A fast hierarchical traversal strategy for multimodal visualization

Maria Ferré
URV Computer Science and Math. Dept.

Anna Puig

Dani Tost
CREB Center of Biomedical Engineering Research, ETSEIB, UPC

ABSTRACT

In the last years there is a growing demand of multimodal medical rendering systems able to visualize simultaneously data coming from different sources. This paper addresses the Direct Volume Rendering (DVR) of aligned multimodal data in medical applications. Specifically, it proposes a hierarchical representation of the multimodal data set based on the construction of a Fusion Decision Tree (FDT) that, together with a run-length encoding of the non-empty data, provides means of efficiently accessing to the data. Three different implementations of these structures are proposed. The simulations results show that the traversal of the data is fast and that the method is suitable when interactive modifications of the fusion parameters are required.

Keywords: Volume Rendering, Decision Tree, Run-length encoding, Multimodal rendering.

1. INTRODUCTION

Multimodal rendering is an important requirement of medical imaging systems, since the correlation of images from different modalities provides important clues on the presence of pathologies and dysfunctions.

Current multimodal rendering methods are mostly 2D and perform the merging on the basis of equivalent slices of the different modalities.[1] A major drawback of these methods is that they do not provide enough clues on the spatial relationships between the features shown in the different modalities. There are three main approaches of 3D multimodal rendering[2]:

- extracting isosurfaces from the different datasets and rendering them simultaneously;
- extracting isosurfaces from one or more modalities and performing a hybrid rendering of the surface models and volume data from another modality;
- simultaneous direct volume rendering of the datasets.

Although the two first approaches are faster, they present as a major drawback their lack of flexibility, since isosurfaces must be extracted first, in a pre-process. We herein focus on the latter approach.

The core of direct multimodal rendering is the fusion of the datasets. Fusion consists of merging data from the different datasets according to a given interpolation scheme. A special case of merging is when only one of the values is selected for each sample and used in the rendering pipeline. Another useful merging is a linear combination of the data. In addition, fusion can be done at different steps of the rendering pipeline: previous to the classification, using property values or property and gradient values (Property Fusion, or PF), post classification using material identifiers (Material Fusion or MF) or at the end of the rendering pipeline merging colors (Color Fusion or CF). Finally, fusion can be done at different

M. Ferre, Send correspondence to Avda. Paisos Catalans, 26
43007 Tarragona
Spain: E-mail: mferre@etse.urv.es, Telephone: (034) 977559688, Fax: 934016050

levels depending on the sampling scheme of the rendering algorithm: (i) at points along viewing rays, in raycasting, (ii) at voxels, in projective approaches such as sorted traversal and splatting of the volume or shear-warp, and (iii) at pixels of color planes in rendering algorithms that slice the volume with planes perpendicular to the viewing direction. In,[3] these different fusion processes and rendering approaches have been analyzed and compared. In this work, it has been shown that the fusion based on property values (PF) provides a finer control on the results. Moreover, this modality is the fastest, because it detects and skips samples that do not match the fusion criteria early in the rendering pipeline. The main drawback of PF is that it is not very user-friendly, because it requires many parameters to be specified and thus, special interface widgets should be designed to make it fully useful.[4]

Another important aspect of multimodal rendering is the alignment of datasets. When the datasets are not aligned, it is necessary to compute the geometrical transformations that reference the local coordinate systems of the different datasets in a common frame. This process is called registration[5].[6] Moreover, after the registration, a resampling process may be necessary in order to set the different data in the same coordinate system and the same resolution. This resampling step may introduce errors and it may increase the size of the final data. However, it is necessary in many rendering strategies.

Ray casting[7] is particularly well suited for non-aligned data because the geometrical transformations from global to local coordinate systems can be applied locally for each sample, although at an extra CPU time cost. According to,[8] the expected performance of other methods such as splatting (for high pixel/voxel ratios), 3D texture mapping and shear-warp is higher. Besides, the key for multimodal rendering is the speed of the visualization, as physicians need to modify the fusion parameters interactively. Image-aligned splatting as well as 3D texturing and shear-warp can be done on non-aligned datasets by merging final color planes. However, this restricts the fusion to merging colors at the end of the pipeline, and thus reduces the exploratory capacities. Finally, sorted traversals of the voxel models, image-aligned splatting and shear-warp supporting any fusion type require the datasets to be aligned.

Synthesizing, the rendering algorithm that seem more suitable for fast interactive explorations of multimodal datasets based on fusion of properties are the shear-warp factorization for low pixel/voxel ratios and splatting for higher ratios. The goal of this paper is to provide a fast rendering method for multimodal datasets. We propose a data structure that accelerates the sorted traversal of the data and that allows interactive modifications of the fusion criteria. We assume that the different sets are originally aligned or have been aligned and resampled at the same resolution and orientation using existing methods.[9]

2. DISCUSSION

Figure 1 illustrates the Property Fusion (PF) rendering pipeline. Let n be the number of voxels of a multimodal dataset composed of m modalities. Let n_i $i = 1...m$ be the number of non-empty voxels in each modality, and $p_i(x)$ the i-th property value at voxel x. Observe that if a voxel value is empty in one modality but not in another, the voxel can be rendered, and thus the number of m-tuples $(p_1(x), p_2(x).., p_m(x))$ of property values that should be taken as input of the fusion process is nu such that $\forall(p_1(x_j), p_2(x_j).., p_m(x_j)) : 1 \le j \le nu : (\exists i : 1 \le i \le m : \neg empty(p_i(x_j)))$. Finally, let nf be the number of m_tuples that match a fusion criterion, and thus are rendered. The cost of the pipeline can be expressed as:

$$C1 = n * m * CA_i + \sum_{i=1}^{m} n_i * CG_i + nu * CF + nf * CP \tag{1}$$

where CA_i and CG_i are respectively the cost of accessing to the i-th property value and of computing its gradient, CF the cost of the fusion and CP the cost of the rest of the pipeline (classification, shading and composition).

Observe that in this pipeline, only nf voxels are actually rendered. Therefore, the ideal pipeline would be that which would directly access to these voxels:

$$C2 = nf * m * CA_i + nf * m * CG_i + nf * CF + nf * CP \tag{2}$$

Furthermore, the ratio of costs r between $C2$ and $C1$ ($r = C2/C1$) depends essentially on the ratios between nf and n, nf and nu and on the relative cost of CP in relation to CA_i, CG_i and CF. These relative costs depend on the type of property, of gradient computation methods, of shading model and on how complex the fusion function is.

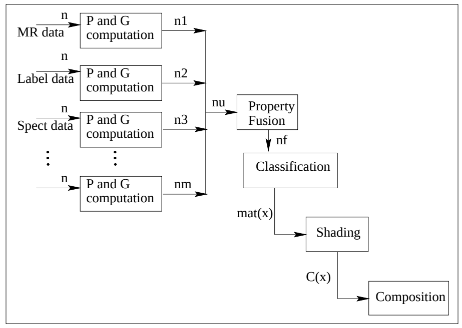

Figure 1. Property Fusion Pipeline.

Before developing the new proposed method exposed in Section 3, we have performed several simulations on different multimodal datasets in order to analyze the magnitude of ratio r in practical applications. Table 1 shows the results of these simulations for a 190*220*178 multimodal MR/LabeledMR/SPECT study of the brain. MR (Magnetic Resonance) data consists of 8-bits per voxel density values showing the patients's head, and more specifically the brain anatomy. The labeled MR set has been constructed from the MR set by segmenting and labelling the different regions of the brain. Finally, the SPECT (Single Positron Emission Tomography) set consists of 3*8-bits RGB values that show the activity of the brain. We have asked to several neurosurgeons which type of queries they usually do or would like to do on such a study. These queries can be roughly classified into three categories:

- $Q1$: Show a weighted average between SPECT and MR values where the MR indicates the presence of brain and MR only everywhere else. This query shows the relationship between the brain anatomy and its activity.
- $Q2$:Show SPECT values inside specific anatomic regions. This query will show SPECT framed into the region.
- $Q3$: Show MR values where the SPECT is intensity falls between a specific range (high values for instance). This query uses SPECT values as a segmentation filter of the MR.

Table 1 shows the results of five simulations on the multimodal dataset: one for query Q1, two for query Q2 selecting a large anatomic region (Q2a) *(left cerebral exterior)* and a small one (Q2b) *(right vessel)* and two for Q3, selecting a wide range of Spect intensities (Q3a) and a narrow one (Q3b).

The values of n, n_i (i=1,..3) and nu are depicted in the upper part of the table along with the ranges of property selected. It can be observed that n_i are quite smaller than n, although Magnetic Resonance data are not precisely of low occupancy in comparison to angiographic (MRA, Magnetic Resonance Angiography) studies. The number of non-empty m-tuples nu is obviously greater than the maximum value of the n_i. The number of voxels that match the fusion criterion nf is shown in the middle part of the table along with the ratio of costs $r = C2/C1$. It can be observed that nf can be almost as large as n as it happens in query $Q1$, or dramatically smaller in query $Q2$ if the selected anatomical region is tiny. The ratio r varies accordingly to these relationships. However, even in the worst case (Q1), it still supposes a reduction of almost 30% of the computational cost. From this, we conclude that it is worth investigating rendering methods that approach the ideal cost $C2$.

There has been several attempts to accelerate the traversal of volume models. Hierarchical data structures used to skip empty space and to codify homogeneous regions include kd-trees[10] and octrees,[7] ,[11] .[12] They present two major drawbacks. First the tree traversal causes a cost overhead. Next, the error associated to the nodes is a global parameter, therefore, the data structure does not provide a local control of the error in a specific region. To overcome these drawbacks, other data structures have been proposed such as shells,[13] distance transforms,[14] ,[15] and run-length encoding,[16] .[17] The shear-warp algorithm, based on a double run-length encoding of the voxel array and of the image scan-line is recognized as the fastest software rendering method. These previous works deal with monomodal datasets. They benefit from the fact that the ratio nu/n is low (being nu the number of non-empty voxels and n the number of voxels, according to the notation used in the previous section). We here extend the use of run-length encoding to multimodal rendering by exploiting the low ratio nf/n, being nf the number of non-empty voxels that match the fusion criteria.

n	7.440.400	190x220x178
n1 (mr)	5.559.241	(2-255)
n2 (spect)	1.436.456	(1-255, 1-145, 1-145)
n3 (label)	1.366.905	(1-35)
nu mr-spect	5.563.025	
nu label-spect	2.336.429	

	nf	$C2/C1$
Q1	2.135.351	0.72
Q2 a	285.270	0.11
Q2 b	48	0.09
Q3 a	1.417.917	0.49
Q3 b	63.505	0.19