice, these image sequences, after vectorization, are arranged in $NM \times L$ matrices. The time invariant images F and $\mathbf{\Delta M}$, after vectorization, give rise to the $NM \times 1$ column vectors $\mathbf{f}$ and $\mathbf{\Delta}M$, respectively. The signal $v(t)$ is arranged in a L column vector, $\mathbf{v}$, and $\mathbf{u}$ is a constant vector of ones with the same dimension.

The probability of $\mathbf{Y}$ is a multivariate Gaussian distribution with mean $\mu = \mathbf{f}\mathbf{u}^T + \mathbf{D} + \mathbf{\Delta M}\mathbf{v}^T$ and covariance diagonal matrix, $\sigma_y^2\mathbf{I}$, because the noise is white,

$$p(\mathbf{Y}) \sim \mathcal{N}(\mu, \sigma_y^2\mathbf{I}). \tag{3}$$

The *Maximum Likelihood* (ML) estimation of the unknown images, $\boldsymbol{\theta} = \{\boldsymbol{f}, \mathbf{D}, \boldsymbol{\Delta m}\}$, may be formulated as follows

$$\boldsymbol{\theta} = \arg\min_{\theta} E_y(\mathbf{Y}, \boldsymbol{v}, \boldsymbol{\theta}) \tag{4}$$

where the energy function

$$\begin{aligned} E_y(\mathbf{Y}, \boldsymbol{v}, \boldsymbol{\theta}) &= -\log p(Y|\boldsymbol{\theta}, \boldsymbol{v}) \\ &= ||\boldsymbol{f}\boldsymbol{u}^T + \mathbf{D} + \boldsymbol{\Delta m}\boldsymbol{v}^T - \mathbf{Y}||^2 + C \end{aligned} \tag{5}$$

is called *Data Fidelity Term*. The optimization task described in equation (4) is an *ill-posed* problem [6] and regularization is needed. By using the *maximum a posteriori* (MAP) criterion the regularization is introduced by the prior distribution of the parameters. In this approach, the new energy function to be minimized is $E(\mathbf{Y}, \boldsymbol{v}, \boldsymbol{\theta}) = -\log p(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{v})p(\boldsymbol{\theta})$ and the estimation process is formulates as follows,

$$\boldsymbol{\theta} = \arg\min_{\theta} E(\mathbf{Y}, \boldsymbol{v}, \boldsymbol{\theta}) \tag{6}$$

where

$$E(\mathbf{Y}, \boldsymbol{v}, \boldsymbol{\theta}) = E_y(\mathbf{Y}, \boldsymbol{v}, \boldsymbol{\theta}) + E_\theta(\boldsymbol{\theta}) \tag{7}$$

with $E_\theta(\boldsymbol{\theta}) = -\log p(\theta)$. Assuming independence in the parameters $\boldsymbol{f}$, $\mathbf{D}$ and $\boldsymbol{\Delta m}$ in $\boldsymbol{\theta}$, the prior term can be decomposed as follows,

$$E_\theta(\boldsymbol{\theta}) = E_f(\boldsymbol{f}) + E_{\Delta M}(\boldsymbol{\Delta m}) + E_D(\mathbf{D}) \tag{8}$$

Here, the parameter images $\boldsymbol{f}$, $\mathbf{D}$ and $\boldsymbol{\Delta}\mathbf{m}$ in $\boldsymbol{\theta}$ are considered *Markov Random Fields* which means its priors are Gibbs distributions [6],

$$p(\boldsymbol{\tau}) = \frac{1}{Z_\tau} e^{-\alpha_\tau U(\boldsymbol{\tau})} \tag{9}$$

where $U(\boldsymbol{\tau})$ is called *Gibbs* energy with $\boldsymbol{\tau} \in \{\boldsymbol{f}, \mathbf{D}, \boldsymbol{\Delta}\mathbf{m}\}$ and α_τ are the prior hyper-parameters. The *Gibbs* energy for 2D images is

$$U(\boldsymbol{\tau}) = \sum_i \left(\delta_h^2(i) + \delta_v^2(i)\right) \tag{10}$$

where i is the index of each pixel of the image and $\delta_h(i)$ and $\delta_v(i)$ are the differences of the i^{th} pixel to its horizontal and vertical neighbors respectively.

In case of time varying *Drift* imaged, $\mathbf{D}$, a third term is added to account for the temporal dimension

$$U(\mathbf{D}) = \sum_{i,t} \left(\delta_h^2(i,t) + \delta_v^2(i,t) + \delta_t^2(i,t)\right) \tag{11}$$

where $\delta_t(i,t) = d_i(t) - d_i(t-1)$.

Equation (7) can be written in the matricial form:

$$\begin{aligned} E\ \ &(\mathbf{Y}, \boldsymbol{v}, \boldsymbol{\theta}) = \\ \tfrac{1}{2}\ \ &Tr[(\boldsymbol{f}\boldsymbol{u}^T + \mathbf{D} + \boldsymbol{\Delta}\mathbf{m}\boldsymbol{v}^T - \mathbf{Y})^T(\boldsymbol{f}\boldsymbol{u}^T + \mathbf{D} + \boldsymbol{\Delta}\mathbf{m}\boldsymbol{v}^T - \mathbf{Y})] + \\ \alpha_f\ \ &[(\phi_h \boldsymbol{f})^T(\phi_h \boldsymbol{f}) + (\phi_v \boldsymbol{f})^T(\phi_v \boldsymbol{f})] + \\ \alpha_m\ \ &[(\phi_h \boldsymbol{\Delta}\boldsymbol{m})^T(\phi_h \boldsymbol{\Delta}\boldsymbol{m}) + (\phi_v \boldsymbol{\Delta}\boldsymbol{m})^T(\phi_v \boldsymbol{\Delta}\boldsymbol{m})] + \\ \alpha_D\ \ &Tr[(\phi_h \boldsymbol{D})^T(\phi_h \boldsymbol{D}) + (\phi_v \boldsymbol{D})^T(\phi_v \boldsymbol{D})] + \\ \alpha_{Dt}\ &Tr[(\boldsymbol{D}\phi_t)(\boldsymbol{D}\phi_t)^T] \end{aligned} \tag{12}$$

where ϕ_h, ϕ_v and ϕ_t are $M \times M$, $N \times N$ and $L \times L$ matrices respectively, used to compute the horizontal and vertical first order differences. Tr stands for the trace of a matrix. α_f, α_m, α_D and α_{Dt} are the priors corresponding to the tissue static magnetization (F), magnetization variation (ΔM) and baseline signal fluctuations, both spatial and temporal (D and Dt).

Equation (12) can be simplified as follows

$$\begin{aligned} &E\ (\mathbf{Y}, \boldsymbol{v}, \boldsymbol{\theta}) = \\ &\tfrac{1}{2}\ Tr[(\boldsymbol{f}\boldsymbol{u}^T + \mathbf{D} + \boldsymbol{\Delta}\mathbf{m}\boldsymbol{v}^T - \mathbf{Y})^T(\boldsymbol{f}\boldsymbol{u}^T + \mathbf{D} + \boldsymbol{\Delta}\mathbf{m}\boldsymbol{v}^T - \mathbf{Y})] + \\ &\quad \alpha_f \boldsymbol{f}^T \Psi \boldsymbol{f} + \alpha_m \boldsymbol{\Delta}\boldsymbol{m}^T \Psi \boldsymbol{\Delta}\boldsymbol{m} + \alpha_D Tr[\boldsymbol{D}^T \Psi \boldsymbol{D}] + \alpha_{Dt} Tr[\boldsymbol{D}\Psi_t \boldsymbol{D}^T] \end{aligned} \tag{13}$$

where $\Psi = \phi_h^T \phi_h + \phi_v^T \phi_v$ and $\Psi_t = \phi_t \phi_t^T$. The minimizers of (14) are the roots of the gradients of E with respect to $\boldsymbol{f}$ and $\boldsymbol{\Delta}\boldsymbol{m}$,

$$\begin{cases} \nabla_{\mathbf{f}} E = (\boldsymbol{f}\boldsymbol{u}^T + \mathbf{D} + \boldsymbol{\Delta}\mathbf{m}\boldsymbol{v}^T - \mathbf{Y})\boldsymbol{u} + \alpha_f \Psi^T \boldsymbol{f} = 0 \\ \nabla_{\boldsymbol{\Delta}\mathbf{m}} E = (\boldsymbol{f}\boldsymbol{u}^T + \mathbf{D} + \boldsymbol{\Delta}\mathbf{m}\boldsymbol{v}^T - \mathbf{Y})\boldsymbol{v} + \alpha_m \Psi^T \boldsymbol{\Delta}\mathbf{m} = 0 \end{cases} \tag{14}$$

Defining the following auxiliary variables

$$\boldsymbol{A_f} = l\boldsymbol{I} + \alpha_f \boldsymbol{\Psi}^T \tag{15}$$
$$\boldsymbol{a_f} = (\mathbf{Y} - \mathbf{D})\boldsymbol{u} \tag{16}$$
$$a_m = \boldsymbol{v^T u} \tag{17}$$
$$\boldsymbol{A_b} = (\boldsymbol{v^T v})\boldsymbol{I} + \alpha_m \boldsymbol{\Psi^T} \tag{18}$$
$$\boldsymbol{a_m} = \mathbf{Y} - \mathbf{D}\boldsymbol{v} \tag{19}$$
$$a_f = \boldsymbol{u^T v} \tag{20}$$

the following simplified version of (14) is obtained,

$$\begin{cases} \mathbf{A_f f} + a_m \boldsymbol{\Delta} \mathbf{m} = \mathbf{a_f} \\ a_f \mathbf{f} + \mathbf{A_m} \boldsymbol{\Delta} \mathbf{m} = \mathbf{a_m} \end{cases} \tag{21}$$

The solution to this system is

$$\begin{cases} \hat{\boldsymbol{\Delta}\mathbf{m}} = (a_f a_m \boldsymbol{I} - \boldsymbol{A_f A_m})^{-1} (a_f \boldsymbol{a_f} - \boldsymbol{A_f a_b}) \\ \hat{\boldsymbol{f}} = (a_m a_f \boldsymbol{I} - \boldsymbol{A_m A_f})^{-1} (a_m \boldsymbol{a_m} - \boldsymbol{A_m a_f}) \end{cases} \tag{22}$$

The stationary point of E w.r.t. $\mathbf{D}$ is computed by finding the roots of the gradient of E (see (14)) with respect to $\mathbf{D}$,

$$\nabla_{\mathbf{D}} E = (\boldsymbol{D} + \boldsymbol{f u^T} + \boldsymbol{\Delta m v^T} - \boldsymbol{Y}) + \alpha_D \boldsymbol{\Psi^T} \mathbf{D} + \alpha_{Dt} \boldsymbol{D \Psi_t} = 0 \tag{23}$$

Defining the following auxiliary variables

$$\boldsymbol{A} = \xi \boldsymbol{I} + \alpha_D \boldsymbol{\Psi^T} \tag{24}$$
$$\boldsymbol{B} = ((1 - \xi)\boldsymbol{I} + \alpha_{Dt} \boldsymbol{\Psi_t}) \tag{25}$$
$$\boldsymbol{C} = -\boldsymbol{Y} - \boldsymbol{f u^T} - \boldsymbol{\Delta m v^T} \tag{26}$$

and considering $\xi = 1/2$, equation (23) can be re-written as follows

$$\boldsymbol{AD} + \boldsymbol{DB} + \boldsymbol{C} = \mathbf{0} \tag{27}$$

which is the well known *Sylvester-Lyapunov* equation, commonly used in control theory [6,7].

The equations (22) and (27) are iteratively computed until convergence is achieved. In each iteration, equation (27) is solved by optimized public routines available in Matlab®[8].

Pair-Wise and Surround Subtractions Implementation

Based on equation (1), pair-wise subtraction was implemented as

$$\boldsymbol{\Delta} \mathbf{m} = \frac{\sum_{i=1,2}^{l-1} Y(i) - Y(i+1)}{l/2} \tag{28}$$

and the surround subtraction as

$$\mathbf{\Delta m} = \frac{Y(1) - Y(2) + \sum_{i=3,2}^{l-1} (Y(i) - \frac{Y(i-1)+Y(i+2)}{2})}{l/2} \tag{29}$$

The slow drift removal for the two subtraction methods was performed using a *high-pass filter*.

3 Experimental Results and Discussion

Tests with synthetic and real data are presented to illustrate the application of the method.

3.1 Synthetic Data

Synthetic ASL data were generated based on a test object with structure similar to the human brain: one axial slice of a real brain mask, segmented into two main regions, *White Matter*(WM) and *Gray Matter*(GM) was used, as shown in Figure 2.

The value of the noise (σ) was set to 1, which adds AWGN with mean value and standard deviation equal to 1, and the magnetization difference ($\mathbf{\Delta M}$) was set to 1 for the GM and 0.5 for the WM. This represents a noise intensity similar to the intensity of the signal. The drift signal ($\mathbf{D}$) is a slow-varying cosine, between -1 and 1, and the background intensity of the image ($\mathbf{F}$) is 10000. The

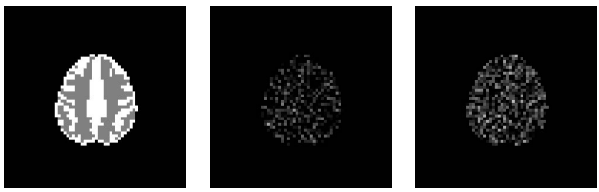

Fig. 2. On the left:synthetic data test object, derived from a real brain mask (64×64 matrix size). The two different regions, colored in white and gray, represent the gray and white matter of the human brain, respectively. Center and right: A control and a labeled acquisition using synthetic data.

proposed algorithm was then tested against the two subtraction methods and the results are displayed in Figure 3. In this Figure, it is evident the reduction of the noise corrupting the image. Image areas with the same intensity are more homogeneous in the image obtained with the proposed algorithm. The edges remain visible and easier to identify. The mean values of Improved SNR (ISNR) and Mean Error (ME) for this test are given in Table 1. In this case, the proposed algorithm obtained an higher value of SNR, approximately $3dB$, and reduced the overall ME by 7%.

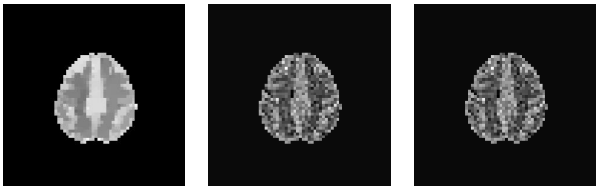

Fig. 3. Perfusion maps obtained with the proposed algorithm using the optimized prior (left), pair-wise subtraction (middle) and surround subtraction (right)

Table 1. Mean values of ISNR and ME for the test realized with the optimized prior

Method	ISNR (dB)	Mean error (%)
Proposed algorithm	16.990	17.807
Pair-wise subtraction	14.026	24.492
Surround subtraction	14.103	24.269

3.2 Real Data

The real data were acquired from an healthy subject, on a 3T Siemens MRI system (Hospital da Luz, Lisboa) using a PICORE-Q2TIPS PASL sequence, with the following parameters: TI1/TI1s/TI2 = 750ms/900ms/1700ms; GE-EPI readout with TR/TE = 2500ms/19ms; 201 repetitions; 9 contiguous axial slices positioned parallel to the AC-PC line, with spatial resolution of $3.5 \times 3.5 \times 7.0$ mm^3 and matrix size $64 \times 64 \times 9$.

The first results obtained are depicted in Figure 4 and were obtained with 200 iterations of the proposed algorithm. They are compared with the images obtained by pair-wise subtraction and surround subtraction.

Fig. 4. Images obtained with the proposed algorithm (left), pair-wise subtraction (center) and surround subtraction (right)

Analyzing Figure 4, the image obtained with the proposed algorithm presents less noise corruption and better distinction of different brain details. The level of smoothing of the image can be controled by changing the values of the priors parameters. Higher values of priors lead to more rigid constraint, hence greater image smoothing.

4 Conclusion

Currently, the perfusion maps from ASL data are usually obtained by image subraction methods. The method here proposed incorporates *a priori* knowledge onto the estimation equation and the averaging procedure is not explicitly done. The trials are put together in a single discrete signal to be used in the estimation of the perfusion map.

As shown in experimental results section, the method proposed outperforms the traditional ones with an increase of 3dB and a reduction of 7% mean squared error. Tested with a set of real data, the images (depicted in Section 3.2) present less noise corruption, as the smoother shape and sharp frontiers suggest.

To improve even further image processing or explore the possibilities, some points of particular interest may be approached in future research, such as automatic prior calculation, so that the algorithm could calculate, on each iteration, the optimal value of the prior.

Validation tests on empirical data are necessary in order to achieve the algorithm validation. In particular, it would be of interest to test the performance of the proposed algorithm in terms of the intra- and inter-subject reproducibility of the perfusion estimates, compared to the most common subtraction methodologies.

References

1. Liu, T.T., Brown, G.G.: Measurement of cerebral perfusion with arterial spin labeling: Part 1. Methods. Journal of the International Neuropsychological Society 13(03), 517–525 (2007)
2. Pollock, J.M., Tan, H., Kraft, R.A., Whitlow, C.T., Burdette, J.H., Maldjian, J.A.: Arterial Spin Labeled MRI Perfusion Imaging: Clinical Applications. Magnetic resonance imaging clinics of North America 17(2), 315 (2009)
3. Aguirre, G.K., Detre, J.A., Zarahn, E., Alsop, D.C.: Experimental Design and the Relative Sensitivity of BOLD and Perfusion fMRI. NeuroImage 15, 488–500 (2002)
4. Buxton, R.B., Frank, L.R., Wong, E.C., Siewert, B., Warach, S., Edelman, R.R.: A general kinetic model for quantitative perfusion imaging with arterial spin labeling. Magnetic Resonance in Medicine 40(3), 383–396 (1998)
5. Moon, T.K., Stirling, W.C.: Mathematical methods and algorithms for signal processing, vol. 13. Prentice Hall, New York (2000)
6. Sanches, J.M., Nascimento, J.C., Marques, J.S.: Medical image noise reduction using the Sylvester–Lyapunov equation. IEEE Transactions on Image Processing 17(9) (2008)
7. Petersen, E.T., Lim, T., Golay, X.: Model-free arterial spin labeling quantification approach for perfusion MRI. Magnetic Resonance in Medicine 55(2), 219–232 (2006)
8. Bartels, R.H.: GW Stewart. Solution of the matrix equation ax+ xb= c [f4]. Communications of the ACM 15(9), 820–826 (1972)