# LANDMARK LOCALISATION IN BRAIN MR IMAGES USING FEATURE POINT DESCRIPTORS BASED ON 3D LOCAL SELF-SIMILARITIES

*Ricardo Guerrero, Luis Pizarro, Robin Wolz, Daniel Rueckert*

Biomedical Image Analysis Group, Department of Computing
Imperial College London, UK

**ABSTRACT**

The identification of anatomical landmarks in the brain is an important task in registration and morphometry. The manual identification and labelling of these landmarks is very time consuming and prone to observer errors, especially when large datasets must be analysed. In this paper we present an approach that describes landmarks based on their intrinsic geometry, rather than their intensity patterns. As the proposed approach moves away from the traditional way to describe landmarks (based on intensities), we show that using this kind of descriptors are well suited for the landmark localisation problem in MR brain images since the intensity information in these images is not quantitative (and intensity normalization is not straight forward). Our results show that for localizing 20 anatomical landmarks in brain MR images, the proposed descriptor performs better in 75% of cases when compared with a Haar feature based classifier and 100% of cases when compared to non-rigid registration.

***Index Terms***— Landmark detection, self-similarity, registration, feature descriptors.

## 1. INTRODUCTION

In recent years several algorithms for landmark localisation have been proposed independently in the medical image analysis, computer vision and machine learning communities, each with specific applications. The detection of landmarks is a crucial step in many medical imaging applications, including registration, shape modelling and morphometry. Approaches to landmark detection can be roughly classified into three main categories: feature-based, structure-based and composite methods.

*Feature-based methods*, in general, rely on a set of local image features located around the point of interest. Landmark detection and localisation is carried out by searching for a set of matching features in unseen images. In [1] the authors used a variation of AdaBoost [2] in combination with Haar-like features to detect human faces in real-time in 2D images using a cascade of classifiers applied to a window sliding over the entire image. In [3], the authors use Marginal Space Learning (MSL) to find matching features within a reduced space in order to detect the left ventricle in 2D MR images. In [4] pixel descriptors are defined based on the self-similarity of the local neighbourhood, the proposed descriptors are used to define and search for objects in different query images. In [5], rotational invariance is introduced to these features to match retinal landmarks.

*Structure-based methods* use prior knowledge about landmark positions in the structure under study, and constrain the landmark search by heuristic rules. The structure is often represented by a learned appearance model, where landmarks are modelled as nodes in a graph and the relation between them are denoted as edges in this graph. Some of the methods that work in this way include Active Shape Models (ASM) [6], Active Appearance Models (AAM) [7], and Constrained Local Models (CLM) [8]. In [9], a combined ASM aimed at finding facial landmarks was proposed to improve accuracy and robustness, in this method two ASM models are build: One for landmarks belonging to the internal part of the face and another for the edge of the face.

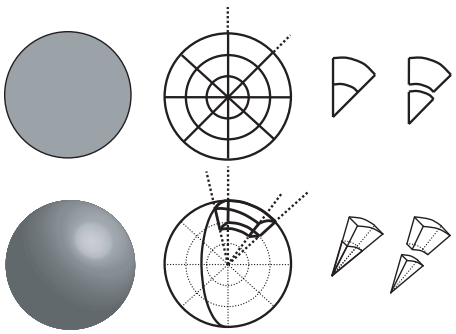


**Fig. 1**. Radial bins used to construct the self-similarity descriptor, in 2D (above) and 3D (below). From left to right, complete area covered by the descriptor, partitioning of the area in radial bins and individual bins.

*Composite methods* integrate as much information as possible about the anatomy the landmarks belong to. In [10] MSL is used to locate landmarks in cardiac MR images where a classifier for each of the landmarks is used to locate candidate positions and then the most probable combination set is chosen. In [11] the authors propose a method that learns landmark characteristics and their spatial relationship from a set of 3D full body human surface scans. This information is then formulated as a Markov network that encodes landmark dependencies. In [12], a boosted regression approach is formulated for medical images. Amongst other applications, the authors test their approach for the detection and localisation of anatomical structures. In [13] the authors add more representational power to the weak learner used in order to improve the performance of [12].

In this work we propose the use of descriptors that define a landmark based on the structural pattern of its neighbourhood. As descriptors we use a modified version of the local self-similarities described in [4]. First, descriptors are found on several training brain MR images. Then when an unseen query image is presented, each training image's descriptor votes on the position of the landmark. Finally, a consensus for the landmark's location is reached by fusing all the predictions made. In our evaluations, the presented approach

This research project was partially funded by CONACyT, SEP and the Rabin Ezra trust.

has been trained on a large dataset of 100 brain MR images from cognitively normal subjects (CN), patients with mild cognitive impairment (MCI) and Alzheimer's disease (AD) from the Alzheimer's Disease Neuroimaging Initiative (ADNI[1]) study database. A different set of 100 images from ADNI is used for testing the proposed approach.

## 2. METHOD

### 2.1. 3D local self-similarity landmark descriptor

In order to characterize landmarks based on their surrounding (local) structures we propose an extended and enhanced self-similarity (SS) descriptor. The local SS descriptor was recently described as an approach for measuring similarity between two visual entities in either images or video [4].

For every pixel in an image a local SS descriptor can be computed. This can be done by computing the similarity between a small square patch around the pixel and every other point (another small square patch) in a larger surrounding circular image region, which results in an internal similarity surface map. This similarity surface is then binned in to a log-polar representation (Figure 1, top row). Each bin is filled with the highest similarity that falls within its supported range. This representation yields three benefits: It compresses the descriptor's length for the pixel. It also accounts for radially increasing affine deformations allowing for invariance to small rotations, shears and scales. Finally, since only the largest similarity gets binned, this accounts for small local non-rigid deformations.

In our work we extend this approach to 3D volumetric images and borrow ideas from the non-local means algorithm [14] to improve the descriptors' robustness. We define spherical regions around each voxel and find matches in a way similar to the approach previously described for 2D. We use sums-of-squared distances (SSD) as a similarity measure, and bin the results according to a log-spherical representation (Figure 1, bottom row). The small descriptor's patches, used to compute the SS, are considered spherical regions surrounding the voxel (typically with radius of 3 or 5 voxels) and each voxel in the sphere is weighted individually using a radially decreasing Gaussian kernel. This has been shown to improve patch similarity measurements [15]. The calculated SSDs are then normalized to form a "correlation volume" $S_q$, that is associated with any voxel $q \in \mathbb{R}^3$, and can be written as:

$$S_q(p) = \exp\left(-\frac{SSD_q(p)}{\max(var_{noise}, var_{auto}(q))}\right) \tag{1}$$

for all $p \in \mathbb{R}^3$ such that $\|p-q\|^2 < \rho^2$. That is for all points within a distance $\rho$ from pivot point $q$. The quatities $(var_{noise}, var_{auto})$ are calculated in the same way as in [4]. The correlation volume, which is defined in Cartesian space, is mapped to a spherical coordinate system $S_q(x,y,z) \rightarrow \tilde{S}_q(r',\theta',\phi')$. Thus, the self-similarity descriptor $SS_q$ is given by the maximum correlation value within each bin $(r_i, \theta_j, \phi_i)$:

$$SS_q(r_i,\theta_j,\phi_i) = \max_{(r',\theta',\phi') \in \mathbb{R}^3} \tilde{S}_q(r',\theta',\phi') \tag{2}$$

where

$$\begin{aligned} r' &\in [r_i, r_{i+1}], & r_i &\in R = \{r_1, ..., r_L\} \\ \theta' &\in [\theta_j, \theta_{j+1}], & \theta_j &\in \Theta = \{\theta_1, ..., \theta_M\} \\ \phi' &\in [\phi_k, \phi_{k+1}], & \phi_k &\in \Phi = \{\phi_1, ..., \phi_N\} \end{aligned} \tag{3}$$

[1] www.loni.ucla.edu/ADNI

and $R$, $\Theta$, $\Phi$ denote the sets of discretised radii, elevation and azimuth angles, each with $L$, $M$ and $N$ values, respectively.

This type of descriptors are specially suited to work with imaging modalities where there is no intensity consistency between images (i.e. MR images) or distinct modalities, since they encode the intrinsic surrounding geometry of the point of interest, and not the intensity distribution. In this way we can characterize anatomical landmarks based on their surrounding structures.

### 2.2. Landmark localisation

With a set of annotated images, where landmark location and landmark descriptors have previously been determined, the landmark can be localized in an unseen images. First, we assume that the brain is in some approximately known orientation and position. Thus a landmark's spatial location is bounded to lie inside a particular volume within the brain. We can reduce the search space to a limited region of interest (ROI), that is defined as a non-zero probability volume of where we expect to find the landmark. This could be done in two different ways: one can define a box where one could expect to see the landmark or one can learn the spatial prior probabilities from the training set (atlases). Here, the latter approach is applied using kernel (parzen window) density estimation. This can be formulated as $P(\mathrm{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_n^d} K_p\left(\frac{\mathrm{x}-\mathrm{x}_i}{h_n}\right)$, where x are the 3D voxel coordinates $(x,y,z)$, $\mathrm{x}_i$ are all the landmark's coordinates, $K_p(\cdot)$ is the window function or the kernel in a $d$-dimensional space such that $\int_{\Re^d} K_p(\mathrm{x})d\mathrm{x} = 1$, $n$ is the number of observations and $h_n > 0$ is the bandwidth parameter that corresponds to the width of the kernel. The kernel function $K_p(\cdot)$ is modelled as a Gaussian function.

In the defined search space we calculate a descriptor for every voxel contained in this ROI, and then find the best possible match for each of the training descriptors. Each training template gives an estimation (vote) for the spatial localisation of the landmark. Each vote will be given a certain weight $w$. For an unseen image $i$, we estimate the localisation $L_i^l$ of the landmark $l$ by fusing all votes with the formula:

$$L_i^l = \frac{\sum_j I_{i,j}^l \cdot w_{i,j}^l}{\sum_j w_{i,j}^l}\ , \tag{4}$$

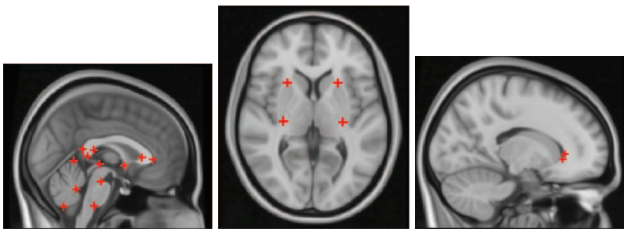
where $I_{i,j}^l$ represents the estimated localisation of the landmark $l$ in the image $i$ as voted by template $j$. In this work we explore several strategies to select the weights $w_{i,j}^l$:

- Simple Average Vote (SAV): all votes have the same weight.
- Sparsity (S): weighting according to the sparsity of the likelihood output (the more sparse the more reliable), cf. (5).
- Normalized Sparsity (NS): same as sparsity, but with the values normalized to the range [0,1].
- Descriptor Similarity (DS): between template and test descriptors.
- $k$ most Similar Descriptors (kSD): only the $k$ most similar templates are considered.

As a sparsity metric we used a measure based on the relationship between the $L_1$ and $L_2$ norms (as described in [16]):

$$S(\mathbf{x}) = \frac{\sqrt{n} - (\sum |x_i|)\sqrt{\sum x_i^2}}{\sqrt{n}-1} \tag{5}$$

where $n$ is the dimensionality of $\mathbf{x}$.



(a) (b) (c)

**Fig. 2**. Anatomical landmarks in the MNI152 atlas. (a) Splenium and genu of corpus callosum, superior and inferior tip of the cerebellum, fourth ventricle, anterior and posterior commisure, and superior and inferior aspect of the pons. (b) Superior and inferior tip of the putamen (left and right). (c) Anterior and inferior tip lateral ventricle (only left side shown).

## 3. EXPERIMENTS AND RESULTS

The images that were used to evaluate the proposed method were obtained from the ADNI database [17]. In the ADNI study, brain MR images were acquired at regular intervals after an initial baseline scan from approximately 200 cognitively normal older subjects (CN), 400 subjects with mild cognitive impairment (MCI), and 200 subjects with early Alzheimers disease (AD). In this work, we used a subset of 1.5T T1-weighted baseline images of 100 randomly chosen subjects for training and another 100 randomly chosen subjects for testing. In both (training and testing) datasets there are 24 AD, 48 MCI and 28 healthy subjects, to faithfully represent the full ADNI dataset. All brain images were skull stripped and affinely aligned to the MNI space. For both the training and testing datasets a total of 20 landmarks (Table 1) were manually selected by an expert observer using three orthogonal views.

Two other commonly used methods for landmark localisation were used as a comparison to the proposed one: (a) Haar feature boosted classifier (HF) and (b) non-rigid image registration (REG). ***Boosted Haar features***. A classifier for each landmark was built using a variation of the Viola-Jones face detector [1]. 3D Haar-like features are calculated within a cuboid region in 3D space from MR brain images. Boosting [2] is used to perform classification based on all calculated Haar-like features using Quadratic Discriminant Analysis (QDA). A weak classifier $h_j(\mathbf{x}_i)$ consists of a feature $f_j$, a threshold $\theta_j$ and a parity $p_j$ indicating the direction of the inequality sign: $h_j(\mathbf{x}_i) = \pi \{p_j f_j(\mathbf{x}_i) < p_j \theta_j\}$. The split of the data produced by the features only needs to perform slightly better than random. At each round of boosting the algorithm picks the best weak classifier, reweighs the training examples and repeats the process for $T$ rounds, or until an early exit criteria is met.

***Non-rigid image registration***. Intensity-based image registration establishes dense point correspondences across images by computing a transformation that map points from one image $I$ to corresponding points in a second image $J$. The transformation $\mathbf{T}$ applied to a point x in $I$, represented by the column vector $\mathrm{x} = (x, y, z)$, yields a transformed point $\mathrm{x}' = (x', y', z')$ in $J$, such that, $\mathrm{x}' = \mathbf{T}(\mathrm{x})$. Intensity-based image registration can be used to propagate the annotation from a reference image to a new image. In this case the annotated reference image acts as a template. Hence, landmarks that are annotated in the template can be propagated to new images. In this work we used the method proposed in [18].

From the training set we calculate 3D SS descriptors, at the landmark's position, that define the surrounding structure. To calculate the mentioned descriptors we used typically small spherical kernels with a radius of three voxels, that are weighted using a second Gaussian kernel (same size) with a sigma of three, and a larger correlation sphere of radius 10. The similarity results were binned in a partitioned sphere with four radial intervals, five inclinations and 10 azimuth angles (see Figure 1, bottom row for visualization). We tested the performance of classifiers built from different numbers of training images (10, 20, 40, 60, 80 and 100)

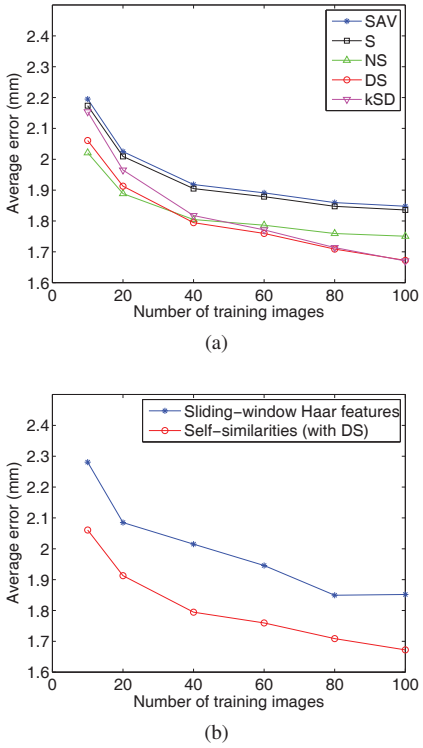



**Fig. 3**. (a) Performance of the proposed classifier built with different amounts of training images and using different fusion techniques. (b) Comparison of the proposed method vs SW.

During testing, we look for the best matches of the training descriptors in the test image and then form a consensus on the final landmark location by fusing the results using several techniques. Figure 3 (a) shows the average (over the 20 landmarks shown in Table 1) landmark error. Using descriptor similarity (most robust results, Figure 3 (a)), a comparison with the boosted Haar features classifier, which is a commonly used approach used in the computer vision and biomedical imaging communities, was carried out (also using the 20 landmarks from Table 1). We show that using less training images we can obtain the same level of accuracy, which is a highly desired quality in the medical image analysis community (as annotated images are scarce and expensive in terms labour involved). Using the same amount of images the proposed method achieves better results. Figure 3 (b) better illustrates this. Another desirable feature of the proposed method is that it does not require any intensity normalization step, as it is required by intensity based approaches (such as SW). The boosted Haar feature classifier was built as a monolithic combination of 100 boosted features, that combine into a final strong classifier.

| | Mean | | | Std. dev. | | |
|---|---|---|---|---|---|---|
| **L** | SS | HF | REG | SS | HF | REG |
| 1 | **1.29** | 1.79 | 3.95 | 0.99 | 1.07 | 1.43 |
| 2 | **1.27** | 1.66 | 2.1 | 0.88 | 0.96 | 0.89 |
| 3 | 2.28 | **1.81** | 2.31 | 1.54 | 0.82 | 1.31 |
| 4 | **1.08** | 1.87 | 1.73 | 0.78 | 1.08 | 1.01 |
| 5 | 2.56 | **1.32** | 1.47 | 2.17 | 0.73 | 0.64 |
| 6 | **1.06** | 1.12 | 2.79 | 0.77 | 0.77 | 1.26 |
| 7 | **1.26** | 1.45 | 1.7 | 2.29 | 0.72 | 0.85 |
| 8 | 3.92 | **2.83** | 2.99 | 1.02 | 1.75 | 1.64 |
| 9 | **0.83** | 1.19 | 5.57 | 0.78 | 0.80 | 2.7 |
| 10 | **2.24** | 2.43 | 4.36 | 1.16 | 1.41 | 1.81 |
| 11 | **1.78** | 2.14 | 2.48 | 1.14 | 1.27 | 1.29 |
| 12 | 2.28 | **2.25** | 3.53 | 1.33 | 1.13 | 1.78 |
| 13 | **1.90** | 2.28 | 2.79 | 1.24 | 1.23 | 1.43 |
| 14 | **0.67** | 1.16 | 1.05 | 0.59 | 0.59 | 0.42 |
| 15 | **0.64** | 0.69 | 1.85 | 0.31 | 0.59 | 0.48 |
| 16 | 2.87 | **2.39** | 3.71 | 2.02 | 1.89 | 1.68 |
| 17 | **1.31** | 2.22 | 3.67 | 0.89 | 1.24 | 1.72 |
| 18 | **1.14** | 1.78 | 3.65 | 0.71 | 0.98 | 1.73 |
| 19 | **1.76** | 2.80 | 4.44 | 1.34 | 1.35 | 2.07 |
| 20 | **1.27** | 2.20 | 4.01 | 0.78 | 1.11 | 1.79 |

**Table 1**. Accuracy of the proposed method on the ADNI database, for the Splenium of corpus callosum (outer aspect, inferior tip and inner aspect (1,2,3)), Genu of corpus callosum (outer and inner aspect (4,5)), Superior and inferior aspect of pons (6,7), Superior and inferior aspect cerebellum (8,16), Fourth ventricle (9), Putamen posterior and anterior (10,11)(left), (12,13)(right), Anterior and posterior commissure (14,15), Anterior tip of lateral ventricle (left and right) (17,18), Inferior tip of lateral ventricle (left and right) (19,20).

Table 1 shows the result of individual landmark localisation error. A five-fold cross validation of the method was carried out in order to asses the results and ensure reproducibility (the average of the five tests is shown, with a variability among tests of $\sim$0.15mm).

## 4. DISCUSSION AND CONCLUSIONS

We have proposed a method that characterizes landmarks in brain MR images using a local 3D SS geometric descriptor (not intensity dependent), around the vicinity of the landmark. Using this landmark representation we can search for and obtain votes on where the landmark is located. Several vote fusion strategies are tested. Also, prior knowledge of the spatial distribution of the landmarks was used to reduce the search space. Our results show that the proposed method significantly outperforms a 3D boosted Haar features classifier and B-spline registration, in the landmark localisation task.

Anatomical landmarks of the brain have a constrained position within the brain and to the relative position of other landmarks. The inferior tip, inner and outer aspect of the splenium of the corpus callosum, are part of the same anatomical structure, so their locations with respect to each other are somewhat correlated. A graphical model of the joint spatial distribution probabilities of related landmarks, in the form of a Markov random field, could be used to represent spatial dependencies. This path was explored in order to assess if we could improve on the accuracy of the SW approach. However, we found that this additional step did not offer any significant benefit. This could be attributed to the fact that even though the landmarks were manually annotated by an expert, there is still some intrinsic error associated with the annotations (due to partial volumes and human errors). We also hypothesize that the accuracy achieved by the method could be higher than the spatial relation between different landmarks (even if they are from the same structure) due to small non-linear deformations.

## 5. REFERENCES


[1] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[2] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[3] Y. Zheng, A. Barbu, B. Georgescu, et al., "Fast automatic heart chamber segmentation from 3D CT data using marginal space learning and steerable features," in *ICCV*, 2007, pp. 1–8.

[4] Eli Shechtman and Michal Irani, "Matching local self-similarities across images and videos," in *CVPR*, 2007.

[5] Y. Zheng, A.A. Hunter, III, J. Wu, et al., "Landmark matching based automatic retinal image registration with linear programming and self-similarities," in *IPMI*, 2011, pp. 674–685.

[6] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models - their training and application," *Computer Vision Image Understanding*, vol. 61, pp. 38–59, 1995.

[7] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, pp. 484–498.

[8] D. Cristinacce and T. Cootes, "Feature detection and tracking with constrained local models," in *BMVC*, 2006, pp. 929–938.

[9] D. Zhou, D. Petrovska-Delacretaz, and B. Dorizzi, "Automatic landmark location with a combined active shape model," in *BTAS*, 2009, pp. 1–7.

[10] X. Lu, B. Georgescu, Marie-Pierre Jolly, et al., "Cardiac anchoring in MRI through context modeling.," in *MICCAI*, 2010, pp. 383–390.

[11] Z. B. Azouz, C. Shu, and A. Mantel, "Automatic locating of anthropometric landmarks on 3D human models," in *3DPVT*, 2006, pp. 750–757.

[12] S. K. Zhou, B. Georgescu, X. S. Zhou, and D. Comaniciu, "Image based regression using boosting method," in *CVPR*, 2005, vol. 1, pp. 541–548.

[13] S. K. Zhou, J. Zhou, and D. Comaniciu, "A boosting regression approach to medical anatomy detection," in *CVPR*, 2007.

[14] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *SIAM Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[15] L. Pizarro, P. Mrázek, S. Didas, et al., "Generalised nonlocal image smoothing," *International Journal of Computer Vision*, vol. 90, no. 1, pp. 62–87, 2010.

[16] P.O. Hoyer and P. Dayan, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.

[17] S.G. Mueller, M.W. Weiner, L.J. Thal, et al., "The Alzheimer's Disease Neuroimaging Initiative," *Neuroimaging Clinics of North America*, vol. 15, no. 4, pp. 869–877, 2005, Alzheimer's disease: 100 years of progress.

[18] D. Rueckert, L. I. Sonoda, C. Hayes, et al., "Nonrigid registration using free-form deformations: Application to brest MR images," .