View-Dependent Text Placement for Augmented Reality using Offset Surfaces

Xuetong Sun, Mukul Agarwal, Hsueh-Chien Cheng and Amitabh Varshney

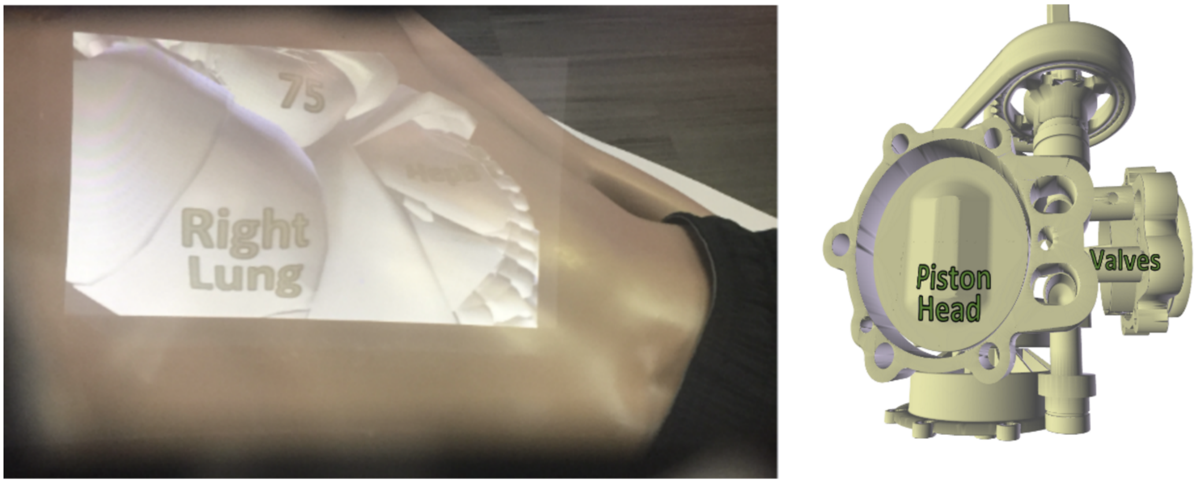


Fig. 1. We visualize view-dependent text that has been wrapped on the offset surfaces placed on an augmented reality view of a mannequin and on the model of an engine. Using our technique, one can visualize text labels such as *right lung*, *piston head*, as well as real-time text information such as the heart rate shown as 75 beats per minute.

Abstract— Text can be used to enhance a large range of augmented reality applications by providing timely context-aware information to the users. Current methods for text placement in augmented reality applications are typically static (i.e. view-independent) and involve placing the text label away from the labeled object using lines or arrows to establish the correspondence between the objects and the labels. In this paper, we present a new method to improve text placement on objects by enhancing object-label association, visibility, readability, as well as making better use of the display real estate. We build upon previous work in visualization literature that uses an offset surface for placing text that wraps around an object and enhance it to real-time, view-dependent display of time-critical information. We have validated our approach by using an optical see-through display with a virtual surgery application on a human mannequin. We envision our system finding use in a variety of augmented reality applications with a need for real-time, view dependent text display including mechanical assembly, repair and maintenance, augmented navigation and surgical procedures.

Index Terms—Augmented reality, text placement, view dependent visualization

1 INTRODUCTION

Modern tracking, display, and graphics processing technologies have greatly improved the usability of virtual and augmented reality. Various recently introduced virtual reality headsets, developed by major companies such as Oculus Rift (Facebook/Oculus), Vive (HTC and Valve) and Playstation VR (Sony), have shown significant improvement over previous devices in field of view, display resolution, and tracking precision. These headsets place users in an immersive virtual environment, where users can see and interact with virtual objects in natural and intuitive ways. Microsoft has recently released an augmented-reality headset, Hololens. This device and other augmented reality systems (e.g. metaAR, Vuzix) allow users to see augmented real-life objects. These augmented reality devices use semi-opaque displays to present directly to the users additional information about the real world to augment their information of the world around them.

Augmented reality technology has immense real-life applications in various fields. One of the benefits of augmented reality is the presentation of critical information where it is needed. In the medical field, augmented reality can directly display information such as pulse rate, blood pressure, location of organs, and other relevant medical conditions of a patient during a surgery. Using augmented reality, a headset can show medical images directly in a surgeon's field of view. Another application can be found in mechanical CAD part assembly or maintenance, where the specifications and functions of each component can be shown around the actual objects. By making critical information more easily accessible, users can focus on their task instead of consulting another data source which may involve visual context switching. In spite of these applications and recent advances, displaying even a moderate amount of textual content remains a challenging problem in augmented reality environments.

The criteria of text placement in augmented reality include well-defined object-text association, text visibility, and readability. In a complex scene where multiple text labels are associated with multiple objects, text placement must address the need for a clear correspondence between the text and the corresponding object to avoid ambiguity. The visibility constraint requires the text, if relevant to the current task, to be visible to the users and not be occluded by other real or

- *Xuetong Sun email xtsun@umiacs.umd.edu*
- *Mukul Agarwal email amukul@umiacs.umd.edu*
- *Hsueh-Chien Cheng email cheng@umiacs.umd.edu*
- *Amitabh Varshney email varshney@umiacs.umd.edu*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org.
Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx/

augmented objects. The text should also have a high contrast against the underlying objects such that they can be seen clearly. Readability requires the text to be presented in a way easily understandable by users. Some of the contributing factors of readability include font size and text orientation. Because augmented reality systems are deployed to alleviate the cognitive burden of users when performing challenging tasks, the rendering of text in augmented reality should make it easy for users to read with as little effort as possible.

Conventionally, text in augmented reality is shown using billboards rendered in the screen space [2, 26, 16]. This technique is sufficient for presenting information to the user in most simple scenes. In a simple scene with few objects and ample spaces among objects, object-label association is readily achieved, and could be further strengthened, by linking the text and the corresponding object with an arrow. Visibility and readability are trivial because text is shown using billboards; it can be made visible and upright to the users at all times. Nevertheless, as the scene grows in complexity, the text may become increasingly cluttered or occluded. Although special care has been taken to manage text placement in such scenarios [2, 1, 5], in general billboard-based approaches are severely limited by scene complexity. Another major disadvantage of billboard text display is that it breaks immersion.

This paper aims to improve text placement in an augmented reality environment. We build upon previous work on offset surfaces, ambient occlusion, and text scaffolds [13] to present an interactive solution for augmented reality applications. Placing text by carefully wrapping it around offset surfaces of key object regions achieves object-label association naturally by spatial proximity. Furthermore, the shape of the text resembles that of the object being augmented. Because the text appears to be affixed onto the objects, they do not occlude objects that are visible before the augmentation. The offset surfaces are crucial because they provide an appropriately displaced, yet proximal surface for text placement.

Our goal is reliably delivering augmented text information to users when performing time-critical tasks. In such scenarios, the text that conveys important information must be visible at all times. When the viewpoint changes as users move around, the text should remain visible as much as possible. To address this requirement, we calculate the visibility of different parts on the offset surface and place text on the parts with high visibility. In addition, we orient the text according to a user's view so that it always appears upright to the user, thereby improving its readability.

We show our proposed system in a surgery scenario; we display a 3D model of organs and overlay relevant information on top of a mannequin. Our system is also applicable to other fields that require real-time view-dependent display of information, such as industrial assembly and maintenance, augmented navigation, and medical training, and surgical procedures.

Contributions

We propose a novel technique to display text in an augmented reality environment that associates the text with the object to be augmented and maximizes reading legibility. The main contributions of our work are:

- we show how offset surfaces provide a natural surface for depicting text in augmented reality applications,
- we use ambient occlusion to efficiently identify the regions that afford high visibility from multiple directions, and
- we show how one can easily alter the orientation of the text on the surface according to the pose of the user so the text always appears upright and can be easily read.

2 Related Work

Here we briefly introduce existing work in augmented reality and its applications. We mainly focus on labeling and annotation problems.

2.1 Augmented Reality Applications

Augmented reality has been characterized as combining real and virtual environments, interactive in real time, and registered in three dimensions by Azuma [3]. Displaying additional information about the scene in textual form to strengthen the experience is one of the most common applications of augmented reality. For more general information regarding augmented reality like enabling technologies and problems it faces, we refer readers to surveys by Azuma and Krevelen *et al.* [3, 49].

Augmented reality has been successfully applied to medical applications. In Kancherla *et al.* [24], authors use an optical see-through headset to overlay computer generated graphics on the real scene to teach elbow anatomy. To help surgeons focus on the patient without having to look away to view medical imaging like ultrasound, Bajura *et al.* [4] display slices of ultrasound images in the physician's field of view. The image slices are positioned and oriented according to the ultrasound probe so they overlay on top of the organs where the images are taken. This work has been extended to showing the internal 3D volume of the human body in Ohbuchi[36]. In [6], Betting *et al.* use two side-by-side cameras to reconstruct the 3D geometry of the scene, which is used to track the headset and display MRI/CT images. During a surgery, augmented reality can display critical information like where to make incisions in the body , similar to Livingston *et al.*[29].

From the industrial point of view, augmented reality can be used to reduce manufacturing and maintenance cost, improve efficiency by eliminating templates, form-board diagrams and other masking devices like Caudell and Mizell[10]. In a more recent work by Henderson and Feiner [22], components of an aircraft engine are labelled in augmented reality, and instructions to assemble them are also displayed. A user study shows that trainees complete the assembly task faster with AR guidance than those who read instructions displayed on a separate screen.

2.2 Augmented reality Labelling and Annotation

A number of augmented reality systems, including ours, involve annotating, labelling, or displaying additional information about the real scene. In [17], Maurice introduced a handheld computer that displays information that is specific to certain objects in the real world. This work envisions displaying the image taken by the computer and labelling the objects in the picture, and displaying their functions. Rekimoto *et al.* [41] realises this vision with NaviCam, which is a handheld camera that displays additional information about the objects in the images that it is acquiring. Feiner *et al.*'s touring machine [16] tracks the location and orientation of users and display the names of the buildings in their field of view.

Rose *et al.* [44] and Azuma *et al.* [2] tackle more complicated scenes. The former labels different parts of the engine with 2D text displayed around the border of the picture and the engine in the center. Association between the parts and texts is illustrated with arrows pointing from parts to texts. As the camera rotates around the engine, the positions of the text are adjusted to avoid overlapping arrows. In [2], names and functions of different buttons of a panel are displayed as text. The positions of the text are automatically generated to be relatively close to the buttons. Special care is taken to ensure the text is evenly distributed in the screen space to avoid overlapping, while still maintaining proximity. When information about the scene is absent, Grasset *et al.* [20] analyze the image of the scene using visual saliency and uses those salient regions as anchor points to place text.

Orlosky *et al.* discuss placing user-centric (like text messages or emails) in dynamic scene while wearing a see-through AR display. Locations to place text are calculated so that the text is readable and does not occlude important information from the scene in Orlosky *et al.*[37, 38]. Iwai *et al.* [23] study text placement in projected AR, so the projected text does not appear distorted because of underlying geometry when viewed from arbitrary view points.

In Madsen *et al.* [32], a user study is conducted to find out what effect label placement could have on temporal coherence. Results show that study participants perform better when the labels are placed in the object space, and when the view management adjustment rate is limited.

2.3 3D labelling and annotation

3D labelling and annotation can be seen as an extension of 2D cartography, where point, line and area features are annotated. The task of generating a label layout which places labels in available space and minimizes the label overlap is proven to be NP-hard[33]. We refer readers to [11] by Christensen *et al.* for a more detailed survey on static 2D label placement. Greedy strategies and other approximations have been proposed.

Preim *et al.* [40] pioneered research that extended label placement to 3D objects. Label placement in 3D has been transformed to a 2D problem by labelling the shadow of the 3D object projected onto a plane in Ritter *et al.* [42]. Some papers compute the locations of the labels sequentially. Bell *et al.* [5] use a greedy heuristic to place text in rectangles near the anchor points that do not overlap existing text in the order from most important to least important. Another order is from those close to the observer to far away in Maass and Döllner[30]. Other papers [1, 47] by Ali *et al. and Stein and Décoret* treat the labels as a whole. Hartman *et al.* [21] use a dynamic potential field and assign attractive or repulsive forces to pairs of elements (the model, labels, and edges) in the scene. The locations of labels are computed by simulating the movement caused by the forces. Götzelmann *et al.* study labelling an animated 3D model[19]. Mühler and Preim [34] annotate 3D medical images for surgery planning by selecting visible anchor points and linking it to text with a connection line or using bent arrorws to point to occluded structures. Götzelmann *et al.* [18] pay special attention to asethetic values by using both internal and external labels while maintaining the balance among unambiguity, readability and frame coherence.

Bürger *et al.* [9] show a way to directly edit a volume scalar field. Annotations can be attached to volumes in the form of textures with resolution independent from that of the volume's.

Li *et al.* in [28, 27] alter the model to be labelled by creating a cutaway or exploded diagram to better show and label the occluded parts of the model.

Very few papers discuss text placement techniques that attach text to the objects. Ropinski *et al.* [43] wrap labels on the surface of 3D objects. They avoid bumpy surfaces. Cipriano *et al.* [13] place static text labels on an offset surface that smoothes the original surface. This technique has been extended to massive CAD models in Prado and Raposo[14]. Maass and Döllner [31] put text labels on parameterized hulls that generalize objects' geometry. Areas on the hulls are tested and selected so the labels placed on them are visible and legible

3 Smooth Offset Surfaces

To improve the association between the target object and the corresponding text information, we wrap the text information (in the form of a text texture) onto the surface of the target. Our technique, which is a form of texture decal discussed later in section 4. Whereas texture decal works well for objects with simple geometry, it presents significant limitations as the object complexity rises. Complex objects may have highly-irregular surfaces resulting in unreadable text. These complex objects are common in real-world applications, for example, folded protein molecules, biological organs such as small intestines, or mechanical CAD assemblies such as engines.

To account for bumpy or discontinuous surfaces, we need to smooth the mesh of objects and simplify the concavities as well as topological features such as holes and tunnels. Here was have adapted the text-scaffold technique of Cipriano *et al.* [13].

The text scaffold technique first voxelizes the original mesh and then performs in sequence dilation, smoothing the volume, and erosion. The newly created mesh, which represents the smooth surface calculated from the original volume, is used as the offset surface. We next discuss these steps in our implementation. Each step of the offset surface creation process is illustrated in fig 2.

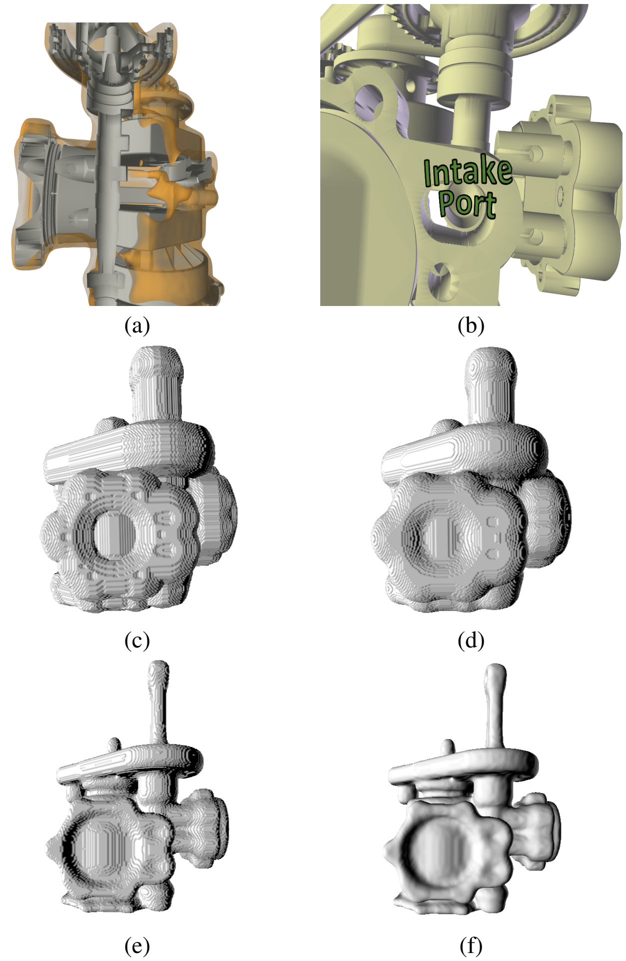


Fig. 2. Here we show the smooth offset surface and the process to create it. The offset surface is shown in (a). The internal grey mesh represents the original surface, while the orange exterior is the offset surface. (b) shows that the offset surface fills the gap on the original object surface. (c) is the voxelized volume after dilation. (d) shows the dilated volume convolved with a gaussian filter. (e) shows the eroded volume. 100 iterations of Taubin smoothing are applied to the extracted surface, which is shown in (f).

3.1 Volume Generation

The input mesh is scaled and translated to fit into a $X \times Y \times Z$ cube. A voxel is assigned value 1 if the interior or surface of the mesh is present in the voxel, and assigned 0 otherwise.

We use the implementation by Min *et al.* [1], which is a variation of the Nooruddin and Turk [35] technique. The volume is represented as a binary $volume(x,y,z)$. Suppose the mesh represent a solid object S,

$$volume(x,y,z) = \begin{cases} 1, & \text{if } (x,y,z) \in S. \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

3.2 Distance Transform

The steps from dilation to erosion are carried out on the distance field calculated on the volume. Here, we use the Chamfer distance transform [45, 7, 8]. The distance field is defined on the volume by specifying the value of each voxel ($df(x,y,z)$) to be its signed Chamfer distance to the boundary of the volume ∂S.

$$df(x,y,z) = \begin{cases} -\inf_{p \in \partial S} ||p-(x,y,z)|| & (x,y,z) \in \partial S \\ \inf_{p \in \partial S} ||p-(x,y,z)|| & \text{otherwise} \end{cases} \quad (2)$$

[1]http://www.cs.princeton.edu/ min/binvox/

We accomplish this in two stages. The initialization stage involves having the voxels inside the solid object getting assigned a value 0 and those outside assigned a value of infinity:

$$df(x,y,z) = \begin{cases} 0 & (x,y,z) \in \partial S \\ \infty & \text{otherwise} \end{cases} \quad (3)$$

The second stage consists of two passes. The forward pass propagates the distance field values by iterating through the voxels from one corner to the other across the diagonal, as:

$$df(x,y,z) = \min_{(i,j,k)\in d_{mat}} (D(x+i,y+j,z+k) + d_{mat}(i,j,k)), \quad (4)$$

where d_{mat} is a $3 \times 3 \times 3$ matrix.

The backward pass propagates in the opposite direction to the forward pass. In the two passes, different entries of d_{mat} are used [45].

3.3 Dilation

Dilation of a magnitude l transforms a solid object S into its Minkowski sum with a cube of side l as:

$$\bigcup_{s\in S} d_l(s), \quad (5)$$

where $d_l(s)$ is the cube of length l centered at s.

With objects defined in a distance field, dilation can be easily achieved by subtracting the dilation magnitude l from each voxel:

$$df_d(x,y,z) = df(x,y,z) - l.$$

This subtraction can be seen as assigning voxels whose distance from ∂S is l, which are outside of the S before dilation, to be the new boundary of S.

We dilate the occluded voxels even more to further smooth the boundary and fill any concavities, holes, and tunnels. The ambient dilation of a voxel $amb(x,y,z)$ is defined as

$$amb(x,y,z) = \max(0, v_{\text{thresh}} - vis_{\text{cur}}(x,y,z)), \quad (6)$$

where v_{thresh} is a predefined threshold and $vis_{\text{cur}}(x,y,z)$ is the portion of the direction from which the voxel (x,y,z) is visible. Ambient dilation is further subtracted from the distance field value of each voxel

$$df_{amb}(x,y,z) = df_d(x,y,z) - amb(x,y,z).$$

In the following, we denote the dilated object with S_d.

3.4 Erosion

Before we erode to recover the object, we smooth the distance field using a Gaussian filter.

To recover the object, we want to erode S_d as much as possible without intersecting the original object S. The erosion value is determined as the maximum distance field value (negative) on the original object boundary

$$\max_{(x,y,z)\in\partial S} df_{amb}(x,y,z).$$

This value is then subtracted from all distance field values.

3.5 Smooth offset surface

After we extract the mesh as the boundary in the distance field, the vertices will appear to be on a lattice because the mesh is extracted from a volume. We use Taubin smoothing [48] to adjust the locations of the vertices.

Each iteration of the Taubin smoothing has two passes. In each pass, every vertex is moved towards, to some degree, the average of its neighbours.

$$p' = p + scale * \frac{1}{|N_p|} \sum_{q\in N_p} q$$

, where N_p are the vertices that share an edge with p. The two passes have different scales, one positive and one negative.

4 Ambient Visibility

When users are performing tasks in augmented reality, time-critical information needs to be visible at all times. An important characteristic of augmented reality systems is the flexibility that allows users to change the viewing position and orientation freely. Nevertheless, such flexibility may cause augmented objects or texts to be occluded by other objects that also appear in a user's field of view. In this work we wrap texts onto an object, assuming that the information about that object is only important when that object itself is visible. This selective delivery of information reduces the amount of text that is visible from a specific view point and can help reduce the visual clutter.

While reducing the visual clutter by culling the unnecessary text display, we also need to ensure that when the relevant objects are in a user's view, their associated text content is maximally visible. There are two aspects to this problem. The first is to provide sufficient contrast with respect to the background [26]. The situation with optical see-through augmented reality is further complicated by real-world lighting [37, 38]. A complete discussion of contrast is beyond the scope of this paper. Here we focus on the second aspect, which is maximizing the number of view points from which the text can be seen.

We calculate the visibility of each region of the object's surface, and place the text on the region that maximizes visibility. Here, we define visibility by measuring the surface exposure to the rays from a certain set of view directions. In a surgery scenario, the set of desirable viewing directions could be the directions from a typical user's height around the operating table. We approximate this using ambient occlusion to calculate overall visibility for each candidate point.

4.1 Ambient occlusion

Ambient occlusion [50] is used in computer graphics to calculate how much a point is obscured from ambient lighting coming from all directions. The more exposed a point is, the brighter it will be. This technique has been widely used in the entertainment industry – both games and movies to create more realistic visual effects [25, 12, 39].

The basic idea of ambient occlusion is to shoot rays from a point p in a predefined range of directions Ω, and compute the ambient visibility of p as the integral of a visibility function over Ω.

$$A(p) = \frac{1}{C} \int_{\Omega} V(p,\omega) * (\vec{n} \cdot \omega) d\omega \quad (7)$$

The visibility function is defined to have value zero if p is occluded in the direction of ω, and one otherwise.

$$V(p,\omega) = \begin{cases} 1, & p \text{ occluded in } \omega \\ 0, & otherwise \end{cases} \quad (8)$$

C is a scalar constant for normalization.

4.2 Ambient visibility

We calculate ambient visibility value $A(p)$ for every vertex p on the offset surface. The visibility of each part P is the average of ambient visibility of all vertices in P.

$$A(P) = \frac{1}{|P|} \sum_{p\in P} A(p) \quad (9)$$

The ambient visibility values, the determined visible areas after thresholding, the selected regions to place text labels, and the model rendered with labels are all shown in Fig 3

5 Text Rotation to Enhance Readability

If the text is statically texture-mapped on the surface, much like a sticker stuck on an object, it maintains view-independence. This unfortunately also means that the text may appear upside down and/or in reverse order when viewing from a different view point. Although users may still be able to read the text, the cognitive burden introduced can be significant. Here we aim to reduce the cognitive load of

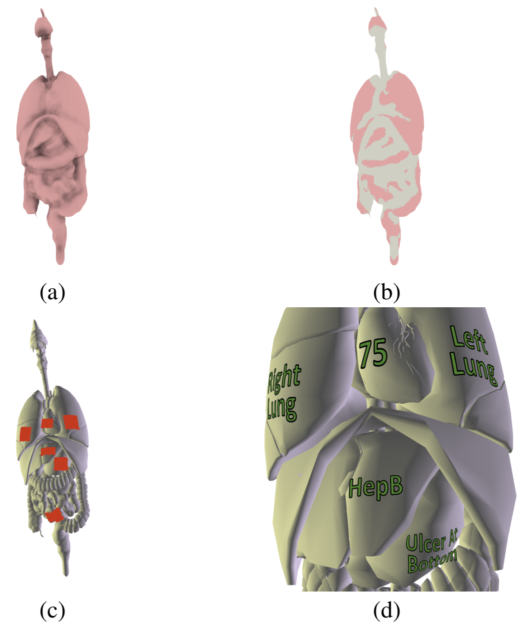


Fig. 3. Choosing surface locations with high visibility to place text. Ambient visibility is calculated as shown in (a). The brighter a point is, the more visible it is. (b) shows the visibility after thresholding. Red shows that the average visibility in the area is above the threshold, while blue indicates that it is below. Regions are selected based on the thresholded visibility to place text, which are shown as red patches. (d) shows displays text overlaid on top of the detailed model.

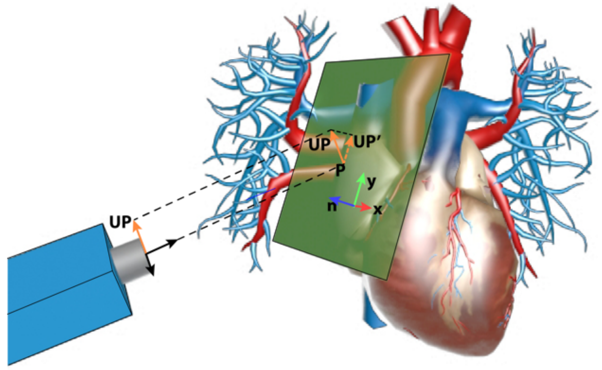


Fig. 4. Orienting the texture according to view. The camera's up vector $\overrightarrow{up}$ is projected to the tangent plane at p to $\overrightarrow{up'}$. The 2D coordinate system of the tangent plane is rotated so that the positive y axis aligns with $\overrightarrow{up'}$. Texture is placed on the tangent plane with the upward direction being positive y.

the users by orienting the text according to the view point so the text always appears upright.

In this section, we first briefly describe the technique we have used to wrap the texture on the surface. We next explain how we orient the text based on the view point of users.

5.1 Displaying Text as Texture decal

We treat the text labels as textures and use the decal technique to wrap them onto the surfaces. Decal has been widely used in computer graphics to supplement the underlying textures with repeating high-frequency details such as holes and grass. Using texture decal greatly reduces the work load of texture mapping. In our work, we implement the discrete exponential map approximation technique [46].

Given a text texture and a patch F on the surface ∂S, the objective is to find for each vertex in F the texture coordinate. The exponential map [15], with a specified point p on F, exp_p, maps points on F to the 2D plane tangent on p. The texture can be placed on the tangent plane, such that each vertex on F can have a texture coordinate.

Let $exp_p(q)$ be the coordinate of point q mapped to the tangent plane at p. Let T_p be the tangent plane at p.

$$exp_p(p) = \vec{0} \tag{10}$$

The 3D coordinate system of the tangent plane at p is characterized by three vectors, the normal, and a pair of tangent-plane basis vectors $(\vec{n}_p, \vec{x}_p, \vec{y}_p)$. Suppose we wish to compute $exp_p(q)$,

$$exp_p(q) = exp_p(r) + (exp_p(q) - exp_p(r)) \tag{11}$$

where $exp_p(r)$ is the known coordinate of r in the tangent plane of p.

We approximate $exp_p(q) - exp_p(r)$ by transforming $exp_r(q) - exp_r(r)$, which is the vector from r to the projection of q onto T_r. $exp_r(q)$ is know to us. To approximate, we rotate T_r so it is co-planar with T_p. Specifically, first rotate T_r to align the normal vectors n_r and n_p. Then use a 2D rotation to align the x and y axes.

$$n_p = Rot_{(n_r \times n_p)} * n_r$$
$$x'_r = Rot_{(n_r \times n_p)} * x_r$$
$$y'_r = Rot_{(n_r \times n_p)} * y_r$$
$$x_p = Rot_{n_p} x'_r$$
$$y_p = Rot_{n_p} y'_r$$

$Rot_{n_r \times n_p}$ and Rot_{n_p} are the two rotation matrices about $n_r \times n_p$ and n_p respectively that satisfie the above equations. $n_r \times n_p$ is the cross product of n_r and n_p. Thus, $exp_p(q) - exp_p(r)$ is approximated as

$$exp_p(q) - exp_p(r) = Rot_{n_p} * exp_r(q) \tag{12}$$

5.2 Rotating the Texture to Improve Readability

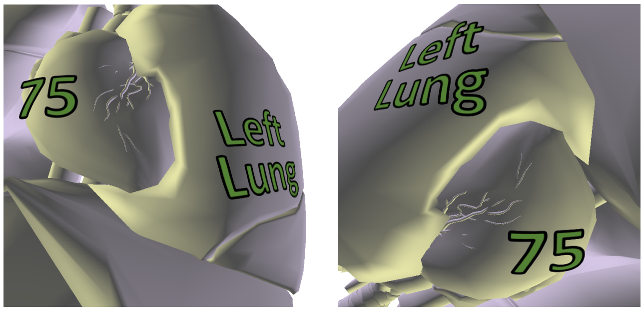
Most languages have an upright direction in which the characters or letters are written or read. Although humans are flexible in reading the text from different directions, doing so usually requires extra cognitive efforts. In our augmented reality system, we reduce such overhead by re-orienting the text. To facilitate reading the text, the up direction of the text $\overrightarrow{up_t}$, which is a 3D vector in space once wrapped onto a surface, should be parallel to the plane spanned by the up $\overrightarrow{up_v}$ and forward $\overrightarrow{f_v}$ direction of the user's view. And the angle between the $\overrightarrow{up_t}$ and $\overrightarrow{up_v}$ should be smaller than $\frac{\pi}{2}$.

To achieve this, we project $\overrightarrow{up_v}$ down to the tangent plane T_p and use it as the positive y direction y_p of the tangent plane T_p. After calculating the exponential map exp_p, the texture is mapped to the T_p such that $\vec{u}$ and $\vec{v}$, the axes of the texture, are parallel to $\vec{x_p}$ and $\vec{y_p}$, respectively (see Fig 4).

To maintain readability, decal must be updated every time the view matrix is changed, which is practically every frame rendered in a head-mounted augmented reality system. Our approach outlined above is quite efficient and can be easily accommodated in the augmented reality processing pipeline.

6 Augmented Reality Application: Medical Inspection Mock-up

To validate the effects of our technique in augmented reality, we have built and tested this in a mock-up augmented reality medical inspection system. The user wears an optical see-through head-mounted display and moves around a mannequin. The relative position and pose of the head-set is tracked. A detailed model of the organs of the human



(a) (b)

Fig. 5. Text orientation changed according to view

body is displayed and aligned with the mannequin in the user's field of view.

To show the benefits of our proposed technique, we compare it with an approach where text appears as billboards floating in the screen space and associated with the objects via arrows. We compare the visibility of text placed on calculated visible surfaces against the visibility of text placed by hand. Finally, we show the necessity of rotating the text depending on user's view by comparing it against static text placement.

6.1 Hardware setup

We use a Vuzix Star 1200 Augmented Reality headset. It is equipped with a color camera, whose position and pose is estimated by tracking markers placed on prespecified locations in the scene. The display on the headset is calibrated to have the text augmentation aligned with the real object in user's field of view. The augmented reality images shown in this paper have been captured from inside the Vuzix headset.

To allow certain level of portability, processing is done on a laptop computer with an Intel i7 CPU at 2.8GHz, 16GB RAM, and NVIDIA Quadro K2100M.

6.2 Comparison against billboard technique

In this work, we propose to wrap text around the surface of objects. To show how our technique is better than screen space text display, we have implemented a simple version of screen space text display where the labels are fixed in screen space.

As can be seen in Figure 1, in our proposed application of medical procedures or industrial maintenance, the patient or equipment to be operated on is quite close to the user (at arm's length). As a result, the object would take up much of the display real-estate of the headset. As far as we know, none of the available optical see-through headsets supports a large field of view. Because of the limits in display real estate, internal labels are used, which if not carefully placed would occlude parts of the object. In augmented reality, the user constantly changes the viewing position and direction. Change of the view would cause the arrows to become crossed. While there are view management techniques, for example in [2, 1, 11], label placement remains a hard problem.

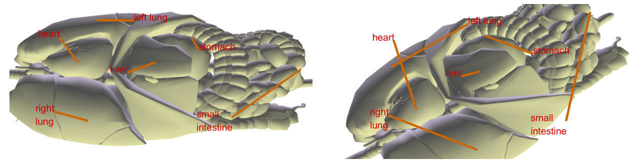
In our implementation of the billboard display, the links between labels and anchor points become crossed after view changes (Fig 6).

6.3 Ambient visibility

To find out whether ambient-visibility-assisted label placement helps improve visibility of labels, we present a comparison between a set of labels placed at areas with high ambient visibility against labels placed at other arbitrarily selected areas.

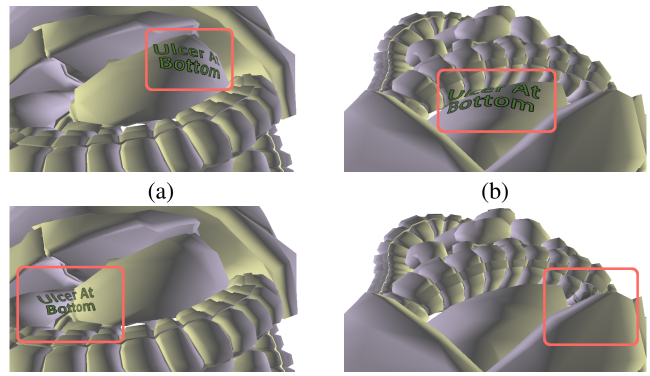
Figure 7 shows a case where the text "ulcer at bottom" indicating the condition of the stomach has larger general visibility placed on the area with high ambient visibility than placed on another area.

For a more systematic evaluation, we sample a selection of view points above and around the object model. The views mimic those of physicians' moving around the operating table looking down at the



(a) (b)

Fig. 6. Visualization of billboard text display in AR headset. In figure (a), we place the labels so the links from labels to anchor points are not crossed. When the view changes, the arrows become crossed, as in (b).



(a) (b)

(c) (d)

Fig. 7. Label placement at areas with large ambient visibility vs other areas. (a) and (b) shows two views of labels placed at areas with large ambient visibility. The text "ulcer at bottom" on the stomach is visible from both of the views. (c) and (d) shows two views of labels placed at arbitrarily selected areas. The same text is visible only from one view.

patient. We place six labels on six organs, and use two sets of label placements. In the first set, the 6 labels are placed on the high ambient visibility areas of the six corresponding organs. In the second set, the same 6 labels are placed on other, randomly-selected, areas of the same 6 organs. We manually counted the total number of visible labels across all the sampled views. In this preliminary study, our finding is that 518 out 600 labels are visible with ambient-visibility-assisted label placement, while 358 out of 600 labels are visible with the arbitrary set of label placements.

6.4 Text Orientation

We next compare the visualization of view-dependent text placement against static text placement. We start out by disabling view-dependent text placement. As we move around the mannequin, the text becomes inverted and unreadable. Then we enable the view dependent text placement. The visuals are shown in fig 8. The texts without view dependent placement are harder to read.

With view-dependent text placement, calculating the decal texture coordinates becomes the most time-consuming part. However, we can still achieve an interactive frame rate of 31 fps with a CPU implementation.

7 Conclusions and Future Work

In this paper, we have explored techniques to display text in augmented reality. Our goals are achieving good object-label association, visibility, and readability. To achieve these goals, we have designed and developed the technique of using offset surfaces from the original object and place labels on areas of the surface visible from a large set of potential view directions. Our approach orients the text in accordance with the view so the texts always appears up right and easily readable to the user.

Initial experiments show that our approach has a good object-label association as the label is always close to and follows the same geom-

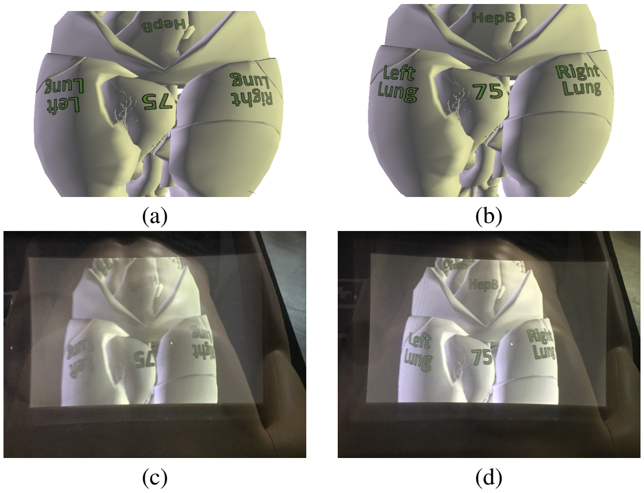


Fig. 8. Visualizing with static and view-dependent text rotation in AR headset. (a) and (c) show the text labels with view dependent label placement disabled, which is enabled in (b) and (d). Light is dimmed for photographing purposes.

etry as the object. Ambient visibility assisted label placement gives a 45% improvement in visibility over arbitrary placements. View-dependent text orientation reduces the mental cognitive load in reading text compared to static text placements.

To better show the advantage our technique has on text display in augmented reality, we need a more systematic evaluation with more user input. We plan to conduct a user study that shows that our text display has more value than existing text display techniques, such as billboards.

We have shown that ambient visibility assisted label placement allows labels to be seen from a large range of viewing directions. We can still encounter cases where the area with the maximum ambient visibility is occluded when viewed from some view directions. For such cases, we plan to explore how to dynamically move the text labels to a new area of the same object that is visible from the current view. We need to be careful with such view adjustment to avoid temporal flickering.

A major problem facing mass deployment of AR is that most AR systems work only in controlled environments where the object and users are tracked with a sufficiently high degree of accuracy. It is impractical to place markers, visible or infra-red, on all the equipment that needs inspection or maintenance. However, the wide use of medical imaging in modern medicine and CAD in modern manufacturing may provide a solution. We aim to implement our text display techniques on an augmented reality platform that can function in more general and uncontrolled environments.

We believe that our technique shows an impressive potential to be used across a wide variety of augmented medical procedures, manufacturing, maintenance, and many other fields.

References

[1] K. Ali, K. Hartmann, and T. Strothotte. Label layout for interactive 3d illustrations. 2005.

[2] R. Azuma and C. Furmanski. Evaluating label placement for augmented reality view management. In *Proceedings of the 2nd IEEE/ACM international Symposium on Mixed and Augmented Reality*, page 66. IEEE Computer Society, 2003.

[3] R. T. Azuma. A survey of augmented reality. *Presence: Teleoperators and virtual environments*, 6(4):355–385, 1997.

[4] M. Bajura, H. Fuchs, and R. Ohbuchi. Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. In *ACM SIGGRAPH Computer Graphics*, volume 26, pages 203–210. ACM, 1992.

[5] B. Bell, S. Feiner, and T. Höllerer. View management for virtual and augmented reality. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 101–110. ACM, 2001.

[6] F. Betting, J. Feldmar, N. Ayache, and F. Devernay. A new framework for fusing stereo images with volumetric medical images. In *Computer Vision, Virtual Reality and Robotics in Medicine*, pages 30–39. Springer, 1995.

[7] G. Borgefors. Distance transformations in digital images. *Computer vision, graphics, and image processing*, 34(3):344–371, 1986.

[8] G. Borgefors. On digital distance transforms in three dimensions. *Computer vision and image understanding*, 64(3):368–376, 1996.

[9] K. Burger, J. Kruger, and R. Westermann. Direct volume editing. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1388–1395, 2008.

[10] T. P. Caudell and D. W. Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on*, volume 2, pages 659–669. IEEE, 1992.

[11] J. Christensen, J. Marks, and S. Shieber. An empirical study of algorithms for point-feature label placement. *ACM Transactions on Graphics (TOG)*, 14(3):203–232, 1995.

[12] P. Christensen. Ambient occlusion, image-based illumination and global illumination. photorealistic renderman application notes. Technical report, Note 35, 2002.

[13] G. Cipriano and M. Gleicher. Text scaffolds for effective surface labeling. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1675–1682, 2008.

[14] R. Deris Prado and A. Barbosa Raposo. Real-time label visualization in massive cad models. In *Computer-Aided Design and Computer Graphics (CAD/Graphics), 2013 International Conference on*, pages 337–344. IEEE, 2013.

[15] M. P. Do Carmo. *Differential geometry of curves and surfaces*, volume 2. Prentice-hall Englewood Cliffs, 1976.

[16] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster. A touring machine: Prototyping 3d mobile augmented reality systems for exploring the urban environment. *Personal Technologies*, 1(4):208–217, 1997.

[17] G. W. Fitzmaurice. Situated information spaces and spatially aware palmtop computers. *Communications of the ACM*, 36(7):39–49, 1993.

[18] T. Götzelmann, K. Ali, K. Hartmann, and T. Strothotte. Form follows function: Aesthetic interactive labels. *Computational aesthetics*, 5, 2005.

[19] T. Götzelmann, K. Hartmann, and T. Strothotte. Annotation of animated 3d objects. In *SimVis*, volume 7, pages 209–222. Citeseer, 2007.

[20] R. Grasset, T. Langlotz, D. Kalkofen, M. Tatzgern, and D. Schmalstieg. Image-driven view management for augmented reality browsers. In *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*, pages 177–186. IEEE, 2012.

[21] K. Hartmann, K. Ali, and T. Strothotte. Floating labels: Applying dynamic potential fields for label layout. In *Smart Graphics*, pages 101–113. Springer, 2004.

[22] S. J. Henderson and S. K. Feiner. Augmented reality in the psychomotor phase of a procedural task. In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, pages 191–200. IEEE, 2011.

[23] D. Iwai, T. Yabiki, and K. Sato. View management of projected labels on nonplanar and textured surfaces. *Visualization and Computer Graphics, IEEE Transactions on*, 19(8):1415–1424, 2013.

[24] A. R. Kancherla, J. P. Rolland, D. L. Wright, and G. Burdea. A novel virtual reality tool for teaching dynamic 3d anatomy. In *Computer Vision, Virtual Reality and Robotics in Medicine*, pages 163–169. Springer, 1995.

[25] H. Landis. Renderman in production. In *ACM SIGGRAPH*, 2002.

[26] A. Leykin and M. Tuceryan. Automatic determination of text readability over textured backgrounds for augmented reality systems. In *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 224–230. IEEE Computer Society, 2004.

[27] W. Li, M. Agrawala, B. Curless, and D. Salesin. Automated generation of interactive 3d exploded view diagrams. In *ACM Transactions on Graphics (TOG)*, volume 27, page 101. ACM, 2008.

[28] W. Li, L. Ritter, M. Agrawala, B. Curless, and D. Salesin. Interactive cutaway illustrations of complex 3d models. In *ACM Transactions on Graphics (TOG)*, volume 26, page 31. ACM, 2007.

[29] M. A. Livingston, W. F. Garrett, G. Hirota, M. C. Whitton, E. D. Pisano, H. Fuchs, et al. Technologies for augmented reality systems: realizing ultrasound-guided needle biopsies. In *Proceedings of the 23rd annual conference on computer graphics and interactive techniques*, pages 439–446. ACM, 1996.

[30] S. Maass and J. Döllner. Efficient view management for dynamic anno-

tation placement in virtual landscapes. In *Smart Graphics*, pages 1–12. Springer, 2006.

[31] S. Maass and J. Döllner. Seamless integration of labels into interactive virtual 3d environments using parameterized hulls. In *Proceedings of the Fourth Eurographics conference on Computational Aesthetics in Graphics, Visualization and Imaging*, pages 33–40. Eurographics Association, 2008.

[32] J. B. Madsen, M. Tatzgern, C. B. Madsen, D. Schmalstieg, and D. Kalkofen. Temporal coherence strategies for augmented reality labeling. 2016.

[33] J. Marks and S. M. Shieber. The computational complexity of cartographic label placement. 1991.

[34] K. Mühler and B. Preim. *Automatic textual annotation for surgical planning*. Citeseer, 2009.

[35] F. S. Nooruddin and G. Turk. Simplification and repair of polygonal models using volumetric techniques. *Visualization and Computer Graphics, IEEE Transactions on*, 9(2):191–205, 2003.

[36] R. Ohbuchi, M. Bajura, and H. Fuchs. Case study: Observing a volume rendered fetus within a pregnant patient. In *Visualization: Proceedings of the IEEE Conference on Visualization*, volume 5. Citeseer, 1998.

[37] J. Orlosky, K. Kiyokawa, and H. Takemura. Dynamic text management for see-through wearable and heads-up display systems. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, pages 363–370. ACM, 2013.

[38] J. Orlosky, K. Kiyokawa, and H. Takemura. Managing mobile text in head mounted displays: studies on visual preference and text placement. *ACM SIGMOBILE Mobile Computing and Communications Review*, 18(2):20–31, 2014.

[39] J. Pantaleoni, L. Fascione, M. Hill, and T. Aila. Pantaray: fast ray-traced occlusion caching of massive scenes. In *ACM Transactions on Graphics (TOG)*, volume 29, page 37. ACM, 2010.

[40] B. Preim, A. Raab, and T. Strothotte. Coherent zooming of illustrations with 3d-graphics and text. In *Graphics Interface*, volume 97, pages 105–113, 1997.

[41] J. Rekimoto. The magnifying glass approach to augmented reality systems. In *International Conference on Artificial Reality and Tele-Existence*, volume 95, pages 123–132, 1995.

[42] F. Ritter, H. Sonnet, K. Hartmann, and T. Strothotte. Illustrative shadows: integrating 3d and 2d information displays. In *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 166–173. ACM, 2003.

[43] T. Ropinski, J.-S. Praßni, J. Roters, and K. H. Hinrichs. Internal labels as shape cues for medical illustration. In *VMV*, volume 7, pages 203–212, 2007.

[44] E. Rose, D. Breen, K. H. Ahlers, C. Crampton, M. Tuceryan, R. Whitaker, and D. Greer. Annotating real-world objects using augmented reality. In *Computer Graphics: Developments in Virtual Environments (Proceedings of CG International95 Conference)*, pages 357–370, 1995.

[45] R. Satherley and M. W. Jones. Vector-city vector distance transform. *Computer Vision and Image Understanding*, 82(3):238–254, 2001.

[46] R. Schmidt, C. Grimm, and B. Wyvill. Interactive decal compositing with discrete exponential maps. *ACM Transactions on Graphics (TOG)*, 25(3):605–613, 2006.

[47] T. Stein and X. Décoret. Dynamic label placement for improved interactive exploration. In *Proceedings of the 6th international symposium on Non-photorealistic animation and rendering*, pages 15–21. ACM, 2008.

[48] G. Taubin. A signal processing approach to fair surface design. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 351–358. ACM, 1995.

[49] D. Van Krevelen and R. Poelman. A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, 9(2):1, 2010.

[50] S. Zhukov, A. Iones, and G. Kronin. An ambient light illumination model. In *Rendering Techniques 98*, pages 45–55. Springer, 1998.