

# A Methodology for Validating a New Imaging Modality with Respect to a Gold Standard Imagery: Example of the Use of 3DRA and MRI for AVM Delineation

Marie-Odile Berger<sup>1</sup>, René Anxionnat<sup>1,2</sup>, and Erwan Kerrien<sup>1</sup>

<sup>1</sup> LORIA/INRIA Lorraine, France

<sup>2</sup> CHU Nancy, France

berger, anxionna, kerrien@loria.fr

**Abstract.** Various medical treatments require an accurate determination of the shape of a considered anatomic structure. The shape is often recovered from several delineations performed on a two-dimensional gold standard imagery. Using true 3D imagery is attractive to supplement this gold standard. However, before using 3D modalities in clinical routine, it must be proved that these modalities are well suited to the delineation task. We propose in this paper a methodology for validating a new imaging modality with respect to a reference imagery.

## 1 Introduction

Numerous clinical treatments, such as radiotherapy, require an accurate definition of the true three-dimensional size and shape of a given anatomic structure. Depending on the pathology, a reference imagery, so-called Gold Standard imagery, is classically used by physicians for the definition of the target. Other imaging modalities, especially emerging modalities, may complement this reference imagery to better understand the architecture of the targeted organ. However, prior to using these modalities for the delineation task, it must be proved that they bring more than or at least equal information to the reference imagery.

This paper focuses on the case where the reference imagery is two-dimensional. This is especially true for vascular imaging (cardiology, neurology,...) where 2D angiographic (DSA) images are still the gold standard. The difficulty of obtaining a good approximation of the shape from its delineation in a small number of views is well known. For this reason, several groups have reverted to using true three dimensional data from either computed tomography (CT) or magnetic resonance imaging (MRI). Unfortunately, these 3D modalities present with a lower spatial and temporal resolution than conventional DSA. A better identification of vascular structures in these images is therefore not guaranteed. This important problem of validating the use of a new imaging modality using standardized protocols has been little addressed in the literature.

In this paper, we propose a methodology for validating a new 3D image modality with respect to a Gold Standard imagery in the clinical context of the delineation task of a target. This methodology is applied to validate the use of 3D image modalities to delineate cerebral arteriovenous malformation (AVM). The clinical context of this application is depicted in section 2. The overall methodology is described in section

3. Application of our framework to the use of 3D rotational angiography (3DRA) and magnetic resonance imaging (MRI) for AVM delineation is shown in section 4.

## 2 Clinical context

The efficiency of our methodology is illustrated by the clinical problem of cerebral arteriovenous malformation delineation. A cerebral arteriovenous malformation (AVM) is defined as a localized arteriovenous shunt consisting of a tangle of capillaries and veins, also called nidus. The treatment of complex AVMs is classically a two stage process. Embolization or endovascular treatment is first performed. It reduces the size of the AVM but is usually not sufficient to cure it completely. This step is then followed by a stereotactic irradiation of the remnant through the intact skull. Therefore, accurate definition of the target is of crucial importance for the efficacy of the treatment.

For many years, 2D angiography has been the gold standard for the diagnosis and the delineation of the AVMs: the AVM is delineated in at least two calibrated views and the volume of interest is limited by the intersection of X-ray cones defined by each delineation [7]. This method only provides an approximation of the shape since only a small number of views is allowable (from 2 to 5 views), due to clinical and technical limitations. This procedure overestimates the volume, by an amount depending on the shape of the AVM and on the considered viewpoints (Fig. 1).

For these reasons, several groups have attempted to supplement conventional angiography with true 3D data for defining the vascular lesion: MRI helps to exclude vulnerable tissues, MR angiography is a fast, non invasive 3D imaging that provides information on vascular and parenchymal brain anatomy important for optimal dose planning [6, 5]. 3DRA has been also used for AVM delineation with promising results [3]. In all these studies, the evaluation often rests on a visual inspection of the delineations and is thus partly subjective. The volume of the target derived from different modalities also often grounds the comparison: because the AVM size is somewhat overestimated in angiography and to avoid irradiation of safe tissues, methods leading to the smallest volumes are preferred. Unfortunately, though unnecessary irradiation of adjacent safe tissue is important to avoid, the reduction of the volume does not necessarily prove that the complete detection of the nidus has been achieved.

All these modalities have a lower temporal resolution than conventional angiography. Though the use of a 3D modality theoretically allows us to increase the accuracy on the target, it must be proved that the information available in the volume effectively allows for a complete detection of the nidus. Indeed, nidus are identified from the arterial phase, just as the draining vein begins to fill. As all the considered 3D modality are acquired during the whole progression of the opaque mixture (around 5 seconds for 3DRA, much more for MRA), it is not clear whether all the parts of the nidus will be appearant in the 3D images, in particular those presenting with a fast blood flow.

If the development of MR and 3DRA will no doubt improve the evaluation of the AVM in the near future, a meticulous assessment of both the target delineation and the treatment outcome has to be performed. This problem has not yet been studied in the medical community and we propose in the next section a general methodology to this aim.

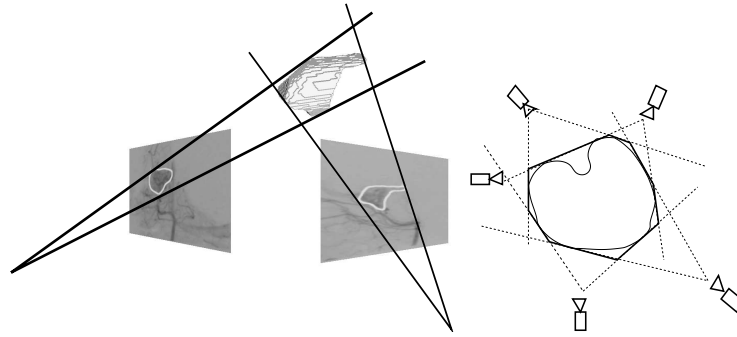


Fig. 1. AVM reconstruction using conventional angiography.

### 3 Methodology

#### 3.1 Evaluating the relevance

This section addresses the general problem of validating a 3D modality with respect to a 2D reference imagery. The shape inferred from the 3D modality (which is called *new 3D shape* in the following) must be proven better than or at least as good as the shape inferred from the reference imagery. Phantoms are commonly used to compare imaging modalities. However, no phantom currently enables a realistic simulation of the complexity of the vascular phenomena occurring in the nidus.

In the absence of ground truth, the only standards available to assess the quality of a new 3D shape are the delineations produced by expert observers in the 2D reference imagery. Comparison can be achieved either directly in 2D - by comparing the projection of the new shape to the expert curves in each reference image - or in 3D. In this latter case, the shape delineated in the 3D modality must be compared to the 3D shape recovered from expert delineations, which is only a rough approximation of the actual shape due to the small number of considered views. In addition, delineating a vascular structure is a difficult task especially when some parts of the boundary are hidden by vessels superposition. Such difficulties prevent from delineating the true boundary and lead to inaccurate reconstruction. Hence, direct comparison in 3D is not well founded.

We therefore propose to validate the use of a 3D imaging modality by a direct comparison of the projected shape with experts delineations. The problem of evaluating a curve with respect to expert delineations was previously addressed in [2]. Their methodology strongly relied on the existence of a metric between curves which is supposed to reflect their similarity. In our clinical application, preliminary investigations indicated that manual delineations suffer from observer bias and large inter and intra-observer variability [1] (Fig. 2): large consensus areas alternate with high variability areas, preventing us from defining a reliable metric between curves and to use [2]. Fortunately, a statistical model of the variability derived from the set of expert curves can be computed [4]: it is then possible to learn what are plausible variations in the shapes and what are not, allowing us to check if a projected shape is compatible with expert delineations.

### 3.2 Building a statistical shape model

Statistical shape models were introduced by Cootes [4] in order to capture variability in shape. Such a model is built by analyzing shape variations over a training set of 2D curves  $C_i$  ( $1 \leq i \leq N$ ) drawn by experts. Dense correspondences were established using a semi-interactive interface. A curve  $C_i$  is represented with a set of  $n$  2D points  $C_i = (x_i^1, \dots, x_i^n)$ . Let  $\bar{C} = \frac{1}{N} \sum_{i=1}^N C_i$  be the mean of the curves, computed from point-to-point correspondences, and  $S = \frac{1}{N-1} \sum (C_i - \bar{C})(C_i - \bar{C})^t$  be the covariance of the data. Following Cootes [4], a linear eigen-model can be built through a principal component analysis (PCA) that provides the  $t$  principal modes of variation. These modes are represented by the  $N \times t$  matrix  $P$  of eigenvectors of  $S$  corresponding to the  $t$  largest eigenvalues. A particular curve  $C$  is then approximated by:

$$C \approx C' = \bar{C} + Pb \quad \text{where } b = P^t(C - \bar{C}) \quad (1)$$

The squared error on this approximation is therefore:

$$r^2(C) = \|C - C'\|^2 = (C - \bar{C})^t(C - \bar{C}) - b^t b \quad (2)$$

The value of  $t$  is classically chosen so that the model represents a suitable proportion of the variation in the training set (95%).

### 3.3 A framework for validation

The relevance of a 3D imaging modality is estimated by comparing the shape recovered with this modality to the statistical 2D shape model built from a set of experts in the same viewing conditions. A two stage process is proposed:

1. **Building the database and the statistical models.** A set of patients is considered. The standard clinical technique is used and 2D calibrated images are acquired using the reference modality.  $N_e$  experts delineate the shape in each view and a statistical model of the 2D shape is built from these expert delineations for each view.
2. **Validating the 3D modality.** The projection of the new 3D shape is computed with the same viewing parameters as used for the acquisition of the reference images. Each obtained 2D shape is then tested to check if it can be considered as an instance of the statistical model built from the experts.

The second stage requires to decide how a new delineation matches the model. A first measure of the quality of fit is classically deduced from the PCA:

$$fit_1(C) = \sum_{i=1}^{i=t} \frac{b_i^2}{\lambda_i} \quad (3)$$

where  $\lambda_i$  is the  $i$ -th largest eigenvalue in the PCA and  $b = (b_i)$ . This criterion only measures the distance of the shape from the mean along the modes of the model but it does not consider the distance of the delineation from the linear model given by the residual in eq. 2.

To cope with this problem, a complementary measure must be defined to decide if this residual is compatible with the delineations of the experts  $\{C_e\}_{e \in Experts}$  for a given reference view. Any measure which involves  $\{r(C_e)\}_{e \in Experts}$  is not a convenient measure of the performance of the model because it comes down to estimate the model from data which were used to build the model. A better measure is given by cross-validation and especially miss-one-out experiments: for each expert  $e$ , a reduced PCA model is built from all the data except  $C_e$ . The residual of this reduced PCA is denoted  $r_e$ .  $r_e^2(C_e)$  is thus a good measure of the compatibility of a given expert with the others. Estimating the compatibility of a new curve  $C$  is then better achieved by comparing  $r_e^2(C)$  with  $r_e^2(C_e)$  on the set of experts. Therefore, we propose to statistically assess a delineation  $C$  using:

$$fit_2(C) = \frac{\sum_{e \in Expert} r_e^2(C)}{\sum_{e \in Expert} r_e^2(C_e)} \quad (4)$$

The complete measure of fit is then defined as the sum  $fit(C) = fit_1(C) + fit_2(C)$ . In order to design a statistical test for acceptance/rejection of new delineations, the distribution of the measure of fit must be studied on the particular clinical application. Using bootstrap techniques [4], we found experimentally that the distribution of fit for AVM detection could be modeled with a  $\chi$  square distribution with  $t$  degrees of freedom.

## 4 Validating 3D modalities for AVM delineation

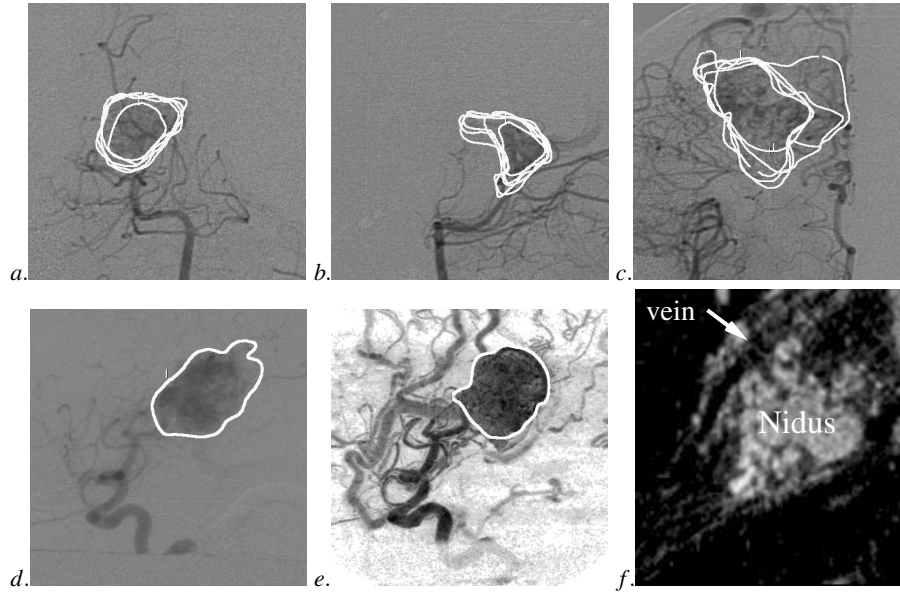
### 4.1 Clinical material

Twelve patients with AVM were treated using embolization and radiotherapy at Nancy University Hospital (France). According to the complexity of the AVM shape, between 2 to 5 2DSA views were considered for delineation (antero-posterior, lateral and oblique views). Overall, 45 2DSA images were available for this study. In addition, the 3DRA and the MRI volumes were available for each patient. These images were acquired in stereotactic conditions, allowing for 2DSA/3DRA/MRI image registration.

The AVMs considered in this study were complex pathological cases. In order to assess a possible bias due to inner medical difficulties, each AVM was labelled with A or B according to the complexity of the delineation task: type A is for fairly difficult AVMs (5 patients) whereas type B is for very difficult AVMs (7 patients).

Three experts in interventional neuroradiology from Nancy University Hospital (France) and Karolinska Institute (Stockholm, Sweden) were involved in this study. They manually delineated the AVMs twice on the 45 2DSA images.

**3D Rotational angiography.** 3DRA is a new emerging imaging modality [8] and physicians are used to analyzing this volume using 3D rendering techniques, such as MIP (Maximum Intensity Projection). However, they are unable to delineate the shape directly in 3D because the appearance of one slice does not allow them to identify the nidus properly (Fig.2.f). Hence, for each patient, MIP images of the 3DRA were generated with the same viewing parameters as the original 2DSA images. An example of a MIP view, together with the original 2DSA image are shown in Fig. 2.d,e. The same



**Fig. 2.** Angiographic views of AVMs

**First row:** Examples of AVM delineation realized by the experts for 2D angiographic views for type A (a and b) and type B (c) AVMs. **Second row:** Example of (d) an original 2DSA view and (e) the MIP view computed from 3DRA along with the expert delineations. (f) is a slice of the 3DRA volume through the AVM.

three experts were asked to delineate the AVMs in these MIP views. These contours were then tested using the statistical model built from the six experts delineations.

**MRI modality.** Though used by several groups for AVM delineations [6, 5], the clinical superiority of MRI over DSA has never been proved. These previous experimental studies concerned small patient series. They suggest that MR can be of great value in widening radiosurgical indications to large AVM but seems inadequate for previously embolized AVMs. Our results will partly confirm these impressions. Our three experts delineated the AVMs slice by slice to get the 3D shapes. These shapes were then projected onto the reference views.

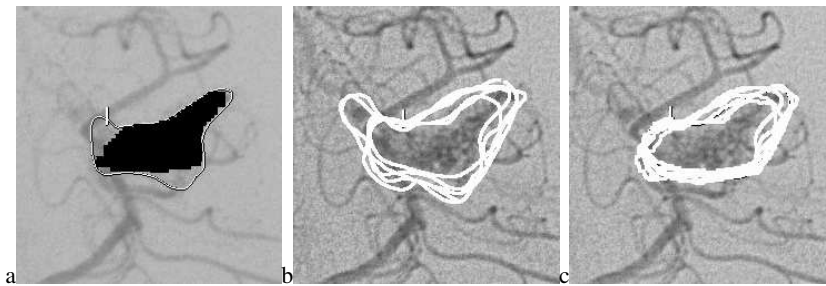
## 4.2 Results

### Validating the use of 3DRA

Results are summarized in table 1 which indicates the acceptance rate of the MIP contours on the whole database (type A and B) and on type A AVMs. As in section 3.3, these results are compared to the acceptance rate of each expert curve when compared to the statistical model built from the other experts.

For type A AVMs, the results are equivalent for DSA and for 3DRA, proving that 3DRA can be used for nidus delineation. On the other hand, delineations are not equivalent for type B AVMs. Differences mainly originate in the large size of type B AVMs and in multiple vessels superposition in 2D, making the true boundary very difficult to assess in a single 2D view. Fortunately, the use of several views often allows us to get a fair target reconstruction despite these errors: an erroneous part delineated in one view is often removed by a contour drawn in another view due to cone intersections (Fig. 3.a). Hence, a more coherent database is given by the reprojection of the 3D shape recovered from expert delineations onto the reference images (see Fig. 3).

Results of the evaluation of 3DRA against this new database are given in Table 2. The global acceptance rates are nearly similar for both modalities. Though the imaging set up for large AVMs must certainly be improved to get a better image quality, these results still prove that 3DRA can be used for AVM delineation.



**Fig. 3.** (a) Reprojection of the recovered 3D shape on a reference image in black along with the expert delineation. (b) The set of expert curve delineations. (c) The new database using the reprojection of the 3D shape recovered from expert delineations for the same view.

### Validating the use of MRI modality

The validation framework was also applied to assess MRI images. The contours were considered in our statistical framework with the same confidence as for 3DRA (95 %). Table 2 shows that globally, only 44% of the contours were accepted. This rate improved to 60% for type A AVMs. It must be noted that the difficulty of the delineation task depends on the considered image modality. Only 3 AVMs were considered as fairly difficult (type A) for MRI delineation. They were also classified type A in DSA. For type A AVMs, the volume recovered from DSA and MRI are quite similar. On the contrary, the volume computed for type B AVMs are significantly larger than those computed from DSA. A careful examination of the target proves that previously embolized areas are included in the MRI delineations. The large number of previously embolized patients in our dataset (10/11) thus explains these poor results. This study proves that MRI cannot be used alone for AVM delineation for previously embolized AVMs. Though this study should be extended to incorporate more non embolized AVMs, these results also tend to prove that the delineation of the nidus in MRI and DSA presents a similar level of difficulty for simple non embolized AVMs.

Contours	global acceptance rate (type A & B)	acceptance for type A AVMs
2DSA	77.5%	77%
3DRA	64%	76%

**Table 1.** Acceptance rate of 3DSA contours compared to the acceptance rate of 2DSA contours computed by cross validation.

Contours	global acceptance rate (type A & B)	acceptance for type A AVMs
2DSA	83%	85%
3DRA	77%	83%
MRI	44 %	60 %

**Table 2.** Acceptance rate of 3DSA and MRI contours with the corrected database.

## 5 Conclusion

In this paper, we have proposed a methodology for validating a new imaging modality with respect to a reference imagery. We have applied this protocol to prove that 3DRA can be used for AVM delineation. We believe that the objective and quantitative evaluation of new image modalities is an important step toward their acceptance in clinical use. We are currently investigating the use of 3DRA to improve AVM delineation through the definition of virtual 2D views which can be generated from the 3DRA volume for any position. Hence, the physician could potentially use an arbitrary number of views to delineate the AVM without further dose exposure for the patient.

## References

1. R. Anxionnat, M-O. Berger, E. Kerrien, S. Bracard, and L. Picard. Intra- and inter-observer variability in the angiographic delineation of brain arterio-venous malformations (AVM). In *CARS, London (UK)*, 1297–1298, 2003.
2. V. Chalana and Y. Kim. A methodology for evaluation of boundary detection algorithms on medical images. *IEEE Trans. Med. Imag.*, 16(5):642–652, 1997.
3. F. Colombo, C. Cavedon, P. Francescon, L. Casentini, U. Fornezza, L. castellan, F. Causin, and S. Perini. Three-dimensional angiography for radiosurgical treatment planning for AVMs. *Journal of Neurosurgery*, 98:536–543, 2003.
4. T. Cootes, C. Page, C. Jackson, and C. Taylor. Statistical grey-level models for object location and identification. *Image and Vision Computing*, 14:533–540, 1996.
5. W. Guo. Application of MR in Stereotactic Radiosurgery. *JMRI*, 8:415–420, 1998.
6. Douglas Kondziolka, Dade Lunsford, Emanuel Kanal, and Lalith Talagala. Stereotactic magnetic resonance angiography for targeting in arteriovenous malformations radiosurgery. *Neurosurgery*, 35:585–591, 1994.
7. M. Soderman, B. Karlsson, L. Launay, B. Thuresson, and K. Ericson. Volume measurement of cerebral arteriovenous malformations from angiography. *Neuroradiology*, 42:697–702, 2000.
8. K. Wiesent, K. Barth, N. Navab, P. Durlak, T. Brunner, O. Schuetz, and W. Seissler. Enhanced 3-d- Reconstruction Algorithm for C-Arm Systems Suitable for Interventional Procedures. *IEEE Transactions on Medical Imaging*, 19(5):391–403, 2000.