

Information Theory-Based Automatic Multimodal Transfer Function Design

Roger Bramon, Marc Ruiz, Anton Bardera, Imma Boada, Miquel Feixas, and Mateu Sbert

Abstract—In this paper, we present a new framework for multimodal volume visualization that combines several information-theoretic strategies to define both colors and opacities of the multimodal transfer function. To the best of our knowledge, this is the first fully automatic scheme to visualize multimodal data. To define the fused color, we set an information channel between two registered input data sets and, afterwards, we compute the informativeness associated with the respective intensity bins. This informativeness is used to weight the color contribution from both initial 1D transfer functions. To obtain the opacity, we apply an optimization process that minimizes the informational divergence between the visibility distribution captured by a set of viewpoints and a target distribution proposed by the user. This distribution is defined either from the data set features, from manually set importances, or from both. Other problems related to the multimodal visualization, such as the computation of the fused gradient and the histogram binning, have also been solved using new information-theoretic strategies. The quality and performance of our approach is evaluated on different data sets.

Index Terms—Multimodal visualization, Multimodal fusion, Transfer function design, Information theory, Kullback-Leibler distance.

I. INTRODUCTION

Multimodal visualization aims at combining the most relevant information from different volumetric data sets into a single one that provides as much information as possible [1]. This technique is of great interest, especially in a medical context where complementary information from different medical devices, such as computed tomography and magnetic resonance, can be combined in a single model to enhance diagnosis and treatment.

Multimodal visualization techniques require two main processes. The first one is the information fusion which reduces the information of spatial-aligned input data sets into a single value. To carry out this fusion different methods have been proposed [1], [2]. The second process is the transfer function definition that assigns graphical attributes (color and opacity) to the fused model to determine which structures of each volume will be visible and how these will be rendered. The definition of this transfer function is a complex task since it is not always easy to understand the relationship between the structures of the input models nor determine which of

them have to be visualized and how. Generally, to tackle this problem, advanced transfer function editing tools are proposed and main decisions are relegated to the user who modifies the parameters until the desired rendering effects are reached. A main drawback of this edition process is the high degree of user interaction which may introduce errors and also makes the reproducibility of the method difficult. To overcome these limitations, the automation of both the fusion and the transfer function design is needed.

In this paper, we present a new approach to automate both the information fusion process and the transfer function design for multimodal data sets. This approach combines several information-theoretic strategies to define colors and opacities. These strategies are based on the information maps introduced by Bramon et al. [3] to represent the informativeness associated with the intensity values of the input data sets. To compute these maps, we establish an information channel between two registered input data sets and calculate the informativeness using two different information measures, that correspond to two different decompositions of the mutual information of the channel.

In order to obtain the fused color, we weight the original 1D transfer functions according to the informativeness associated to each intensity. This fusion is analyzed using different color spaces and color fusion strategies. While in Bramon et al. [3] the information maps were only used to select the most informative color from two input data sets, in this paper they are used to weight the fusion of the colors. Then, to define the opacity function, we have extended the approach presented by Ruiz et al. [4] to deal with multimodal information. Similar to this previous work, we propose an optimization procedure that minimizes the informational divergence between the visibility distribution (i.e., the normalized visibility histogram) captured by a set of viewpoints and a target distribution proposed by the user. The target distribution represents an importance-based description of what the user expects to be visualized. It is important to emphasize that, in this stage, the extension to multimodality forces us to introduce two preliminary steps: a binning strategy to reduce the number of bins of the data sets and a new gradient fusion method to obtain a single value for the gradient magnitude associated with each voxel.

The main contribution of our approach is the definition of a general framework for the automatic transfer function definition in multimodal visualization. It is general in the sense that it is not limited to specific image modalities nor to particular anatomical regions, and, thus, it can be applied to any type of multimodal image pair. This feature is very valuable in real medical environments. As far as we know, this

Manuscript received June 19, 2012. This work has been funded in part by grants from the Spanish Government (Nr. TIN2010-21089-C03-01) and from the Catalan Government (Nr. 2009-SGR-643). This work has also been supported by SUR of DEC of Generalitat de Catalunya (Catalan Government).

Roger Bramon, Marc Ruiz, Anton Bardera, Imma Boada, Miquel Feixas, and Mateu Sbert are with the University of Girona. E-mail: roger.bramon@udg.edu

is the first attempt to define an automated pipeline that finds an optimal transfer function for two multimodal data sets.

This paper is organized as follows. In Section II, we review related work on multimodal visualization and some applications of information theory to visualization. In Section III, we describe the information maps computation. In Section IV, we overview the proposed approach. In Sections V and VI, we explain in detail the main processes of our proposal: color fusion and opacity computation, respectively. In Sections VII and VIII, we show the experimental results and discuss the strengths of our method. Finally, in Section IX, we present our conclusions and future work.

II. RELATED WORK

In this section, we present previous work on multimodal volume rendering and review some information-theoretic applications in visualization.

A. Multimodal Volume Rendering

The main goal of multimodal visualization is to provide in a single image the most important features of different input data sets [1], [2]. To reach this goal, a fusion process that combines the input data is required. For each position, this process can consider single or multiple properties. In the first case, the property can be selected by a user-defined criterion, as proposed by Burns et al. [5] and Brecheisen et al. [6], or by an automatic method, such the one introduced by [3]. In the second case, the fusion can occur at different levels of the volume rendering pipeline [1], [7]. Cai and Sakas [1] defined three levels: image level intermixing, when two images are merged; accumulation level intermixing, when sample values are calculated in each volume along a ray and their visual contributions are mixed; and illumination model level intermixing, which consists in opacity and intensity calculation at each sampling point directly from a multi-volume illumination model. This approach requires multidimensional transfer functions capable of balancing the visual contributions from the input data sets.

Although multidimensional transfer functions are commonly used for volume visualization, their definition is not trivial. The concept of 2D transfer function, where the second dimension is given by the gradient magnitude, was introduced by Levoy [8]. More general multi-dimensional transfer functions were suggested by Kindlmann and Durkin [9], and Kniss et al. [10]. Kniss et al. [11] also proposed an extension of preintegrated volume rendering for multidimensional transfer functions, which was limited to transfer functions specified by Gaussian primitives. Tory et al. [12] proposed the use of an interface based on parallel coordinates to explicitly represent the visualization parameter space of a transfer function. Haidacher et al. [13] introduced the decomposition of mutual information for transfer function design in multimodal volume visualization. They proposed a new 2D space for manually defining transfer functions. Bruckner and Möller [14] introduced isosurface similarity maps to present structural information of a volume data set by depicting similarities

between individual isosurfaces quantified by mutual information. The maps are used to guide the transfer function design and the visualization parameter specification. Based on the mutual information as a measure of the isosurface similarity between different modalities, Haidacher et al. [15] defined a similarity space that provides a concise overview of the differences between modalities, and also serves as the basis for an improved selection of features.

To guide the transfer function design, different authors have proposed to use the data set visibility. Correa and Ma [16] introduced the notion of visibility histogram, which represents the contribution of each sample in the final resulting image, as an interactive aid to generate effective transfer functions. Correa and Ma [17] also generalized the notion of visibility histogram along a number of dimensions and proposed a semiautomated method that progressively explores the transfer function space towards the goal of maximizing the visibility of important structures. Ruiz et al. [4] also used the visibility as a main parameter to be considered for the transfer function specification. They proposed an information-theoretic framework for automatic transfer function design that, based on a user-defined target distribution, obtains the opacity transfer function whose visibility distribution minimizes the informational divergence to the target. Our purpose is now to extend this approach to multimodal volume visualization aiming to automate as much as possible the multimodal transfer function design. In this extension, the information maps proposed by Bramon et al. [3] play a fundamental role to define the fusion strategy.

B. Information Theory in Visualization

In 1948, Claude E. Shannon published a paper entitled “A mathematical theory of communication” [18] that marks the beginning of information theory. In this paper, he introduced the concepts of entropy and mutual information that have been used in many fields, such as physics, computer science, neurology, image processing, and computer graphics. The application of information theory to computer graphics and scientific visualization has been reviewed by Sbert et al. [19], Chen and Jänicke [20], and Wang and Shen [21].

Information theory has been applied to different areas in scientific visualization, such as view selection, flow visualization, time-varying volume visualization, multimodal visualization, and transfer function design. Next, we refer some applications to these areas. In view selection, Bordoloi et al. [22] and Takahashi et al. [23] introduced the entropy to evaluate the quality of a viewpoint, and Viola et al. [24] proposed the mutual information of the information channel between a set of viewpoints and a set of objects to calculate the representativeness of a viewpoint. In flow visualization, Xu et al. [25] used entropy to measure the information content in the local regions across a vector field and conditional entropy to evaluate the effectiveness of streamlines to represent the input vector field, and Lee et al. [26] used entropy for viewpoint selection and view-dependent streamline placement. In time-varying volume visualization, Ji and Shen [27] applied entropy to dynamic view selection, and Wang et al. [28] introduced the conditional

entropy to quantify the information a data block contains with respect to other blocks in the time sequence. Finally, different works mentioned in Section II-A have used mutual information for multimodal visualization [13], [14], [15], [3] and applied the informational divergence for transfer function design [4].

III. INFORMATION MAPS

Since the concept of information map, introduced by Bramon et al. [3], constitutes the kernel of our approach, in this section we briefly review it.

The relationship between two multimodal data sets can be represented by a communication channel $X \rightarrow Y$ between the random variables X (input) and Y (output), which represent, respectively, the set of intensity bins \mathcal{X} of the data set \mathbf{X} and the set of intensity bins \mathcal{Y} of the data set \mathbf{Y} . The three basic components of this channel are the *input distribution* $p(X) = \{p(x)\} = \left\{\frac{n(x)}{N}\right\}$, where $n(x)$ is the number of voxels corresponding to bin x and N is the total number of voxels, the *conditional probability matrix* $p(Y|X) = \{p(y|x)\} = \left\{\frac{n(x,y)}{n(x)}\right\}$, where $n(x,y)$ is the number of voxels with intensity x such that the corresponding voxel in the data set \mathbf{Y} has intensity y , and the *output distribution* $p(Y) = \{p(y)\} = \left\{\frac{n(y)}{N}\right\}$, where $n(y)$ is the number of voxels corresponding to bin y .

From this channel, the *mutual information* $I(X;Y)$ between the two data sets is defined by

$$I(X;Y) = H(Y) - H(Y|X), \quad (1)$$

where $H(Y)$ and $H(Y|X)$ are, respectively, the entropy of Y and the conditional entropy of Y when X is known [29]. Mutual information provides us the amount of information that is transferred or shared between X and Y .

To quantify the specific information associated with each intensity value, $I(X;Y)$ can be decomposed as

$$I(X;Y) = \sum_{x \in \mathcal{X}} p(x) I(x;Y), \quad (2)$$

where $I(x;Y)$ is the *specific information* of x . Thus, $I(X;Y)$ can be seen as a weighted average over individual contributions from particular intensities. Three specific information measures, called surprise (I_1), predictability (I_2) and entanglement (I_3), were previously introduced in the field of neural systems to investigate the information associated with stimuli and responses (see [30], [31]). Bramon et al. [3] introduced these measures in the field of multimodal fusion and concluded that the best performance was achieved by a procedure that combines the measures predictability and entanglement. Taking this fact into account, we focus our attention on these two measures, that will be used in this paper to produce the information maps of each data set.

From Equation 1 and 2, the specific information I_2 [30], called also *predictability* in [3], is defined by

$$\begin{aligned} I_2(x;Y) &= H(Y) - H(Y|x) \\ &= - \sum_{y \in \mathcal{Y}} p(y) \log p(y) + \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x), \end{aligned} \quad (3)$$

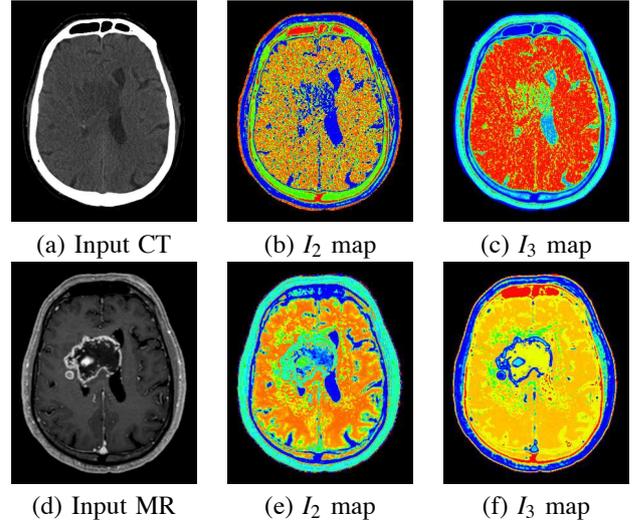


Fig. 1. From left to right, the original CT and MR head data sets and their corresponding I_2 and I_3 information maps.

where $H(Y|x)$ expresses the entropy of Y when the output x is known. The specific information $I_2(x;Y)$ expresses the change in uncertainty about Y when x is observed. Note that $I_2(x;Y)$ can take negative values. This means that certain observations x do increase our uncertainty about the state of the variable Y . Intensity values x with high $I_2(x;Y)$ greatly reduce the uncertainty in Y and, thus, they are very significant in the relationship between X and Y .

Butts [31] introduced the stimulus specific information I_3 , also obtained from the decomposition of $I(X;Y)$. This measure, called *entanglement* in [3], is defined by

$$I_3(x;Y) = \sum_{y \in \mathcal{Y}} p(y|x) I_2(y;X). \quad (4)$$

A large value of $I_3(x;Y)$ means that the intensity values of Y associated with x are very informative in the sense of $I_2(y;X)$. That is, the most informative input values x are those that are related to the most informative outputs y . Note that $I_3(x;Y)$ can also take negative values.

Thus, for each data set, we can obtain two information maps given by the specific information measures I_2 and I_3 , respectively. To avoid negative values in the information maps, the value range of each map has been shifted so that its minimum value is equal to 0. These information maps will enable us to fuse the initial gradients of both data sets into a single value, and the colors of both transfer functions into a single color. Figure 1 shows the I_2 and I_3 information maps for the CT and MR head data sets. These maps have been colored using a thermal scale, where warm colors (red) correspond to high values of the evaluated measure and cool colors (blue) to low ones.

IV. OVERVIEW

In Bramon et al. [3], given two input data sets, the information maps were proposed to select for each voxel the most informative source data set that has to be visualized. In that case, there were no fusion at the voxel level since only

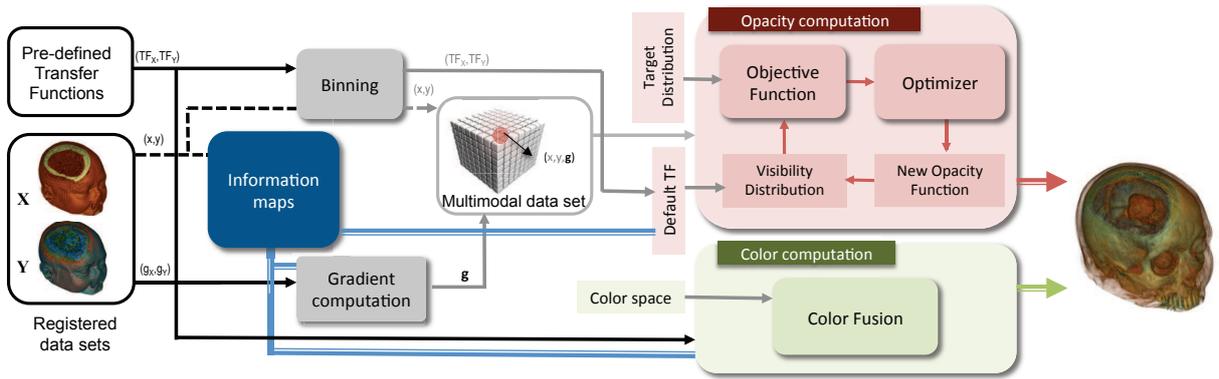


Fig. 2. Main processes of the proposed multimodal visualization approach.

the information of one of the input data sets is visualized, discarding the other one. On the other hand, we propose now to apply the information maps to fuse the input data sets and to create a multimodal visualization where both inputs are represented at each voxel.

The objective of our approach is the automation of the multimodal transfer function design. Given two registered volume data sets, \mathbf{X} and \mathbf{Y} , their pre-defined 1D transfer functions, TF_X and TF_Y , and their information maps, our approach is composed of two main steps (see Figure 2):

- 1) **Color fusion.** This process weights the contribution of the colors provided by the initial 1D transfer functions to obtain the final fused color function. The combination of colors is guided by the information maps.
- 2) **Opacity computation.** This step computes the final opacity function using an iterative strategy that minimizes the informational divergence (or Kullback-Leiber distance) between the visibility distribution captured by a set of viewpoints and a target distribution proposed by the user to obtain the color opacity function.

A more detailed description of these steps is given in the next sections.

V. COLOR FUSION

To obtain the final fused color, the contribution of the colors provided by the initial 1D transfer functions is guided by the I_2 and I_3 information maps of the input data sets. The study carried out by Bramon et al. [3] showed the good performance of an asymmetric fusion strategy based on $I_2(x;Y)$ and $I_3(x;Y)$. In this approach, for each pair of matched voxels with intensities x and y , the graphical attributes of x were selected when $I_2(x;Y) > I_3(x;Y)$, and the ones of y when $I_2(x;Y) \leq I_3(x;Y)$. In our method, this approach is extended to fuse the gradient values. Thus, given a reference data set \mathbf{X} , the gradients of the voxels of \mathbf{X} and \mathbf{Y} are respectively weighted by $I_2(x;Y)$ and $I_3(x;Y)$ from the previously computed information maps. As we have mentioned in Section III, the values of I_2 and I_3 have been shifted to avoid negative values.

At each voxel, the fused data set takes a color c that is a combination of colors $c_X(x)$ and $c_Y(y)$ coming from the initial

transfer functions. The fused color c is defined by

$$c(x, y) = \frac{I_2(x;Y)c_X(x) + I_3(x;Y)c_Y(y)}{I_2(x;Y) + I_3(x;Y)}, \quad (5)$$

where, as stated in Section III, $I_2(x;Y)$ measures the predictability of the intensity value x over the variable Y , and $I_3(x;Y)$ gathers the predictability of the intensity values of Y associated with intensity x . Note that the proposed color fusion strategy will generate new colors and this may lead to misinterpretation when viewing the final color transfer function. This limitation is inherent to any color fusion technique. We have studied different color fusion techniques. First, Equation 5 has been applied using RGB and CIELab color spaces. CIELab (abbreviation for the CIE 1976) color space is perceptually uniform and has been designed to approximate human vision. Second, we have also studied the hue-preserving color blending strategy proposed by Chuang et al. [32] in HSL color space. They proposed a perception-guided compositing operator for color blending, denoted by \oplus , which maintains the same rules for achromatic compositing as standard operators, but it modifies the computation of the chromatic channels in order to preserve the hue of the input colors. This strategy requires to slightly modify Equation 5 replacing the traditional component-wise addition by the new operator:

$$c(x, y) = \frac{I_2(x;Y)c_X(x) \oplus I_3(x;Y)c_Y(y)}{I_2(x;Y) + I_3(x;Y)}. \quad (6)$$

Figure 3 presents, in the first row, the original MR-T1 and MR-T2 data sets and, in the second row, from left to right, the results obtained using the RGB and CIELab spaces, and the hue-preserving color blending strategy. Observe that the results using RGB and CIELab color spaces are very similar, although a more natural hue transition and a more uniform color distribution are obtained using CIELAB color space. On the other hand, note that the hue-preserving color blending strategy tends to produce gray values that can be hard to be interpreted as Chuang et al. described in [32]. From these results, we consider that the CIELab color space is the best option because it enables to identify the origin of the colors better than using the hue-preserving color blending strategy.

From Equation 5, observe that if $I_2(x;Y) > I_3(x;Y)$ then x is more informative than y and, thus, has to have a greater

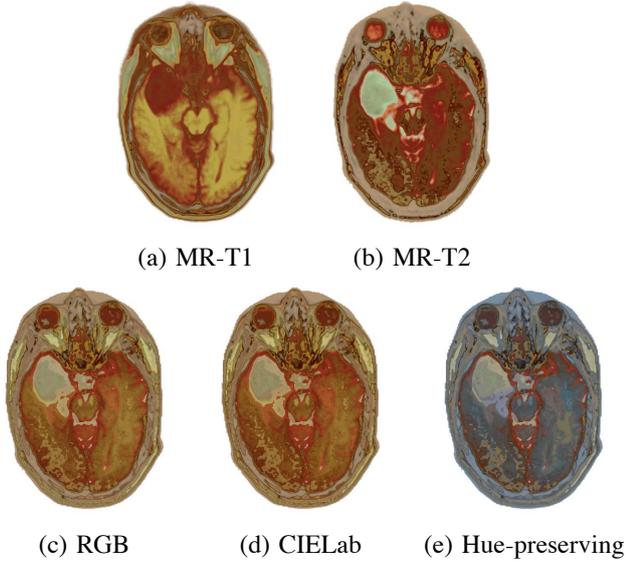


Fig. 3. Multimodal visualization of (a) MR-T1 and (b) MR-T2 in (c) RGB and (d) CIELab color spaces, and (e) the hue-preserving color blending strategy.

contribution to the final result, while $I_2(x;Y) < I_3(x;Y)$ indicates that the values y corresponding to x are more informative than x and must have a greater contribution. Remember that $I_3(x;Y)$ has a low value when the values y corresponding to x are uninformative in the sense of I_2 . Note the asymmetric role of \mathbf{X} and \mathbf{Y} since both measures I_2 and I_3 are taken from \mathbf{X} . This means that, prior to the fusion, we have to select the reference data set. From the experiments carried out in Bramon et al. [3], it can be seen that the best results are achieved when the reference data set corresponds to the one whose structures of interest are more contrasted.

To quantify the contrast of a data set, for each voxel, the variance of the intensities on a small window centered in the voxel is computed. This value can be seen as a measure of local non-uniformity. Thus, the mean of the local variance for all the voxels can be used as an inverse measure of contrast. The lower the mean local variance the higher the contrast. For normalization purposes, we compute this measure on the segmented volumes, since in this case both data sets take values in the same intensity range. In our framework, the most contrasted image is taken as the reference image by default, but the user can easily modify this automatic selection.

VI. OPACITY COMPUTATION

To calculate the opacity values of the multimodal transfer function, we present a method that is based on the transfer function design technique for single data sets introduced by Ruiz et al. [4]. In this approach, opacities are obtained by an optimization procedure that minimizes the informational divergence between the average projected visibility distribution from all viewpoints and a target distribution which expresses an importance-based description of what the user expects to be visualized. The main modifications to extend this approach to multimodal visualization are due to the fact that we have to consider pairs of intensity values at each voxel instead of

single values. In addition to the mathematical reformulation of the method, two new steps will be required to solve the high-dimensionality associated to the problem: the binning of the intensities and the gradient fusion.

A. Multimodal Opacity Optimization

The main steps of this process are represented in the opacity computation module of Figure 2. This process begins with a default multimodal transfer function, obtained from a weighted average of the opacity values from the 1D input transfer functions TF_X and TF_Y . Similar to the color fusion (see Equation 5), the weights are given by the I_2 and I_3 information maps. This new 2D multimodal transfer function is used to compute the visibility distribution for a set of viewpoints. Then, the informational divergence or Kullback-Leibler distance [29] between the obtained visibility distribution and the target distribution is evaluated. The target distribution represents an importance-based description of what the user expects to be visualized, i.e., the probability of each bin in the final visualization. From the informational divergence value, the optimizer, based on the steepest gradient descent algorithm, assesses a new transfer function in the direction of the divergence gradient. The process is repeated until the value of the informational divergence is below a given threshold or a given number of iterations has been performed.

The computation of the informational divergence is carried out in the framework on an information channel $V \rightarrow B$ between random variables V and B that are respectively defined over the alphabets \mathcal{V} (set of viewpoints) and \mathcal{B} (set of bins), where each bin corresponds to the set of voxels that have the same pair (x,y) of intensity values or the same triplet (x,y,g) of intensities and gradient. It is assumed here that all the volume data sets are centered in a sphere of viewpoints and the camera is looking at the center of this sphere. The main elements of the channel $V \rightarrow B$ are the *conditional probabilities* $p(b|v)$, given by the normalized projected visibility of intensity bin b over a viewpoint v , the *input probability* $p(v)$, given by the normalized projected visibility of the data set over a viewpoint v , and the *output probability* $p(b)$, given by $p(b) = \sum_{v \in \mathcal{V}} p(v)p(b|v)$ that expresses the average projected visibility of intensity bin b from all viewpoints. For more details, see [4].

In this paper, three different target distributions have been used:

- *Occurrence of the intensities*: the target distribution obtained from the occurrence of each intensity bin b is defined as

$$q(b) = \frac{\text{occur}(b)}{\sum_{i \in \mathcal{B}} \text{occur}(i)}, \quad (7)$$

where $\text{occur}(b)$ stands for the occurrence of bin b . This approach requires that each intensity bin, i.e., each pair of intensities (x,y) , is visualized according to its probability in the volume data set. Note that the original resolution of the intensities can not be used (as it was in the original paper of Ruiz et al.) due to the high number of different pair combinations. Thus, a binning strategy has to be applied. In our framework, we used an information

bottleneck-based approach, which is described in more detail in Section VI-B.

- *Gradient magnitude*: the previous target distribution is extended by using the 3D transfer function generated by the intensity pair and the gradient. Using this extension, the target distribution obtained from the gradient values weighted by the occurrence distribution is defined by

$$q(b) = \frac{grad(b)occur(b)}{\sum_{i \in \mathcal{B}}(grad(i)occur(i))}, \quad (8)$$

where $grad(b)$ stands for the gradient component g of the bin b . Note that \mathcal{B} represents now the joint variable (x, y, g) . In this case, the voxels with a high gradient (i.e., those that are borders of anatomical structures) can be highlighted. Note that each input data set has a different gradient magnitude and a fusion scheme is needed also in this case. We propose to fuse them based on the information maps. Section VI-C describes this technique in more detail.

- *Importance function*: the previous target distributions can be weighted by an importance function $imp(b)$ defined by the user. For instance, weighting the second one by importance, we obtain the following target distribution:

$$q(b) = \frac{imp(b)grad(b)occur(b)}{\sum_{i \in \mathcal{B}}(imp(i)grad(i)occur(i))}. \quad (9)$$

In this way, a priori knowledge of the data, such as the intensity range of the relevant structures, is combined with statistical features of the data.

The informational divergence or Kullback-Leibler distance [29] measures the distance between the visibility distribution and a target distribution $q(B)$. From this measure, two different approaches can be defined depending on how the visibility is estimated:

- *Global informational divergence (GID)*, which is defined as

$$D_{KL}(p(B), q(B)) = \sum_{b \in \mathcal{B}} p(b) \log \frac{p(b)}{q(b)}, \quad (10)$$

where $p(b)$ is the average projected visibility of intensity bin b from all viewpoints, and thus $p(B)$ represents the mean visibility of each intensity bin considering all the viewpoints.

- *Viewpoint informational divergence (VID)*, which only considers the current viewpoint v . Thus, Equation 10 becomes

$$D_{KL}(p(B|v), q(B)) = \sum_{b \in \mathcal{B}} p(b|v) \log \frac{p(b|v)}{q(b)}, \quad (11)$$

where $p(B|v)$ represents the visibility of each intensity bin considering only the current viewpoint. Note that this measure is view-dependent and will have to be recomputed each time the viewpoint changes.

Ruiz et al. [4] proposed to add an opacity constraint term to the information divergence to ensure a high degree of opacity of the final transfer function. In our framework this term has not been added since the method does not lead to very transparent results without this term. Thus, our objective

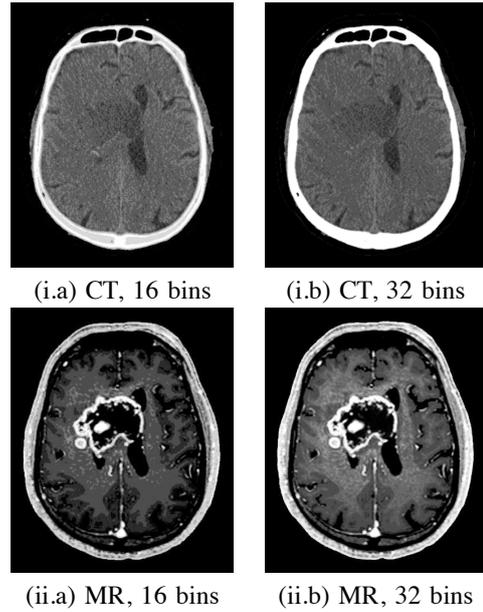


Fig. 4. CT and MR head data sets of Figure 1(a) and 1(d) are shown after applying the binning step with (a) 16 bins and (b) 32 bins.

is to minimize the informational divergence by modifying the opacities of the multimodal transfer function. This optimization procedure is performed using the steepest gradient descent method and using an estimation of the gradient of the informational divergence to speed up the process. For more details see [4].

B. Binning Algorithm

Given the information channel between two registered data sets presented in Section III, the number of bins of each data set is reduced by applying the one-sided clustering algorithm introduced by Bardera et al. [33]. On the one hand, the necessity of this process is due to the computational difficulty of dealing with the high number of bins that result from the combination of two input data sets. On the other hand, the one-sided clustering algorithm, designed for multimodal image segmentation, allows us to obtain a more accurate result than a regular binning approach.

This binning algorithm, based on the agglomerative information bottleneck method [34], is a greedy hierarchical clustering algorithm that merges the histogram bins of one data set by minimizing the loss of mutual information between both data sets. The main idea behind the algorithm is that the final segments of one data set correspond to the structures that are most relevant from the perspective of the other data set, called control data set. For more details, see [33]. Figure 4 shows, for the original CT and MR head data sets of Figure 1, the results obtained after applying the binning process with 32 and 16 bins. Observe how the main structures of the original images have been preserved.

C. Gradient Computation

In volume rendering, the gradient is needed to obtain the normals for the shading calculation. In addition, the gradient

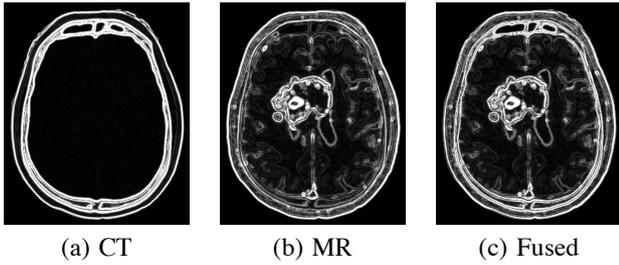


Fig. 5. From left to right, the representation of the gradient magnitudes of the input CT and MR head data sets, and the fused data set.

magnitudes can be used to define the transfer function. In the multimodal visualization scenarios, each data set contributes with a gradient value and, therefore, a strategy to fuse these multiple values in a single one is required.

To obtain the fused gradient, we use the fusion strategy proposed in Section V for color fusion. Thus, the fused gradient magnitude g in voxel (i, j, k) is defined by

$$g(i, j, k) = \frac{I_2(x; Y)g_X(i, j, k) + I_3(x; Y)g_Y(i, j, k)}{I_2(x; Y) + I_3(x; Y)}, \quad (12)$$

where $g_X(i, j, k)$ and $g_Y(i, j, k)$ stand for the gradient magnitudes in the voxel (i, j, k) of data sets X and Y , respectively. The fused gradient direction is also computed in a similar way.

For the computation of g_X and g_Y , the 4D linear regression algorithm proposed by Neumann et al. [35] has been applied to the original data sets (before the binning step). With this method we obtain a more accurate gradient approximation than using the standard finite difference method [35]. Figure 5 shows the gradient magnitude for the input CT and MR head data sets, and the fused gradient. Note that the fused gradient preserves the main structures of the input models without disruptive discontinuities.

Observe that the gradient associated with the intensity value x of the reference data set contributes more when its predictability is greater than the predictability of the intensity values y associated with x , and vice versa. As we discussed in Section V, due to the asymmetric role of \mathbf{X} and \mathbf{Y} , prior to the fusion we have to select the reference data set.

VII. RESULTS

In this section, we present a set of experiments that have been carried out to evaluate the proposed approach. We have considered two testing data sets, the first composed of medical data and the second of industrial data.

A. Medical Applications

For the medical experiments, we have used CT, MR, and PET data sets from the Osirix database [36] and we have analyzed both the CT-MR and the CT-PET fusions. In the CT-MR fusion, CT detects dense structures, such as bones, giving the general shape of objects but few details on the soft tissues, while MR images are used to depict the morphology of soft tissues being rich in detail. Generally, in CT-MR fusion, physicians want to see the dense structures from CT and the soft tissues from MR. In the CT-PET fusion, PET provides

information of metabolism activity patterns while CT provides high quality spatial context information. Generally, in the CT-PET fusion, physicians want to see the functional active areas from PET, and bone and other anatomical structures from CT.

The proposed approach has been integrated in a multimodal visualization platform. Its user interface, developed using Qt [37], integrates two lateral viewers to present the input data sets and a central viewer with the multimodal visualization. The user interacts with the main viewer and all the actions are reproduced to the other ones. We use GPU-based ray casting to render the input models and CPU-based ray casting to render the fused data set based on VTK [38]. Note that multimodal transfer functions have, in general, three input variables: the intensities of both input data sets and the gradient magnitude and, for each triplet, a color and an opacity scalar value have to be shown. The visualization of this information is not a simple task and physicians, who are not very used to deal with this kind of information, could have some difficulties to correctly interpret them. To overcome this limitation, in our experiments we always provide to the users the multimodal visualization together with the original input data sets.

In our experiments, we have used by default the global informational divergence (GID), a stopping threshold value of the informational divergence measure equal to 0.001, and 6 uniformly distributed viewpoints. The first experiment evaluates the CT-MR fusion using a CT ($512 \times 512 \times 174$) and MR ($176 \times 224 \times 244$) head data sets. In a preprocessing step, these data sets have been registered and the MR-head has been resampled to the CT resolution using linear interpolation (see Figures 6(a)). To apply the proposed approach, the CT data set has been considered as the reference data set, since it is more contrasted than MR data set. Different number of intensity and gradient bins have been used in order to evaluate the effect of the binning process. Figures 6(b-c) show the obtained results using both the target distributions given by occurrence and occurrence weighted by gradient, respectively. To better illustrate the results, a cutting plane at the level of the damaged area has been set. With respect to the target distributions, note that when only occurrences are taken into account (Figure 6(b)) no insight of the lesion is visible. On the contrary, when gradient is considered (see Figure 6(c)), the method assigns a lower opacity around the damaged area and this is perfectly delineated. This effect is due to the contrast injected to the patient, in order to enhance the lesion detection. Therefore, for data sets with highly contrasted structures, the proposed approach will achieve better results using the occurrences weighted by gradient as the target distribution. Figure 6(d) has been obtained using occurrences weighted by gradient and assigning importance 1 to the lesion and 0.5 to the rest for the MR, and 1 to the bone and 0.2 to the rest for the CT. In this way, the importance of each pair of the fused data set is obtained by multiplying the importances of each single data set. As it can be seen, the bone and the lesion are notably highlighted in the final rendering. Finally, we can evaluate the effect of the binning process by comparing Figures 6(i.b-i.d) and Figures 6(ii.b-ii.d). We observe that the different number of bins only slightly affects the final colors of the transfer function. Thus, although the binning process

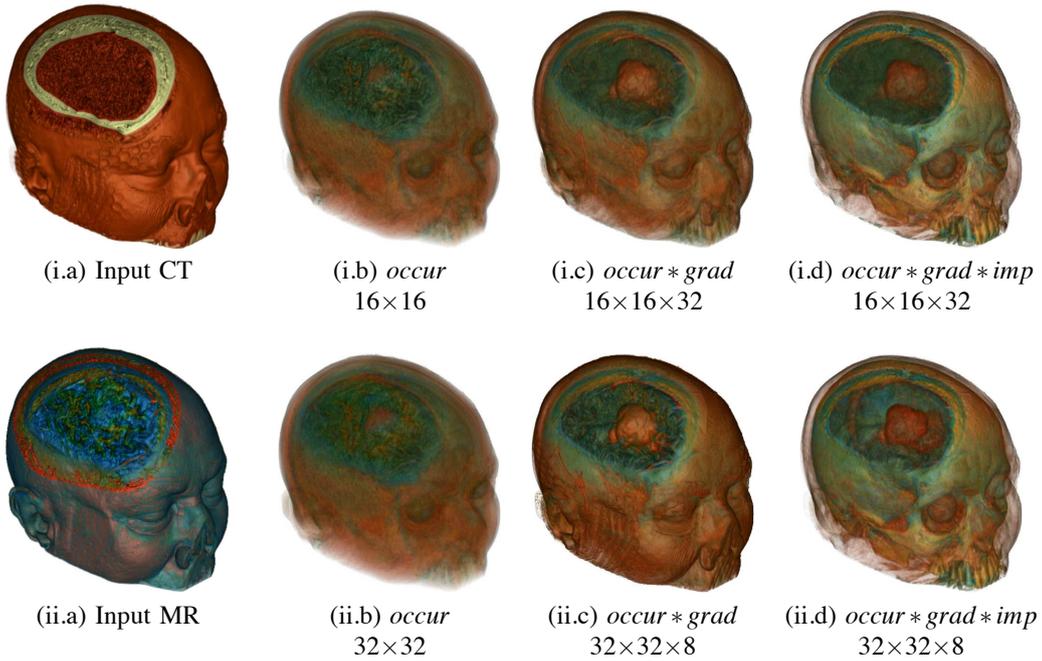


Fig. 6. Multimodal visualization of (i.a) CT and (ii.a) MR data sets using different target distributions: (b) occurrence, (c) occurrence weighted by gradient, and (d) occurrence weighted by gradient and importance. Results (i.b-d) are obtained using 16 non-uniform intensity clusters for each data set and 32 uniform bins for the gradient magnitude, and (ii.b-d) using 32 non-uniform intensity clusters for each data set and 8 uniform bins for the gradient magnitude.

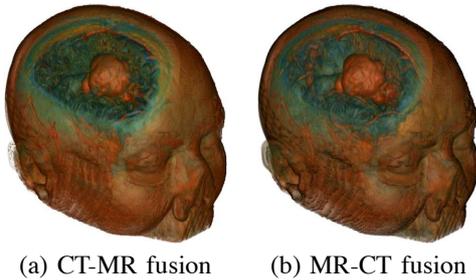


Fig. 7. Comparison of multimodal visualizations of CT and MR head data sets of Figure 6(a) using occurrence weighted by gradient and considering (a) CT and (b) MR as the reference model, respectively.

implies a loss of information, it has no relevant impact to the final result. Using the same pair of data sets, we have also evaluated the difference of using either CT or MR as the reference data set. As we can see in Figure 7, the results are very similar and, thus, the selection of the reference model does not substantially affect the quality of the final rendering.

The next experiment evaluates the CT-PET fusion considering the PET as the reference data set since the PET is more contrasted than the CT. The original data sets (see Figures 8(a-b)) are correctly registered and have a resolution of $(168 \times 168 \times 344)$. In this experiment we use 64 non-uniform intensity clusters for each data set. To obtain the fusion we use the target distribution given by the occurrence, assigning importance 0.7 to the bone of the CT and 0.1 to the rest, and 0.8 to the high activity area of the PET and 0.1 to the rest. The result is shown in Figure 8(c). As it was expected, the integration of the anatomical context from CT makes the interpretation of PET information easier.

The third experiment also evaluates a CT-PET fusion. These data sets are registered and have a resolution of $(512 \times 512 \times 267)$. As in the previous case, PET is considered as the reference for the computation of the information maps, and we use 32 non-uniform intensity clusters for each data set and 8 uniform bins for the gradient magnitude. Figure 9 shows the multimodal visualization using the target distribution of occurrences weighted by gradient and assigning importance 0.7 to the CT bone and 0.1 to the rest. Note how the assignation of importance and the application of the gradient improve considerably the skeleton visualization.

B. User Evaluation

To evaluate the proposed approach in a medical context, we have presented the obtained results to a group of experts from the Hospital Josep Trueta of Girona. The validation of multimodal visualization is a difficult task due to the lack of ground truth data. Moreover, observer's evaluation can be influenced by the diagnostic situation, the observer's experience, training, and preference. Therefore, our evaluation have been based on the capability of the expert to obtain information from the testing images that could be relevant for the diagnosis.

In a first evaluation, we have presented the CT-MR fused data sets obtained with our approach (Figure 6) and also with the classical weighted average visualization, and checkerboard visualization (alternatively visualizing one voxel of each input model) manually modulating the opacities to generate comparable results to our approach. All experts agreed that the most valued image is the Figure 6-ii.c since perfectly delineates the right intra-cerebral mass providing a visualization similar to

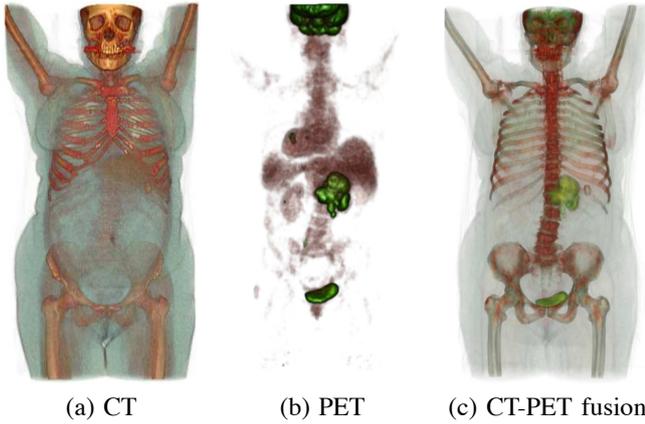


Fig. 8. Multimodal visualization of (a) CT and (b) PET data using occurrence weighted by importance considering the (c) PET as the reference model.

the one obtained with a parietal craniotomy. Moreover, this result was not reproducible with the classical visualization. Figure 6-ii.d has been less valued since loses details of the pathologic mass although it better represents the bone structure and vascular details. Figure 6-ii.a has been directly discarded since it does not provide relevant information for the diagnosis.

In a second experiment, they have analyzed the CT-PET fused data (Figure 8) obtained with our approach and the standard methods previously described. Experts have considered that the active areas of the PET are better represented with the proposed approach than with the standard methods, since these areas have a higher image contrast.

As a conclusion, experts have pointed out the quality of our images and have considered them especially useful for surgical and radiotherapy planning, and for treatment monitoring.

C. Industrial Applications

To show the wide applicability of our method, this has also been tested with an industrial data set. In the industrial area, the dual energy CT (DECT), which performs a high and low energetic measurement simultaneously, has become a novel technique for dimensional measurement of industrial components. The high energy scan is almost free of artifacts but suffers from reduced precision and noise, and the low energy scan has high precision but is affected by severe artifacts [39]. The purpose of fusion is to combine the advantages of both models in a single one. Figures 10(a-b) show the low and high energy scans of a 400 V power connector with a resolution of $(256 \times 256 \times 895)$. The transfer functions used in these visualizations have been obtained with the method proposed by Ruiz et al. [4]. Figures 10(c-d) illustrate the corresponding I_2 and I_3 information maps of both scans. As it was expected, the low energy information maps present severe artifacts while the high energy ones are free of artifacts but suffer from noise. Since the presence of artifacts makes the fusion more difficult, we take the high energy scan as the reference data set. Figure 11 shows the multimodal fusion of DECT data set using the target of occurrence weighted by gradient with both the VID (i.e., only one view is considered) and the GID (6 and 20 views are considered) measures. In

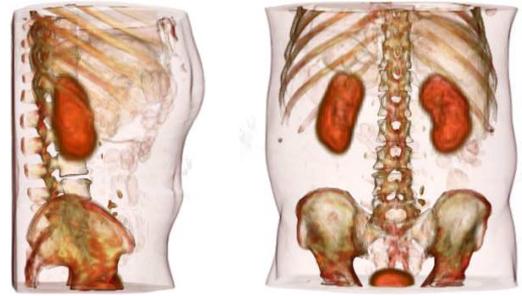


Fig. 9. Two different views of a multimodal visualization of CT-PET fusion using as a target distribution the occurrence weighted by gradient and importance.

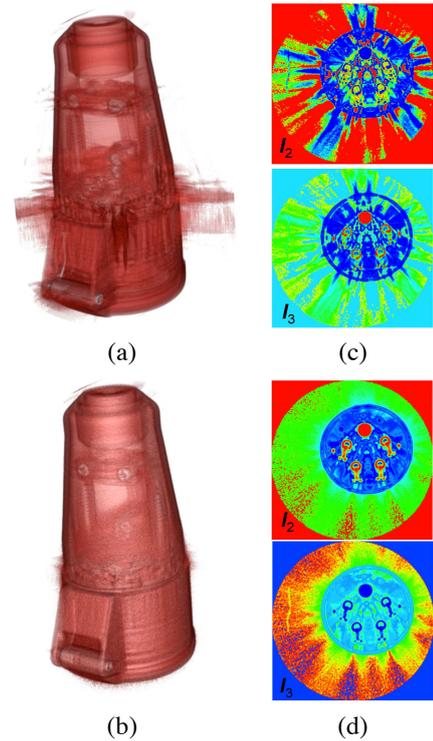


Fig. 10. Visualizations of the (a) low energy and (b) high energy CT scans of a power connector and (c-d) their corresponding I_2 , I_3 information maps.

these experiments we use 16 non-uniform intensity clusters for each data set and 32 uniform bins for the gradient magnitude. When the GID measure is used (see Figures 11(c-d) and (e-f)), a unique transfer function is obtained, while with the VID measure (see Figure 11(a-b)) a new transfer function is defined for each viewpoint. Note how the transfer functions are clearly dependent on the selected viewpoints. In the case of one viewpoint, all structures are visible from each viewpoint, while considering more viewpoints, occlusions do not allow us to perceive all structures from a single viewpoint. We can also observe that the difference using 6 or 20 viewpoints is minimal and, hence, the use of 6 viewpoints is a good trade-off between quality and speed for the global informational divergence.

Table I collects the computation time in seconds for each step of the proposed approach and different data sets. From

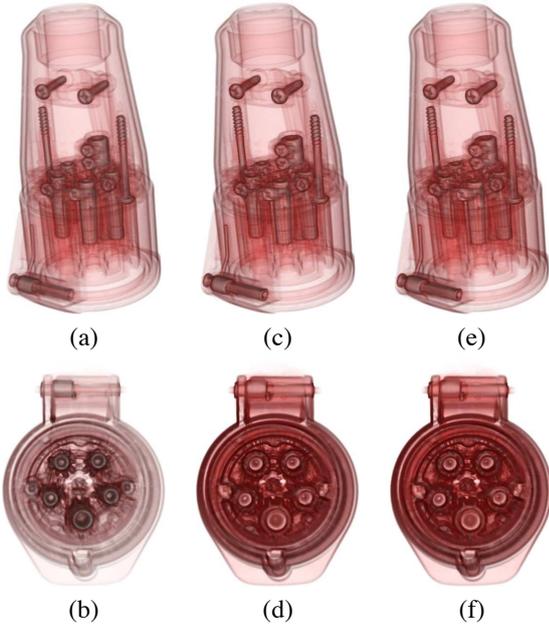


Fig. 11. Multimodal visualization of a dual energy CT scan of a power connector with the target of occurrence weighted by gradient considering (a-b) one view, (c-d) 6 views and (e-f) 20 views.

left to right, columns report evaluated data sets with different configurations and target distributions, data preparation steps (information maps, binning, and gradient), color computation, and opacity computation. In this last column, we considered two different distances to stop the process ($d < 0.01$ and $d < 0.001$) and, for each configuration, we collect the computation time in seconds and the number of iterations required by the opacity process. The performance of our method only benefits from the GPU in the implementation of the visibility computation. Note that in most cases the results converge in less than 50 iterations. All the experiments were carried out on a PC equipped with an Intel Core 2 Quad Q9550 CPU, 4GB of RAM, and a NVIDIA GeForce GTX 280 graphics card.

VIII. DISCUSSION

As we have mentioned in Section II, some approaches have been proposed to assist in multimodal transfer function design. Some previous works [13], [14], [15] present a simplification of the multimodal transfer function space to facilitate the manual definition, even though this is still required. Some other approaches have been proposed for automatic transfer function design. These approaches, that only consider one input data set, deal with the problem of minimizing a cost function while optimizing the opacity values in the transfer function definition. For instance, Correa et al. [17] and Ruiz et al. [4] propose to minimize, respectively, an energy function and the informational divergence between a given visibility function and the visibility obtained with the transfer function.

In our framework, we define a general pipeline to solve the problems related to the multimodal visualization. First, the informativeness of the intensity values of both input data sets is used to obtain a fused gradient function, that is required to compute the illumination and to define the

transfer function. Second, a non-regular histogram binning strategy is proposed to reduce the number of entries of the joint histogram required for the optimization of the transfer function opacities. Third, the informativeness values together with the original colors of both 1D transfer functions are used to generate the color assignment in the multimodal transfer function definition. And fourth, the opacities of the multimodal transfer function are automatically computed from a target distribution by minimizing the informational divergence.

The theoretical fundamentals used in this paper are based on information theory. This theory is used to relate different random variables by defining an information channel between them. Note that, while the information maps and the binning algorithm are obtained from the information channel created between the two input data sets, the informational divergence is computed in the context of an information channel between a set of viewpoints and the bins of the multimodal data set.

In our approach, the problems related to the multimodal visualization are solved by defining a few number of parameters. First, for the binning process, the final number of bins has to be fixed. As it has been shown in Figure 6, the final results are not very sensitive to this parameter and, in our experiments, we have used a default value of 32 bins. Second, the reference image has to be chosen. From our tests, we have observed that the best results are achieved when the most contrasted image is considered as the reference one. For instance, in the visualization of a CT-PET image pair, the best performance is obtained when the PET image is taken as the reference one instead of the CT image. Finally, in the optimization process, a target distribution have to be defined. The choice of the target distribution requires that the user decides which features have to be enhanced.

Focusing on real medical applications, a current limitation of our approach, which is inherent to any color fusion strategy, is the color interpretation of the multimodal visualization. Since new colors are generated, the physician could have difficulties to interpret them. To tackle this problem we can simultaneously visualize the original input data sets with their transfer functions and emphasize the explored area in the fused model to the original data sets.

IX. CONCLUSIONS

We have introduced a novel pipeline to automate the information fusion and the transfer function design required in multimodal visualization. The proposed approach, that combines several information-theoretic strategies to define colors and opacities, is basically composed of the following processes. First, the information maps between two input registered data sets are computed. Second, the fused color is computed from the combination of the original colors using the information maps. Finally, the opacity values are generated by minimizing the informational divergence between the visibility distribution and a target distribution proposed by the user. Before this optimization process, a binning step has been applied to reduce the number of bins of the input data sets and both gradients from the input data sets have been fused to a single gradient value. As future work, we will study the generalization of

Data set	Data preparation			Color computation	Opacity computation (sec./it)	
	Information maps	Binning	Gradient		$d < 0.01$	$d < 0.001$
binsX×binsY×gradient (target)						
CT-MR 32×32×1 (1) 32×32×8 (2)	1.27	10.32	10.40	0.01	(44.67, 26) (110.53, 30)	(48.97, 28) (173.49, 47)
CT-PET 64×64×1 (3)	0.83	4.34	17.74	0.02	(134.79, 34)	(458.45, 101)
Prostatix 32×32×8 (4)	0.87	5.82	15.02	0.01	(88.13, 36)	(128.96, 53)
DECT 16×16×32 (2) 1 viewpoint 16×16×32 (2) 6 viewpoints 16×16×32 (2) 20 viewpoints	1.98	18.41	14.35	0.01	(15.98, 22) (96.61, 29) (316.29, 31)	(30.93, 40) (158.60, 48) (507.32, 50)

TABLE I

TIME COST IN SECONDS REQUIRED FOR THE MAIN STEPS OF THE FUSION PROCESS. TARGET DISTRIBUTIONS ARE: (1) OCCURRENCE, (2) OCCURRENCE WEIGHTED BY GRADIENT, (3) OCCURRENCE WEIGHTED BY IMPORTANCE, AND (4) OCCURRENCE WEIGHTED BY GRADIENT AND IMPORTANCE.

this approach to the visualization of more than two data sets. This extension requires a detailed analysis of the mutual information decomposition for more than two variables. We will also investigate the improvement of the color fusion strategy in order to facilitate the interpretation of the color in a multimodal visualization.

ACKNOWLEDGMENT

The DECT data sets are courtesy of Christoph Heinzl.

REFERENCES

- [1] W. Cai and G. Sakas, "Data intermixing and multi-volume rendering," *Computer Graphics Forum.*, vol. 18, pp. 359–368, 1999.
- [2] B. Wilson, E. B. Lum, and K.-L. Ma, "Interactive multi-volume visualization," in *Computational Science – ICCS 2002*, 2002, pp. 102–110.
- [3] R. Bramon, I. Boada, A. Bardera, J. Rodriguez, M. Feixas, J. Puig, and M. Sbert, "Multimodal data fusion based on mutual information," *TVCG*, vol. 18, no. 9, pp. 1574–1587, 2012.
- [4] M. Ruiz, A. Bardera, I. Boada, I. Viola, M. Feixas, and M. Sbert, "Automatic transfer functions based on informational divergence," *TVCG*, vol. 17, no. 12, pp. 1932–1941, 2011.
- [5] M. Burns, M. Haidacher, W. Wein, I. Viola, and E. Gröller, "Feature emphasis and contextual cutaways for multimodal medical visualization," in *EuroVis*, 2007, pp. 275–282.
- [6] R. Brecheisen, A. Vilanova, B. Platel, and B. M. terHaarRomeny, "Flexible GPU-based multi-volume ray-casting," in *Proceedings of Vision, Modelling and Visualization*, 2008, pp. 303–312.
- [7] M. Ferre, A. Puig, and D. Tost, "A framework for fusion methods and rendering techniques of multimodal volume data," *Computer Animation and Virtual Worlds*, vol. 15, no. 2, pp. 63–77, 2004.
- [8] M. Levoy, "Display of surfaces from volume data," *IEEE Computer Graphics and Applications*, vol. 8, no. 3, pp. 29–37, 1988.
- [9] G. Kindlmann and J. W. Durkin, "Semi-automatic generation of transfer functions for direct volume rendering," in *Proceedings of the 1998 IEEE symposium on Volume visualization*, 1998, pp. 79–86.
- [10] J. Kniss, S. Premoze, C. Hansen, and D. Ebert, "Interactive translucent volume rendering and procedural modeling," in *Visualization, 2002. VIS 02. IEEE*, 2002, pp. 109–116.
- [11] J. Kniss, S. Premoze, M. Ikits, A. Lefohn, C. Hansen, and E. Praun, "Gaussian transfer functions for multi-field volume visualization," in *Visualization, 2003. VIS 03. IEEE*, 2003, p. 65.
- [12] T. Möller, "A parallel coordinates style interface for exploratory volume visualization," *TVCG*, vol. 11, no. 1, pp. 71–80, 2005.
- [13] M. Haidacher, S. Bruckner, A. Kanitsar, and M. E. Gröller, "Information-based transfer functions for multimodal visualization," in *Visual Computing for Biology and Medicine*, 2008, pp. 101–108.
- [14] S. Bruckner and T. Möller, "Isosurface similarity maps," *Computer Graphics Forum*, vol. 29, no. 3, pp. 773–782, 2010.
- [15] M. Haidacher, S. Bruckner, and M. E. Gröller, "Volume analysis using multimodal surface similarity," *TVCG*, vol. 17, no. 12, pp. 1969–1978, 2011.
- [16] C. D. Correa and K.-L. Ma, "Visibility-driven transfer functions," in *PacificVis*, 2009, pp. 177–184.
- [17] —, "Visibility histograms and visibility-driven transfer functions," *TVCG*, vol. 17, pp. 192–204, 2011.
- [18] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.
- [19] M. Sbert, M. Feixas, J. Rigau, I. Viola, and M. Chover, *Information Theory Tools for Computer Graphics*. Morgan & Claypool Publishers, 2009.
- [20] M. Chen and H. Jänicke, "An information-theoretic framework for visualization," *TVCG*, vol. 16, pp. 1206–1215, 2010.
- [21] C. Wang and H.-W. Shen, "Information theory in scientific visualization," *Entropy*, vol. 13, no. 1, pp. 254–273, 2011.
- [22] U. Bordoloi and H.-W. Shen, "View selection for volume rendering," in *Visualization, 2005. VIS 05. IEEE*, 2005, pp. 487–494.
- [23] S. Takahashi, I. Fujishiro, Y. Takeshima, and T. Nishita, "A feature-driven approach to locating optimal viewpoints for volume visualization," in *Visualization, 2005. VIS 05. IEEE*, 2005, pp. 495–502.
- [24] I. Viola, M. Feixas, M. Sbert, and M. E. Gröller, "Importance-driven focus of attention," *TVCG*, vol. 12, no. 5, pp. 933–940, 2006.
- [25] L. Xu, T.-Y. Lee, and H.-W. Shen, "An information-theoretic framework for flow visualization," *TVCG*, vol. 16, no. 6, pp. 1216–1224, 2010.
- [26] T.-Y. Lee, O. Mishchenko, H.-W. Shen, and R. Crawfis, "View point evaluation and streamline filtering for flow visualization," in *IEEE Pacific Visualization Symposium 2011*, 2011, pp. 83–90.
- [27] G. Ji and H.-W. Shen, "Dynamic view selection for time-varying volumes," *TVCG*, vol. 12, no. 5, pp. 1109–1116, 2006.
- [28] C. Wang, H. Yu, and K.-L. Ma, "Importance-driven time-varying data visualization," *TVCG*, vol. 14, no. 6, pp. 1547–1554, 2008.
- [29] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.
- [30] M. R. Dewese and M. Meister, "How to measure the information gained from one symbol," *Network: Computation in Neural Systems*, vol. 10, no. 4, pp. 325–340, 1999.
- [31] D. A. Butts, "How much information is associated with a particular stimulus?" *Network: Computation in Neural Systems*, vol. 14, pp. 177–187, 2003.
- [32] J. Chuang, D. Weiskopf, and T. Möller, "Hue-preserving color blending," *TVCG*, vol. 15, no. 6, pp. 1275–1282, 2009.
- [33] A. Bardera, J. Rigau, I. Boada, M. Feixas, and M. Sbert, "Image segmentation using information bottleneck method," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1601–1612, 2009.
- [34] N. Slonim and N. Tishby, "Agglomerative information bottleneck," in *Proceedings of NIPS-12 (Neural Information Processing Systems)*. MIT Press, 2000, pp. 617–623.
- [35] L. Neumann, B. Cséfalvi, A. König, and E. Gröller, "Gradient estimation in volume data using 4D linear regression," *Computer Graphics Forum*, vol. 19, no. 3, pp. 351–358, 2000.
- [36] "OsiriX database," OsiriX Imaging Software, <http://osirix-viewer.com/datasets>.
- [37] "Qt," Qt Project, <http://qt-project.org>.
- [38] "Visualization toolkit," Kitware Inc., <http://www.vtk.org>.
- [39] C. Heinzl, J. Kastner, and M. E. Gröller, "Surface extraction from multi-material components for metrology using dual energy CT," *TVCG*, vol. 13, no. 6, pp. 1520–1527, 2007.