



Contents lists available at ScienceDirect

## Journal of Cranio-Maxillo-Facial Surgery

journal homepage: [www.jcmfs.com](http://www.jcmfs.com)

# Accuracy of multimodal image fusion for oral and maxillofacial tumors: A revised evaluation method and its application



Lei-Hao Hu, Wen-Bo Zhang, Yao Yu, Xin Peng\*

Department of Oral and Maxillofacial Surgery, Peking University School and Hospital of Stomatology, 22# Zhongguancun South Avenue, Beijing 100081, China

## ARTICLE INFO

## Article history:

Paper received 1 March 2020

Accepted 28 May 2020

Available online 10 June 2020

## Keywords:

Multimodal image fusion

Oral and maxillofacial tumors

Overall fusion accuracy

Tumor volume fusion accuracy

## ABSTRACT

**Objective:** To develop a revised evaluation method for accuracy of multimodal image fusion for oral and maxillofacial tumors and explore its application for comparing the accuracy of three commonly used fusion algorithms, automatic fusion, manual fusion, and registration point-based fusion.

**Materials and methods:** Image sets of patients with oral and maxillofacial tumor were fused using the iPlan 3.0 navigation system. Fusion accuracy included two aspects: (1) overall fusion accuracy: represented by the mean value of the coordinate differences along the x-, y-, and z- axes ( $\Delta x$ ,  $\Delta y$ , and  $\Delta z$ ), mean deviation (*MD*), and root mean square (*RMS*) of six pairs of landmarks on the two image sets; (2) tumor volume fusion accuracy: represented by Fusion Index (*FI*), which was calculated based on the volume of tumor delineated on the two image sets.

**Results:** Eighteen pairs of image sets of 17 patients were enrolled in this study. The  $\Delta x$  and  $\Delta y$  values for the three algorithms were less than 1.5 mm. The  $\Delta z$  values for automatic fusion, manual fusion and registration point-based fusion was 1.049 mm, 1.864 mm and 1.254 mm. The *MD* for automatic fusion, manual fusion and registration point-based fusion was 1.978 mm, 2.788 mm and 1.926 mm. Significant differences existed in  $\Delta z$  for manual fusion and that for automatic fusion ( $P = 0.058$ ), in *MD* for manual fusion and that for automatic fusion ( $P = 0.087$ ), and in *MD* for manual fusion and that for registration point-based fusion ( $P = 0.069$ ). The *FI* for automatic fusion, manual fusion, and registration point-based fusion was 0.594, 0.520, and 0.549; the inter-algorithm differences were not significant ( $P = 0.290$ ).

**Conclusion:** The automatic fusion and the registration point-based fusion were more accurate than manual fusion, and therefore were recommended to be used in multimodal image fusion for oral and maxillofacial tumors.

© 2020 Published by Elsevier Ltd on behalf of European Association for Cranio-Maxillo-Facial Surgery.

## 1. Introduction

Oral and maxillofacial tumors are often located in a deep region of the orofacial area, and their diagnosis depends on radiologic examinations. Computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography–CT (PET-CT), which are the most commonly used modalities in clinical settings, have their own advantages and disadvantages. CT images can clearly show bony erosion of the tumor, but the contrast of the soft tissue involved is sometimes presented less clearly on the CT scan. MRI addresses the limitations of CT by providing fair contrast of the

soft tissue and has been widely used in deep parts of the oral and maxillofacial region, but it is difficult to evaluate the bony involvement. In contrast, PET can yield information regarding tumor metabolism, and provides better visualization of tumor extension over a complicated anatomical distortion, which often occurs in cases with a postoperative recurrent tumor. However, the slice thickness and spatial resolution of PET are often not as optimal as those in CT or MRI, which restrict the application of PET in the diagnosis of the primary tumor (Arya et al., 2014; Queiroz and Huellner, 2015; Sekine et al., 2017a).

With the development of imaging processing techniques and computer science, the multimodal image fusion method, which afforded the advantages associated with different imaging modalities, became available. By using a specific algorithm, two pairs of images can be integrated into one, and doctors can observe tumor infiltration clearly because of the enhanced contrast between

\* Corresponding author. Fax: (8610)62173402.

E-mail addresses: [kevinhlh@126.com](mailto:kevinhlh@126.com) (L.-H. Hu), [michaelzhang1016@126.com](mailto:michaelzhang1016@126.com) (W.-B. Zhang), [759964604@qq.com](mailto:759964604@qq.com) (Y. Yu), [pxpengxin@263.net](mailto:pxpengxin@263.net) (X. Peng).

muscular, vascular, and bony tissues. Moharir et al. first applied multimodal image fusion for the detection of oral and maxillofacial tumors (Moharir et al., 1998). Since then, similar methods have often been used in tumor diagnosis (Feichtinger et al., 2008; Kanda et al., 2013; Loeffelbein et al., 2014; Sekine et al., 2017a, 2017b), preoperative surgical planning of tumor ablation (Dai et al., 2012; Kraeima et al., 2015, 2018; Yu et al., 2017; Zrnc et al., 2018), intra-operative surgical navigation (Leong et al., 2006; Feichtinger et al., 2010; Chien et al., 2012; Yu et al., 2017; Zrnc et al., 2018), and tumor delineation for radiotherapy planning (Moore et al., 2004; Fortunati et al., 2014; Nix et al., 2017).

Retrospective multimodal image fusion (especially, imaging using a heterogeneous scanner) has the intrinsic drawbacks of inconsistent patient position, possible deformation of tumor volume, and inaccurate fusion. When performing multimodal image fusion, the operating staff needs to ensure high-quality and accurate fusion. Despite the wide use of multimodal image fusion in the diagnosis and treatment of oral and maxillofacial tumors, relatively few studies have focused on the accuracy of fusion for these tumors (Daisne et al., 2003; Wang et al., 2009; Al-Saleh et al., 2016). It is worth noting that the evaluation methods for fusion accuracy were different among these studies, which often focused on the overall image fusion accuracy by estimating the coordinates of a few pairs of anatomical landmarks and calculating the deviation between the coordinates. Such evaluation methods could afford fusion accuracy from an overall viewpoint; however, the fusion accuracy regarding gross tumor volume remained unclear. In some cases in which retrospective multimodal image fusion is based on data obtained from a heterogeneous scanner, the overall fusion accuracy could be low. However, in such cases, image fusion of two pairs of images can be achieved by co-registering the gross tumor volume, especially for radiotherapy planning. Therefore, a more comprehensive evaluation method is needed for fusion accuracy, including overall fusion accuracy and tumor volume fusion accuracy.

To this end, this study aimed to develop a revised evaluation method for accuracy of multimodal image fusion for oral and maxillofacial tumors and to apply this method to compare the accuracy of three commonly used multimodal image fusion algorithms, automatic fusion, manual fusion, and registration point-based fusion.

## 2. Materials and methods

### 2.1. Patients

In this study, we selected patients with a diagnosis of an oral and maxillofacial tumor who were referred to our department from January to August 2019. The inclusion criteria were as follows: (1) patients in whom the tumor was located in a deep oral area (gingiva of the posterior teeth, soft palate, base of the tongue, or parapharynx) or deep maxillofacial area (the maxilla, maxillary sinus, mandible ramus, zygomatic bone, skull base, or infratemporal fossa), and infiltrated at least two anatomical regions, regardless of whether it was the primary tumor or benign; (2) patients who had undergone at least two modalities of radiologic examination preoperatively and for whom complete Digital Imaging and Communications in Medicine (DICOM) files of at least two imaging modalities among CT, MRI, and PET-CT were available; and (3) patients whose radiological scans extended superiorly from above the orbit to inferiorly below the chin. Any patients who met those criteria could be included in this study, no matter what the parameter (slice thickness, spatial resolution, and etc) of their image scans were. The exclusion criteria were as follows: (1) patients whose image sets showed that they could not bite when being scanned, which meant that the dental arches of the patient were

not in stable occlusion, because of which the anatomical site of the tumor and surrounding tissues may shift; and (2) patients for whom the time interval between two radiological scans was more than 20 days, since there might be tumor deformation caused by growth. Altogether 17 patients and 18 pairs of image sets were included in this study.

### 2.2. Image acquisition and multimodal image fusion

DICOM files of patients' radiological examinations were imported into the navigation system (iPlan 3.0, BrainLAB, Feldkirchen, Germany). Three fusion algorithms provided by the navigation system were used for each pair of image sets: automatic fusion, manual fusion, and registration point-based fusion.

In automatic fusion, the operating staff set the region of interest (ROI) before image fusion. The ROI was set to oral and maxillofacial area. The principle of automatic fusion is the maximization of mutual information in the ROI. Two image sets were fused automatically based on anatomical structures common to both image sets without any manual manipulation.

In manual fusion, the operating staff used the "coarse" button to translate or rotate one image set to manually match the other image set as much as possible. The easily recognized bony structures of the oral and maxillofacial area (such as the sella turcica, orbital floor, maxillary sinus, or zygomatic arch) were preferentially co-registered when fusing two image sets.

In registration point-based fusion, at least three pairs of anatomical landmark points were marked on two image sets correspondingly as registration points by the operating staffs. The registration points included the base of the sella turcica, the center of the eyeballs, top of the pyriform aperture, or root apex. Registration point-based fusion was then automatically performed to make the distance between corresponding registration points as short as possible.

The image set with the lesser slice thickness was defined as the "fixed image set," which meant that the image set remained unchanged, and the other image set was designated as the "moving image set," as it would be translated or rotated while using the "fixed image set" as a reference (Oliveira and Tavares, 2014) (Fig. 1).

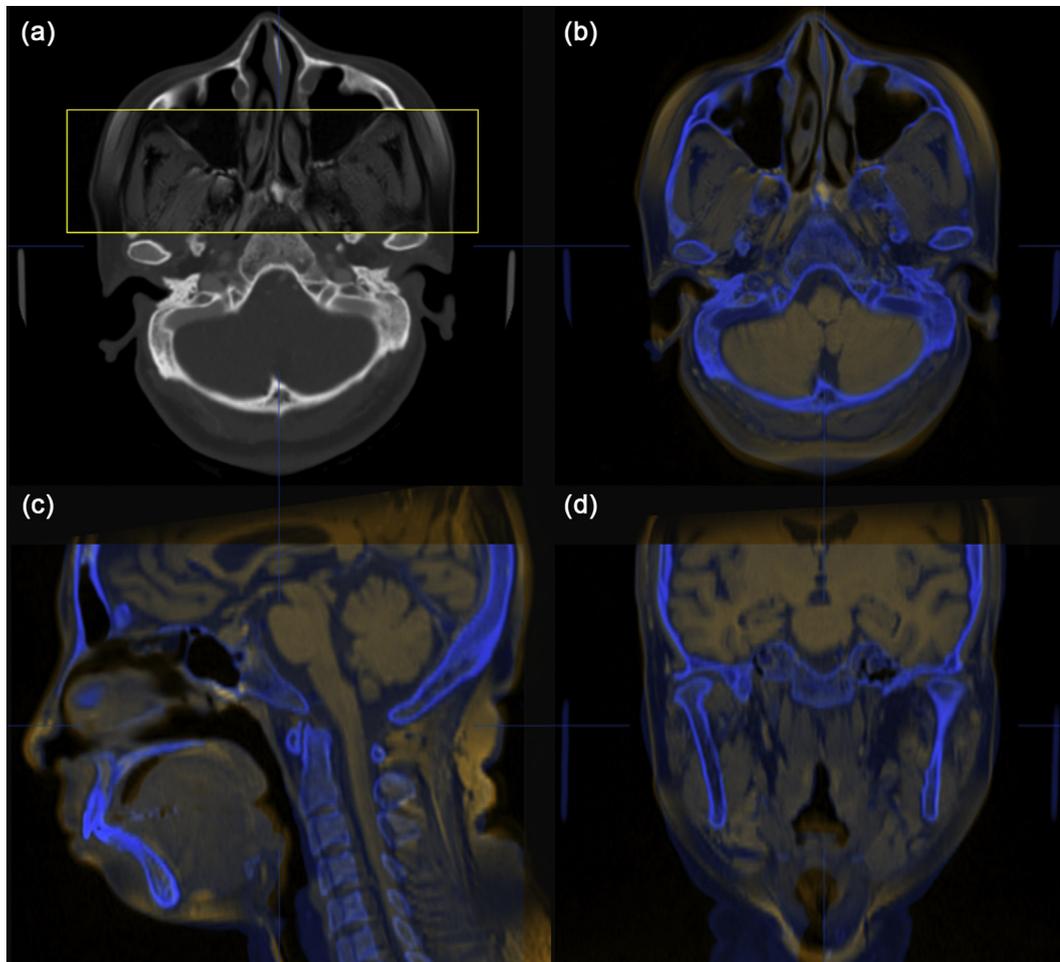
The fusion process was finished after consensus was reached by two oral and maxillofacial surgeons with more than 3 years of experience in using the iPlan 3.0 software. For each pair of image sets, two fusion projects were generated by using one fusion algorithm; altogether, six fusion projects were finally generated.

There were two kinds of multimodality image pairs for which PET-CT was used: PET-CT/contrast-enhanced CT (ceCT) and PET-CT/MRI. The multimodal image fusion process for these two pairs was the same as that of CT/MRI. In PET-CT/ceCT multimodal imaging, the vessels around the tumor along with the extent of the tumor could be clearly visualized, which could give surgeons additional information about the relationship between the tumor and vessels (Yu et al., 2017).

### 2.3. Evaluation of fusion accuracy

The fusion accuracy of one fusion project was evaluated twice by an operating staff member with who was a well-experienced oral and maxillofacial surgeon with more than 3 years of experience in using the iPlan 3.0 software. The evaluating staff did not participate in the fusion process, which might have helped avoid information bias.

Fusion accuracy included two aspects, overall fusion accuracy and tumor volume fusion accuracy.



**Fig. 1.** Result of multimodal image fusion. (a) The figure shows the result of CT/MRI image fusion. A smooth curve border can be observed in the yellow window. (b–d) The axial, sagittal and coronal view of fused images. The CT image was defined as a “fixed image” and presented in blue color, while the MRI image was designated as a “moving image” and presented in orange color.

#### 2.4. Overall fusion accuracy

Six pairs of anatomical landmarks were selected on both image sets as boundary marks of the overall image. The locations of the six anatomical landmarks are shown in Table 1 and Fig. 2. Using the coordinate system of the “fixed image set,” the operating staff recorded the three-dimensional coordinates of the six landmark pairs (Fig. 3). The coordinates of points on the “fixed image set” were  $(x_{i1}, y_{i1}, z_{i1})$  ( $i=1, 2, \dots, 6$ ), while those on the “moving image set” were  $(x_{i2}, y_{i2}, z_{i2})$  ( $i=1, 2, \dots, 6$ ). Overall fusion accuracy was evaluated using the mean values of the coordinate differences along the x-, y-, and z-axes ( $\Delta x$ ,  $\Delta y$ , and  $\Delta z$ ), mean deviation ( $MD$ ), and root mean square ( $RMS$ ) of the six landmark pairs on the two image sets. The values were calculated using the following formulas:

$$\Delta x = \left| \frac{\sum_{i=1}^6 (x_{i1} - x_{i2})}{6} \right|$$

$$\Delta y = \left| \frac{\sum_{i=1}^6 (y_{i1} - y_{i2})}{6} \right|$$

$$\Delta z = \left| \frac{\sum_{i=1}^6 (z_{i1} - z_{i2})}{6} \right|$$

$$MD = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$$

$$RMS = \sqrt{\frac{\sum_{i=1}^6 (x_{i1} - x_{i2})^2 + (y_{i1} - y_{i2})^2 + (z_{i1} - z_{i2})^2}{6}}$$

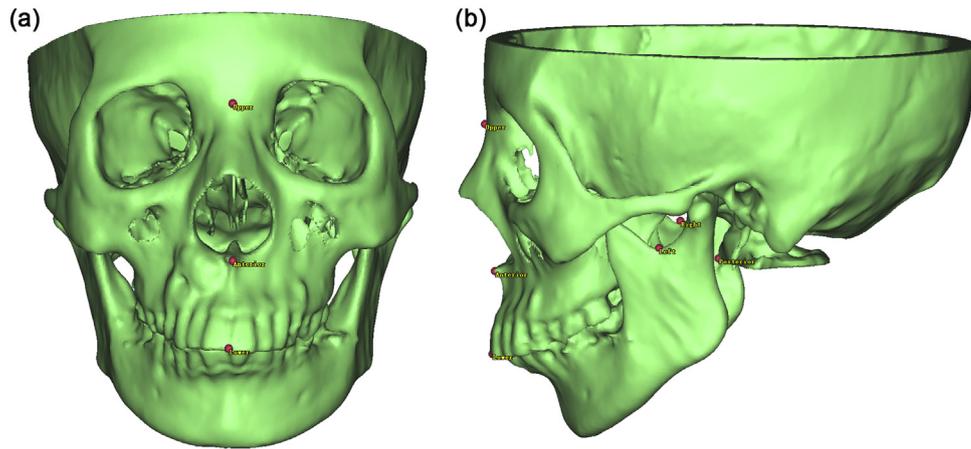
**Table 1**

Location of landmark points used for the evaluation of overall fusion accuracy.

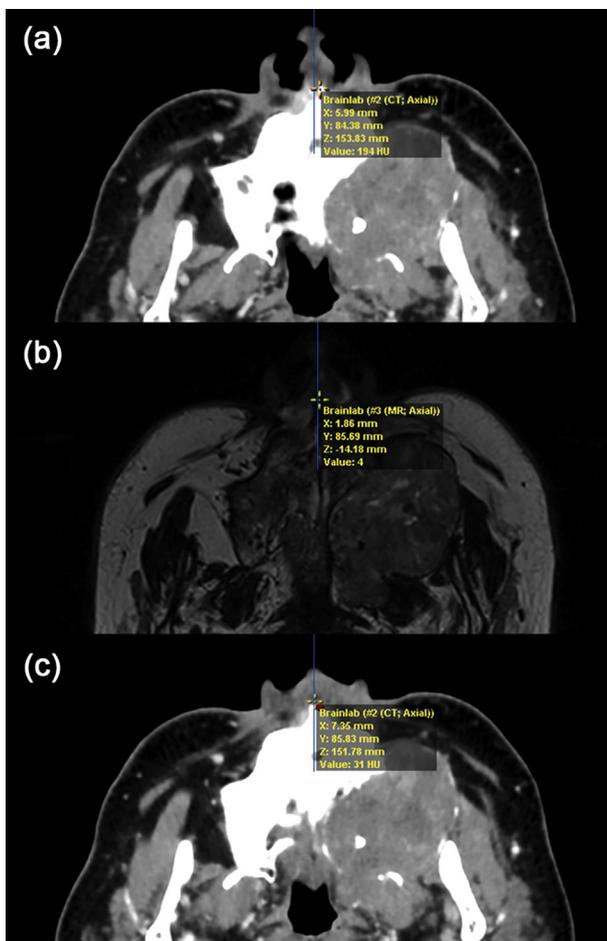
Boundary landmark	Location
(A) Upper	Interior point of nasion
(B) Lower	Tangency point of the upper central incisors
(C) Anterior	Former point of the anterior nasal spine
(D) Posterior	Former point of atlas
(E) Left	Tangency point of the left incisura mandibulae
(F) Right	Tangency point of the right incisura mandibulae

#### 2.5. Tumor volume fusion accuracy

After image fusion, tumor delineation was finished layer by layer on each single-modality image by using the “brush” and “eraser” functions in the “object creation” module. The tumor delineation



**Fig. 2.** Locations of landmark points used for evaluation of overall fusion accuracy. (a) The upper, the lower and the anterior boundary landmark. (b) The posterior, the left and the right boundary landmark.



**Fig. 3.** Process for evaluating overall fusion accuracy. The image shows the coordinates of the anterior anatomical landmark—the former point of the anterior nasal spine—on CT (a) and MRI images (b). By using the coordinate system of the “fixed image set” (CT), the coordinates of the landmark on “moving image set” (MRI) were measured as shown in (c).

was based on the distinctive radiologic characteristics of the tumor, including abnormal density on CT, high or low signal intensity on MRI, or hypermetabolism on PET. This process was finished under the guidance of a well-experienced radiologist. Three-dimensional

reconstruction of tumor objects was then completed, and the volume of the tumor was calculated by the iPlan software automatically. The volume of the tumor in the “fixed image” was named as  $V_F$  while that in the “moving image” was named as  $V_M$ . By using the “advanced manipulation” function, the intersected part of two objects was generated automatically, the volume of which was named as  $V_{F+M}$  (Fig. 4). Tumor volume fusion accuracy was indicated by the Fusion Index (*FI*) (Hu et al., 2018). *FI* was calculated using the following formula:

$$FI = \frac{V_{F+M}}{V_F} \times \frac{V_{F+M}}{V_M}$$

## 2.6. Statistical analysis

For each fusion algorithm, altogether four groups of results ( $\Delta x$ ,  $\Delta y$ ,  $\Delta z$ , *MD*, *RMS*, and *FI*) of one fused pair of image sets were obtained, and the mean value of four results of each indicator was calculated as the final result of fusion accuracy. For comparing the accuracy between three fusion algorithms, one-way analysis of variance (ANOVA) was performed using IBM SPSS Statistics v24.0 (IBM Corp., Armonk, NY) to confirm any significant difference in accuracy among the three fusion algorithms, and the LSD test was used as the post hoc test.

## 3. Results

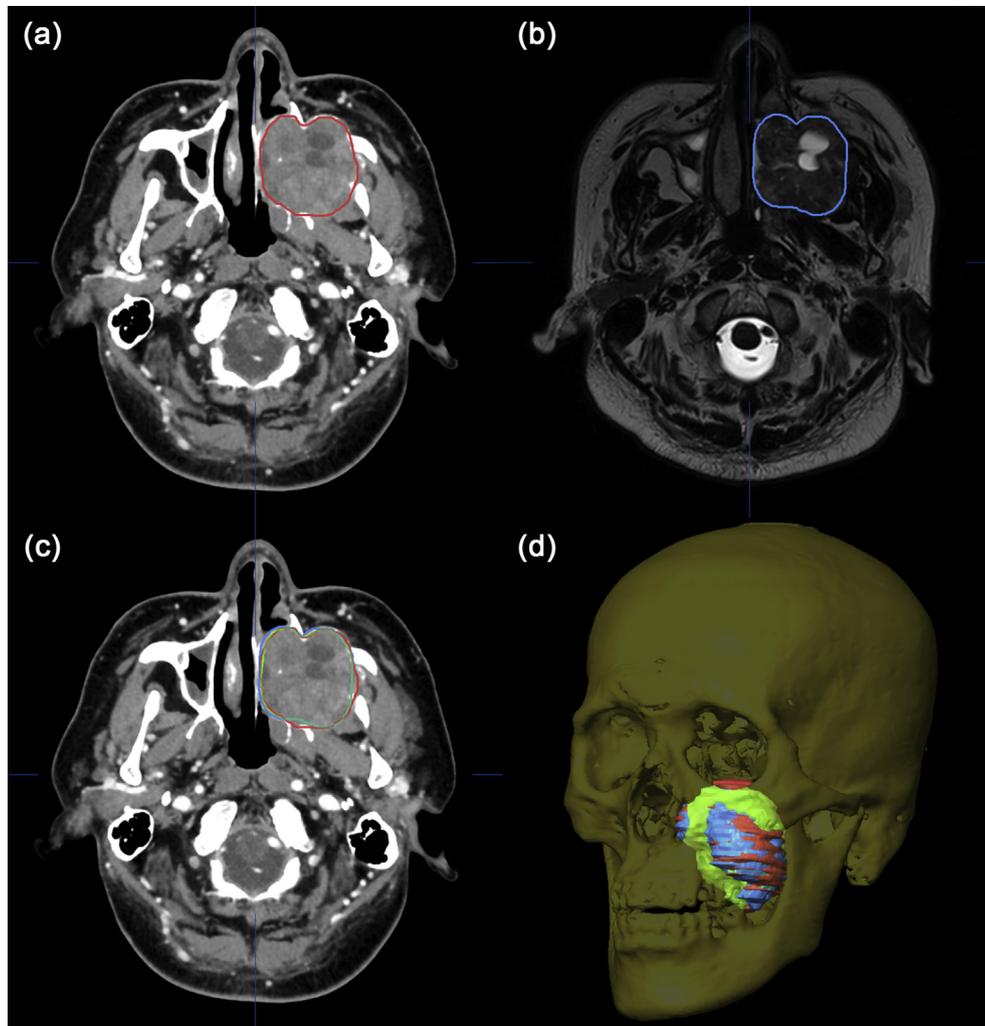
### 3.1. Overview of fusion results

A total of 18 pairs of image sets from 17 oral and maxillofacial tumor patients were fused in this study. By using three algorithms, 54 fusion projects were successfully carried out. Table 2 shows an overview of the 18 pairs of image sets.

### 3.2. Fusion accuracy of the three algorithms

The fusion accuracies of the three algorithms determined by using the evaluation methods mentioned above are shown in Table 3 and Figs. 5 and 6.

For overall fusion accuracy,  $\Delta x$  and  $\Delta y$  values of the three algorithms were less than 1.5 mm, as  $\Delta z$  were the values of automatic fusion and registration point-based fusion, while the  $\Delta z$  of manual



**Fig. 4.** Process of tumor delineation and evaluation of tumor volume fusion accuracy. (a) Tumor was delineated on the CT image as a red circle, and the volume of the tumor was  $V_F$ . (b) Tumor was delineated on the MRI image as a blue circle, and the volume was  $V_M$ . (c) The intersection of two objects of the tumor on the single-modality image was generated automatically by using “advanced manipulation,” as shown by a green circle, and the corresponding volume was  $V_{F+M}$ . (d) Three-dimensional reconstruction of three tumor objects.

**Table 2**

Overview of 18 pairs of image sets.

Categories	Amount
Modalities	
CT/MRI	11
PET-CT/ceCT	5
PET-CT/MRI	2
Location of tumor	
Maxilla	13
Mandible	5
Slice thickness of the fixed image set	
≤1.25 mm	15
>1.25 mm	3
Slice thickness of the moving image set	
≤2 mm	11
>2 mm	7
Dental artefacts	
Absence	15
Presence	3
<b>Total</b>	<b>18</b>

fusion was 1.864 mm. *MD* Values of automatic fusion and registration point-based fusion were less than 2 mm, and that of manual fusion was 2.788 mm. *RMS* Value of automatic fusion (3.776 mm) was less than that of manual fusion (4.518 mm) and registration point-based fusion (3.912 mm). One-way ANOVA showed no significant difference in overall fusion accuracy values among the three algorithms. Nevertheless, LSD post-hoc tests showed significant differences in the  $\Delta z$  value of manual fusion and automatic fusion, *MD* value of manual fusion and automatic fusion, and *MD* value of manual fusion and registration point-based fusion ( $P < 0.10$ ).

For tumor volume fusion accuracy, the *FI* values of the three algorithms were 0.594 (automatic fusion), 0.520 (manual fusion), and 0.549 (registration point-based fusion), while one-way ANOVA showed no significant difference in *FI* values among the three algorithms ( $P = 0.29$ ).

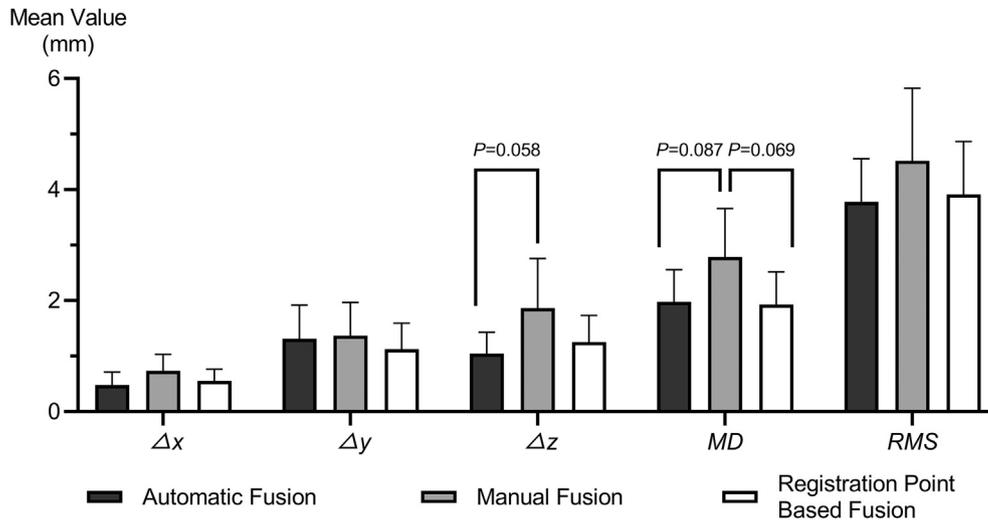
Therefore, the automatic fusion and registration point-based fusion algorithms were considered to show higher overall fusion accuracy compared to manual fusion, while the tumor volume

**Table 3**

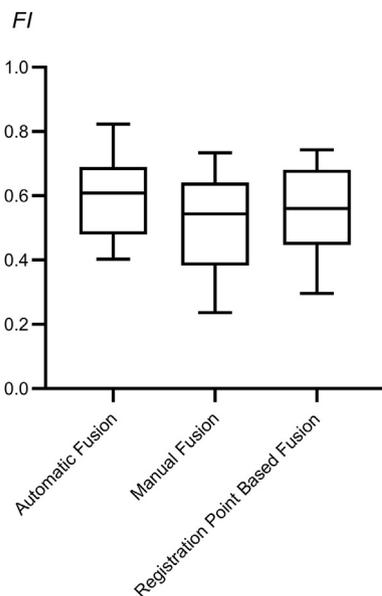
Fusion accuracy of the three fusion algorithms (mean values and 95% confidence intervals) (\*: &lt;0.10).

Fusion algorithms		AF	MF	RPBF	P	$P_{AF-MF}$	$P_{MF-RPBF}$	$P_{AF-RPBF}$
OFA	$\Delta x$ (mm)	0.477 (0.239–0.714)	0.735 (0.440–1.031)	0.553 (0.337–0.770)	0.298	0.132	0.286	0.652
	$\Delta y$ (mm)	1.311 (0.707–1.915)	1.368 (0.769–1.967)	1.122 (0.652–1.591)	0.792	0.881	0.516	0.617
	$\Delta z$ (mm)	1.049 (0.672–1.426)	1.864 (0.969–2.759)	1.254 (0.774–1.734)	0.140	0.058*	0.152	0.627
	MD (mm)	1.978 (1.397–2.558)	2.788 (1.921–3.656)	1.926 (1.336–2.516)	0.124	0.087*	0.069*	0.912
	RMS (mm)	3.776 (3.000–4.551)	4.518 (3.209–5.828)	3.912 (2.961–4.862)	0.430	0.290	0.386	0.846
TVFA	FI	0.594 (0.531–0.657)	0.520 (0.444–0.595)	0.549 (0.479–0.620)	0.290	0.120	0.529	0.348

OFA: Overall fusion accuracy; TVFA: Tumor volume fusion accuracy; MD: Mean deviation; RMS: Root mean square; FI: Fusion Index; AF: Automatic fusion; MF: Manual fusion; RPBF: Registration point-based fusion.



**Fig. 5.** Overall fusion accuracy of the three fusion algorithms. The  $\Delta z$  value of manual fusion was significantly higher than that of automatic fusion, and the MD value of manual fusion was also significantly higher than that of automatic fusion and registration point-based fusion ( $P < 0.10$ ).



**Fig. 6.** Tumor volume fusion accuracy of the three fusion algorithms. No significant difference was observed among the three fusion algorithms ( $P = 0.290$ ).

fusion accuracy for the three fusion algorithms were approximative.

#### 4. Discussion

Single-modality imaging techniques such as CT, MRI, and PET have been widely used for the diagnosis of oral and maxillofacial tumors, but their limitations and drawbacks have also been revealed by previous studies (Wang et al., 2009; Feichtinger et al., 2010; Arya et al., 2014). By integrating single-modality images into the same coordinate system, multimodal image fusion combined the advantages of single-modality images and showed its superiority in better visualization of tumor infiltration, especially for tumors located in deep anatomical regions and those with large scale (Queiroz and Huellner, 2015; Sekine et al., 2017a). For all of the tumor image data included in our study, the tumors were in the deep oral and maxillofacial areas and had infiltrated to at least two anatomical regions, where we believed multimodal image fusion could play a crucial role.

As a part of digital surgery techniques, multimodal image fusion could be the basis for other virtual planning tools such as tumor contouring, ablation design, and surgical navigation. Therefore, quality assurance is a key topic in multimodal image fusion projects. Generally speaking, comprehensive quality assurance should

receive attention from the image scanning stage, including stabilization of patient position, noise reduction, and minimization of the interval between different image scanning modalities (Mutic et al., 2001). Since our study covered retrospective heterogeneous multimodal image fusion, in which the initial stage for the operating staff was image fusion, not image scanning, we could only modify the inclusion and exclusion criteria to allow images with relatively high quality of scanning in our study; for example, images in which the patient was not in centric occlusion during scanning were excluded, and if the time interval between the two image scans was over 20 days, the image sets were excluded to avoid deformation of the tumor. Considering the high quality of image scanning, image fusion could be expected to show relatively high accuracy.

There are many commercially available software platforms with multimodal image fusion function, for example, iPlan (BrainLAB, Feldkirchen, Germany), Stryker Navigation, and Nucletron Oncentra Masterplan. iPlan from the company BrainLAB, the navigation platform used in our department, contains an “image fusion” module. After image fusion and 3D object creation, the navigation plan can be easily created in iPlan without transferring any data to other planning platforms. A few of studies reported on the accuracy of image fusion using BrainLAB iPlan CMF software. Thani et al. (2011) reported the accuracy of CT/MRI automatic fusion for determining the location of deep brain stimulation (DBS) electrode in neurosurgery, and the discrepancy between the CT and MRI fusion was  $1.6 \pm 0.2$  mm. Yu et al. (2017) reported the accuracy of image fusion between FDG-PET/CT and contrast enhanced CT for only one patient who experienced recurrent maxillary SCC, in which the registration point based fusion was used. The shift of the fused image between two image sets was  $0.77 \pm 0.53$  mm. The current situation is that no studies have tested the accuracy of image fusion by iPlan CMF for oral and maxillofacial tumors based on a relatively large sample size. This was the reason why we proposed a fusion accuracy evaluation method by using iPlan software.

The image sets with thinner slice thickness was set as “fixed image set.” Although the “fixed image set” fluctuated among CT, MRI, and PET, the result of image fusion would not alter if the “fixed image set” and “moving image set” interchanged. In BrainLAB iPlan CMF, the slice thickness of the fused image equals that of the image set with less slice thickness, and the missed slices of another image set can be enriched automatically by multiplanar reconstruction of the software. If the image set with higher slice thickness was set as “fixed image set,” the result of image fusion remained unchanged, and the fusion accuracy would not differ. We set a standard for the selection of “fixed image set” in this study only for the convenience of further explanation.

In the “image fusion” module of iPlan, automatic fusion, manual fusion, and registration point–based fusion were the commonly used fusion algorithms. The overall fusion accuracy of the three fusion algorithms was around 2 mm, which was close to that reported in previous studies (Mutic et al., 2001; Wang et al., 2009). According to the User Guide of iPlan, automatic fusion was finished automatically based on the theory of maximization of mutual information. Previous studies have reported its advantages over surface-based fusion algorithms and manual fusion algorithms (West et al., 1999; Dean et al., 2012), including lower time requirement, no requirement of interactions, relatively steady consistency, and satisfactory accuracy. In this study, the performance of automatic fusion and registration point–based fusion was slightly better than that of manual fusion, which agreed with the results of previous studies. For manual fusion, the dominant deviation was from  $\Delta z$ , which meant that craniocaudal deviation played an important role in the MD for manual fusion. This was mainly

because the registration of images in the axial view was easier than that in the sagittal or coronal views.

The method for evaluating the accuracy of multimodal image fusion can be classified as qualitative and quantitative. Dai et al. used “grade of clarity” as a mean of qualitative assessment of the accuracy of CT/MRI image fusion for seven patients of jaw tumor (Dai et al., 2012). Such method was subjective and not quantitative, which restricted its application. As for quantitative evaluating method, Ulin et al. used a benchmark case (a pediatric low-grade glioma) developed by the Quality Assurance Review Center, to assess the variability (accuracy) of CT/MRI image fusion (Ulin et al., 2010). The average error of manual fusion in their study (1.1 mm) was obviously smaller than ours (2.788 mm). Taking the standard deviation of their study (1 standard deviation = 2.2 mm) into consideration, the result of ours was acceptable. The accuracy of manual fusion depends greatly on the operating staff’s ability to recognize anatomical structure and their experience with radiological diagnosis, which means that the accuracy of manual fusion could fluctuate a lot. Meanwhile, the average error of registration point–based fusion in their study (3.1 mm) was obviously higher than ours (1.926 mm). Since Ulin et al.’s study did not reveal more details about their manipulation of registration point–based fusion, we could not comment regarding how this happened. For one thing, our registration points were located mainly at maxilla and were steady for image fusion, which could be the reason why our results looked better. Besides, a few in vivo and in vitro studies have also attempted to quantifiably evaluate fusion accuracy. Daisne et al. showed that the accuracy was within 2.1 mm for 50% of fusions and within 5.8 mm for 95% of the fusions performed in four head-and-neck tumor patients (Daisne et al., 2003). Wang et al. indicated that the accuracy of CT/MRI image fusion was 0.56–1.04 mm three-dimensionally in nasopharyngeal carcinoma patients (Wang et al., 2009). Al-Saleh et al. verified the accuracy of CBCT/MRI fusion in vitro, which was  $0.21 \pm 1.24$  mm (Al-Saleh et al., 2016).

In most of these studies, fusion accuracy was indicated by the Euclidean distance (ED) between central points of the same volume of interest (VOI) on two single-modality images. This method is relatively convenient because coordinates of only two points need to be estimated in one VOI. Nevertheless, the extent to which the central point of the VOI represents the overall VOI remains controversial; ED could thus show only the offset value, and could not reveal the deviation of specific landmarks three-dimensionally. Besides, the process to identify the central point of a VOI in the human body could be difficult and ambiguous if the VOI had an irregular shape, for example, in the maxilla or mandible. With respect to multimodal image fusion for oral and maxillofacial tumors, this “one-point evaluation method” may not be as efficient as the “six-point evaluation method” described above. The six anatomical landmarks chosen in this study were evenly distributed among oral and maxillofacial areas; could represent the upper, lower, anterior, posterior, left, and right boundaries of the overall image; and could be easily distinguished even by a junior resident.

Notably, the six anatomical landmarks chosen for evaluating the fusion accuracy in this study were all bony landmarks originating from the maxilla, mandible, and vertebra, which we thought could be the basis for selection of the landmark points. Bony landmarks show less deformation than landmarks of soft tissue; therefore, their locations will be relatively consistent on different modalities of image sets. In fact, the anatomical landmarks used to evaluate overall fusion accuracy may be altered under special circumstances—such as in edentulous patients or in postoperative patients with defects or distortions in some anatomical structures—as long as these landmarks are bony and evenly distributed in the oral and maxillofacial region. In addition, repeatedly marking landmarks could also decrease the possibility of incorrect marking.

However, it is questionable as to whether fusion accuracy can be completely represented by only one indicator. Oliveira et al. suggested that the drawbacks of measuring the coordinates of specific landmarks were greatly dependent on the fiducial localization, and that incorrect localization of the anatomical landmarks could yield inaccurate results (Oliveira and Tavares, 2014). In essence, retrospective multimodal image fusion could easily result in an intrinsic deviation because of the distortion of the image scanner, different image scan parameters and patient positions, and so forth. Even when two pairs of single-modality images are registered and fused under the same coordinate system, the coordinates of specific landmarks may differ substantially between the two pairs of single-modality images; thus, fusion may show low accuracy if it relies only on evaluation of overall fusion accuracy. An evaluation method using data from another dimension was therefore needed to make the evaluation more comprehensive. In this study, we introduced tumor volume fusion accuracy to complement the drawbacks of the “overall fusion accuracy–based” approach. Theoretically, tumor volume on one image would be nearly equal to the tumor volume on the other image of the same patient. The larger the intersection of the two tumor volumes, the higher the *FI* (Fig. 7). We performed linear correlation analysis between *MD* and *FI* by using SPSS software (SPSS Inc., Chicgo, IL). Pearson correlation between *MD* and *FI* was  $-0.415$ , the significance of which was  $0.002$ . A moderate negative correlation could be observed between *MD* and *FI*, which meant that *FI* would be higher if *MD* became smaller; thus, a moderate positive correlation existed between overall fusion accuracy and tumor volume fusion accuracy. We could conclude that for a fusion project, the tumor volume fusion accuracy would be normally be acceptable if the overall fusion accuracy was high.

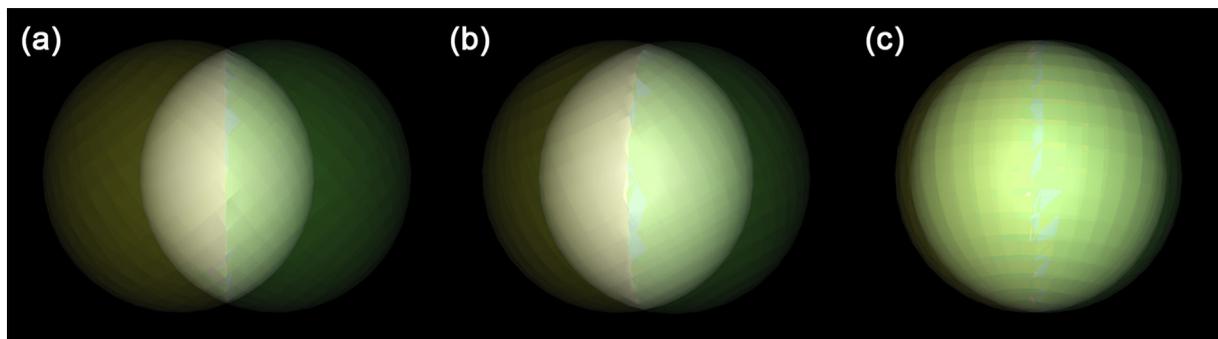
When evaluating tumor volume fusion accuracy, the process of tumor delineation was completed under the guidance of a well-experienced radiologist. In fact, the volume of tumor on different image modalities would not be the same, since the radiologic characteristics of tumor on different image modalities were not the same, and they were marked manually by operating staff. However, the volume of tumor on different image modalities would not differ a lot because they originate from the same tumor. As for repeatability coefficient, the intraclass correlation (ICC) of Fusion Index was  $0.699$ , which meant the evaluation method for “Tumor Volume Fusion Accuracy” was moderately repeatable.

Nevertheless, the tumor delineation process should receive more attention, since *FI* was sensitive to changes in tumor volume on the single-modality image. Different types of tumors have their own unique features on radiographic manifestation, and the criteria for tumor infiltration on different image modalities are not identical. As a result, definitions of the range of the tumor could

differ across image modalities for different types of tumors. Therefore, we would like to propose the following suggestions for the use of *FI*: (1) We strongly recommend that for evaluating tumor volume fusion accuracy, tumor delineation should be performed by a well-experienced surgeon, radiologist, or physicist, and *FI* should be repeatedly measured to minimize the random error. (2) *FI* is a secondary indicator of fusion accuracy, while the coordinate difference and the mean deviation should be measured first. When a patient’s position differs between two image sets, or noise and artifacts are observed in fused images, *FI* could accurately reveal the fusion accuracy. (3) For the fusion of a tumor with small volume, *FI* could fluctuate greatly with the movement of the image sets. Hence, better applicability of *FI* would be observed in the fusion of image sets of large tumors.

Automatic fusion, manual fusion, and registration point–based fusion were three basic fusion algorithms in multimodal image fusion. To ascertain the accuracy of every single algorithm was the aim of this study, so we did not combine two algorithms together in this study in order to avoid the interaction of different algorithms. When applied to clinical work, multimodal image fusion was usually completed by automatic fusion and, if necessary, followed with manual fusion in case obvious deviation between different images was observed. Such a process was completed with the cooperation of a team consisting of surgeon, radiologist, and engineer, and so was the accuracy evaluation. In this study, the fusion process and evaluating process were under the guidance of a well-experienced radiologist. We have many excellent surgeons with lots of experience in using the iPlan 3.0 software in our department, since they had been trained in software manipulation systematically by an engineer team in our hospital.

The fusion algorithms of BrainLAB iPlan software used in our study were essentially rigid image fusion. We have noticed the gradually developing trend of using deformable image fusion in the head-and-neck area. Some previous studies evaluated the accuracy of deformable image fusion for the purpose of radiotherapeutic treatment planning in the brain (Fortunati et al., 2014), pelvis (Zambrano et al., 2013), or prostate. Although Fortunati et al. showed the superiority of CT/MRI deformable image fusion over rigid fusion in head-and-neck tumors (Fortunati et al., 2014), we still believe that rigid fusion algorithms can play an important role in treatment of oral and maxillofacial tumors because of the more stable anatomy and less organ deformation in head and neck area. Furthermore, we performed multimodal image fusion for the purpose of virtual surgical planning, not for radiotherapeutic planning, which is sensitive to distortions of the volume of the organs or tumors. In comparison to deformable image fusion, rigid image fusion is more accessible and less time-consuming; therefore, it can



**Fig. 7.** Schematic diagram of the Fusion Index (*FI*). Yellow and green transparent spheres represent the tumor volume on a single-modality image, and the intersected parts of the two spheres represent the intersected tumor volumes of the two single modalities and are colored in silver. The more the two spheres intersect, the larger the *FI* will be. (a) *FI* = 0.3. (b) *FI* = 0.5. (c) *FI* = 0.8.

be more conveniently used for virtual surgical planning. Besides, BrainLAB was a software platform that could not only finish an accurate image fusion project but could also act as a navigation system, which meant that virtual surgical planning and navigation project could be finished one-step after image fusion. This could decrease the possible error when planning data was transferred between different platforms. As far as we know, other software packages that enable deformable image fusion cannot achieve such a one-step surgical planning function. Nevertheless, the evaluation method in this study is also applicable to deformable image fusion. Evaluation of tumor volume fusion accuracy could be crucial for deformable image fusion, considering the possible changes in tumor volume. We are greatly interested in evaluating the accuracy of deformable image fusion by using the method in this study.

Finally, we believe that there were some factors other than the fusion algorithms that influenced the fusion accuracy. We noticed in some fusion projects containing the PET-CT modality that the performance of automatic fusion was not good, while manual fusion and registration point-based fusion showed better performance. This reminded us that there may be some properties of PET images that could influence the accuracy of auto-detection by fusion software. Whether the low spatial resolution and relatively large slice thickness of PET-CT modality would influence the performance of auto-fusion remained a mystery. Besides, since our study aims to raise a revised evaluation method of fusion accuracy, we would like to include image sets as much as possible in order to enrich the sample size of our study. Any patient who had data from a CT/MRI/PET-CT scan could be included in this study, whether or not that patient was diagnosed with a malignant tumor. Because of the relatively low visibility of malignant tumors in image scanning, the tumor volume was not easily defined, and therefore the tumor volume fusion accuracy may be low for malignant tumors. Furthermore, the sample size of this study was still relatively small, and it is necessary to include more cases to further validate our conclusions. Further in-depth studies need to be carried out to examine the factors that could possibly influence fusion accuracy, such as slice thickness, spatial resolution, nature of the tumor, and tumor volume, which is underway by researchers in our institution.

## 5. Conclusion

A revised method that included both overall fusion accuracy and tumor volume fusion accuracy was proposed to evaluate the accuracy of multimodal image fusion for oral and maxillofacial tumors. With this method, the automatic fusion and registration point-based fusion were more accurate than manual fusion, and therefore were recommended for use in multimodal image fusion for oral and maxillofacial tumors.

## Ethics approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of our institutional review board and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declaration of Competing Interest

None.

## Acknowledgment

We appreciate the professional editor at Elixigen for revising and modifying the English language of this manuscript.

## References

- Al-Saleh MA, Punithakumar K, Jaremko JL, Alsufyani NA, Boulanger P, Major PW: Accuracy of magnetic resonance imaging-cone beam computed tomography rigid registration of the head: an in-vitro study. *Oral Surg Oral Med Oral Pathol Oral Radiol* 121: 316–321, 2016
- Arya S, Rane P, Deshmukh A: Oral cavity squamous cell carcinoma: role of pre-treatment imaging and its influence on management. *Clin Radiol* 69: 916–930, 2014
- Chien CY, Chuang HC, Huang SH, Lin WC, Huang HY: A pilot study of segmental mandibulectomy with surgical navigation using fluorine-18 fluorodeoxyglucose positron-emission tomography/computed tomography. *Laryngoscope* 122: 2205–2209, 2012
- Dai JW, Wang XD, Dong YF, Yu HB, Yang DL, Shen GF: Two- and three-dimensional models for the visualization of jaw tumors based on CT-MRI image fusion. *J Craniofac Surg* 23: 502–508, 2012
- Daisne J-F, Sibomana M, Bol A, Cosnard G, Lonnet M, Grégoire V: Evaluation of a multimodality image (CT, MRI and PET) coregistration procedure on phantom and head and neck cancer patients: accuracy, reproducibility and consistency. *Radiother Oncol* 69: 237–245, 2003
- Dean CJ, Sykes JR, Cooper RA, Hatfield P, Carey B, Swift S, et al: An evaluation of four CT-MRI co-registration techniques for radiotherapy treatment planning of prone rectal cancer patients. *Br J Radiol* 85: 61–68, 2012
- Feichtinger M, Aigner RM, Karcher H: F-18 positron emission tomography and computed tomography image-fusion for image-guided detection of local recurrence in patients with head and neck cancer using a 3-dimensional navigation system: a preliminary report. *J Oral Maxillofac Surg* 66: 193–200, 2008
- Feichtinger M, Pau M, Zemmann W, Aigner RM, Karcher H: Intraoperative control of resection margins in advanced head and neck cancer using a 3D-navigation system based on PET/CT image fusion. *J Craniomaxillofac Surg* 38: 589–594, 2010
- Fortunati V, Verhaart RF, Angeloni F, van der Lugt A, Niessen WJ, Veenland JF, et al: Feasibility of multimodal deformable registration for head and neck tumor treatment planning. *Int J Radiat Oncol Biol Phys* 90: 85–93, 2014
- Hu J, Dong XQ, Lin Q, Zhang Y, Zhang JF, Zhang Y: Medical image registration and fusion technology applied in cancer radiotherapy. *Chin Med Equip J* 39: 75–78, 2018
- Kanda T, Kitajima K, Suenaga Y, Konishi J, Sasaki R, Morimoto K, et al: Value of retrospective image fusion of (1)(8)F-FDG PET and MRI for preoperative staging of head and neck cancer: comparison with PET/CT and contrast-enhanced neck MRI. *Eur J Radiol* 82: 2005–2010, 2013
- Kraeima J, Dorgelo B, Gulbitti HA, Steenbakkers R, Schepman KP, Roodenburg JLN, et al: Multi-modality 3D mandibular resection planning in head and neck cancer using CT and MRI data fusion: a clinical series. *Oral Oncol* 81: 22–28, 2018
- Kraeima J, Schepers RH, van Ooijen PM, Steenbakkers RJ, Roodenburg JL, Witjes MJ: Integration of oncologic margins in three-dimensional virtual planning for head and neck surgery, including a validation of the software pathway. *J Craniomaxillofac Surg* 43: 1374–1379, 2015
- Leong JL, Batra PS, Citardi MJ: CT-MR image fusion for the management of skull base lesions. *Otolaryngol Head Neck Surg* 134: 868–876, 2006
- Loeffelbein DJ, Souvatzoglou M, Wankerl V, Dinges J, Ritschl LM, Mücke T, et al: Diagnostic value of retrospective PET-MRI fusion in head-and-neck cancer. *BMC Cancer* 14: 846, 2014
- Moharir VM, Fried MP, Vernick DM, Janecka IP, Zahajsky J, Hsu LG, et al: Computer-assisted three-dimensional reconstruction of head and neck tumors. *Laryngoscope* 108: 1592–1598, 1998
- Moore CS, Liney GP, Beavis AW: Quality assurance of registration of CT and MRI data sets for treatment planning of radiotherapy for head and neck cancers. *J Appl Clin Med Phys* 5: 23–33, 2004
- Mutic S, Dempsey JF, Bosch WR, Low DA, Drzymala RE, Chao KSC, et al: Multimodality image registration quality assurance for conformal three-dimensional treatment planning. *Int J Radiat Oncol Biol Phys* 51: 255–260, 2001
- Nix MG, Prestwich RJD, Speight R: Automated, reference-free local error assessment of multimodal deformable image registration for radiotherapy in the head and neck. *Radiother Oncol* 125: 478–484, 2017
- Oliveira FP, Tavares JM: Medical image registration: a review. *Comput Methods Biomech Biomed Eng* 17: 73–93, 2014
- Queiroz MA, Huellner MW: PET/MR in cancers of the head and neck. *Semin Nucl Med* 45: 248–265, 2015
- Sekine T, Barbosa FG, Delso G, Burger IA, Stolzmann P, Ter Voert EE, et al: Local resectability assessment of head and neck cancer: positron emission tomography/MRI versus positron emission tomography/CT. *Head Neck* 39: 1550–1558, 2017a
- Sekine T, de Galiza Barbosa F, Kuhn FP, Burger IA, Stolzmann P, Huber GF, et al: PET+MR versus PET/CT in the initial staging of head and neck cancer, using a trimodality PET/CT+MR system. *Clin Imaging* 42: 232–239, 2017b

- Thani NB, Bala A, Swann GB, Lind CR: Accuracy of postoperative computed tomography and magnetic resonance image fusion for assessing deep brain stimulation electrodes. *Neurosurgery* 69: 207–214, 2011 ; discussion 214
- Ulin K, Urie MM, Cherlow JM: Results of a multi-institutional benchmark test for cranial CT/MR image registration. *Int J Radiat Oncol Biol Phys* 77: 1584–1589, 2010
- Wang XS, Li LG, Hu CS, Qiu JJ, Xu ZY, Feng Y: A comparative study of three CT and MRI registration algorithms in nasopharyngeal carcinoma. *J Appl Clin Med Phys* 10: 3–10, 2009
- West J, Fitzpatrick JM, Wang MY, Dawant BM, Maurer Jr CR, Kessler RM, et al: Retrospective intermodality registration techniques for images of the head: surface-based versus volume-based. *IEEE Trans Med Imaging* 18: 144–150, 1999
- Yu Y, Zhang WB, Liu XJ, Guo CB, Yu CY, Peng X: Three-dimensional image fusion of (18)f-fluorodeoxyglucose-positron emission tomography/computed tomography and contrast-enhanced computed tomography for computer-assisted planning of maxillectomy of recurrent maxillary squamous cell carcinoma and defect reconstruction. *J Oral Maxillofac Surg* 75: 1301, 2017
- Zambrano V, Furtado H, Fabri D, Lutgendorf-Caucig C, Gora J, Stock M, et al: Performance validation of deformable image registration in the pelvic region. *J Radiat Res* 54: i120–i128, 2013
- Zrnc TA, Wallner J, Zemann W, Pau M, Gstettner C, Brcic L, et al: Assessment of tumor margins in head and neck cancer using a 3D-navigation system based on PET/CT image-fusion—a pilot study. *J Craniomaxillofac Surg* 46: 617–623, 2018