LaTeX Author Guidelines for 3DIMPVT Proceedings

Anonymous 3DIMPVT submission

Paper ID ****

Abstract

Automated 3D modeling of building interiors is used in applications such as virtual reality and environment mapping. Texture mapping these models accurately is vital towards visualizing the data gathered by modeling systems and increases overall usability. The localization of cameras in 3D scenes often suffers from inaccuracies, resulting in visible discontinuties when different images are projected onto a texture. Previous approaches at minimizing these discontinuities suffer from error accumulation when stitching together multiple images and do not robustly handle wide ranges of camera locations and angles. We propose two approaches for reducing discontinuities during texture mapping, tailored towards a backpack-mounted data acquisition system.

1. Introduction

Three-dimensional modeling of indoor environments has a variety of applications such as training and simulation for disaster management, virtual heritage conservation, and mapping of hazardous sites. Manual construction of these digital models can be time consuming, and as such, automated 3D site modeling has garnered much interest in recent years.

The first step in automated 3d modeling is the physical scanning of the environment's geometry. An indoor modeling system must be able to calculate camera locations within an environment while simulatenously reconstructing the 3D structure of the environment itself. This problem is studied by the robotics and computer vision communities as the simultaneous localization and mapping (SLAM) problem, and is generally solved using a combination of laser range scanners, cameras, and inertial measurement units (IMUs).

The aim of this paper is to present a solution for texture mapping the 3D models generated by indoor modeling systems, with specific attention given to a human-operated system with higher localization errors and greater variance in camera locations. The paper is organized as follows. Section 2 provides an overview of the backpack modeling system from which data and examples used throughout this paper are from. Section 3 describes the general problem of 3D texture mapping and reviews existing approaches. Section 4 describes challenges posed by the backpack modeling system and presents our two approaches tailored for walls and floors/ceilings respectively. Section 5 compares results and presents conclusions.

2. Backpack Modeling System

Human-operated data acquisition systems provide unique advantages in terms of agility and portability over vehicular-mounted systems. Unfortunately, human-operated systems suffer from a lack of automation and stability, resulting in higher localization variance and leading to the problems discussed in Section 4.

2.1. Data Acquisition Hardware

The backpack modeling system from which our data was collected contains five 2D laser range scanners, two cameras, an orientation sensor, and an IMU (Should I even mention the IMU at all?). The laser scanners are 40Hz Hokuyo UTM-30LX 2D laser scanners with a 30-meter range and a 270° field of view. These scanners are mounted orthogonally to one another. The two cameras are Point Grey Grasshopper GRAS-50S5C units equipped with fisheye lenses, resulting in a 180° field of view. The IMU, a Honeywell HG9900, is a strap-down navigation-grade sensor which combines three ring laser gyros with bias stability of less than 0.003°/hour and three precision accelerometers with bias of less than 0.245mm/sec^2. The HG9900 provides highly accurate measurements of all 6 DOF at 200Hz and thus serves as our ground truth for the experiments in this paper.]The orientation sensor (OS), an InterSense InertiaCube3, provides orientation parameters at a rate of 180Hz.

2.2. Environment Reconstruction

Using a combination of these sensors and multiple localization and loop-closure algorithms, the backpack is localized over its data collection period, and a 3D point cloud of our environment is constructed. Approximate normal vectors for each point in the point cloud are then calculated by gathering neighboring points within a small radius and running a Principal Component Analysis procedure. These normal vectors allow for the classification and grouping of adjacent points into structures such as walls, ceilings, floors, and staircases. A RANSAC algorithm is then employed to fit polygonal planes to these structured groupings of points, resulting in a fully planar model. This planar model, consisting of multiple 2D polygonal planes in 3D space, along with the set of images captured by the backpack's camera, can be considered the input data on which our approach runs.

2.3. Image Subsampling

Before beginning the texture mapping process, it is prudent to reduce the amount of potential images being considered for texture mapping each plane. The backpack modeling system takes 5 pictures/second from both cameras, resulting in each plane being present in (and capable of being textured mapped by) thousands of images, at a wide variance of distances and angles. Thus, for the sake of efficiency, it makes sense to associate each plane with a much smaller sample of images. When creating this subset of images for each plane, there are three criteria we keep in mind. First, we want such a subset of images to be capable of texturing the entirety of a plane, so as to ensure there are no holes in our final texture. Second, we want all our images to be taken from a relatively close distance to the plane, and at as much of a direct, head-on angle as possible. This ensures that the image projects squarely onto our desired plane, with minimal inclusion of other planes, and we get a texture with high resolution. Third, we desire an amount of overlap between the images chosen. This overlap helps us with preprocessing as well as postprocessing techniques discussed in Sections x and y.

To meet these three criteria, we tesselate each plane into [X] triangles of [Y] size. For each triangle, we find all images that can be projected upon the entirety of the triangle, accounting for occlusion with other planes as well, using standard ray-polygon intersection tests. Of these images, we then select the best one according to the heuristic in equation X, favoring closer distances and more head-on viewing angles. We also enforce that no two images are chosen such that they were taken within 0.5 meters of eachother. This results in a greatly-reduced set of images such that each one is objectively the "best" location-wise for some number of areas on the plane, addressing the first two criteria. Furthermore, since cameras are only at good angles to triangles near their center of projection, images will be chosen such that there is high overlap between them. This process reduces the amount of candidate images for each plane from many thousands to around 50, at fairly even spacing. With this completed, each of our planes is associated with a sufficiently sized list of candidate images, and

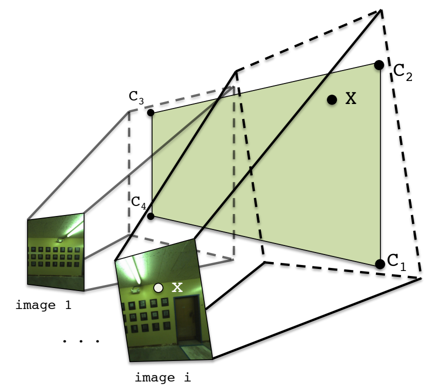

Figure 1. The plane is specified in 3D space by the four corners C_1 to C_4. Images are related to the plane through the camera matrices $P_{1..M}$.

we are ready to begin texture mapping.

3. Texture Mapping Overview

The geometry of the the texture mapping problem for indoor 3D modeling is shown in Figure 1. We are given a set of M images. Each image has a camera matrix P_i for $i = 1..M$, which translates a point in the world coordinate system to a point in image i's coordinates. A camera matrix P_i is composed of the camera's intrinsic parameters, such as focal length and image center, as well as the extrinsic parameters which specify the rotation and translation of the camera center's position with respect to the world coordinates at the time that image i was taken. These extrinsic parameters are determined by the localization hardware and algorithms as part of the indoor modeling system. A point X on the plane can be related to its corresponding pixel x in image i through the following equation:

$$x = project(P_i X)$$

where

$$X = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \text{ and } project(X) = \begin{pmatrix} x/z \\ y/z \end{pmatrix}$$

We are also given a plane generated by the surface reconstruction system, which is to be texture mapped by these images. The plane is defined by four corner points C_1 to C_4 in world coordinates and a normal vector indicating the front facing side of the plane. The challenge is to texture the plane using these images, while eliminating any discontinuities or seams that would suggest that the plane was not composed of a single continuous image.

3.1. Best-Candidate Mapping

Ignoring the fact that the camera matrices $P_{1..M}$ are inaccurate, one can texture map the plane simply by discretiz-

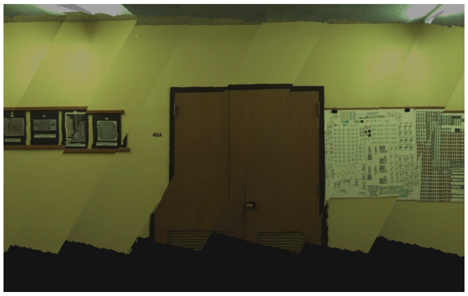

Figure 2. The result of naive texture mapping based on the imprecise camera matrices estimated by the localization system.

Figure 3. Image mosaicing.

ing the plane into tiles or triangles and texturing each tile or triangle seperately. It makes sense to texture each tile or triangle by selecting a best candidate image. This is usually the nearest and/or most direct facing image to the tile or triangle.

As Figure 2 demonstrates, the best candidate mapping leads to significant misalignment between successive tiles. This suggests that while the errors in the localization system are quite low, they are not pixel accurate. For photorealistic texture mapping, either the camera matrices need to be refined such that the localization is pixel accurate, or image stitching techniques need to be applied to provide this illusion. The following are two existing techniques to solve this problem.

3.2. Image Mosaicing

When images are taken of a plane from arbitrary overlapping positions, they are related by homography [**?**]. Thus, existing homography-based image mosaicing algorithms are applicable [**?**]. However, errors can compound when long chains of images are mosaiced together using these approaches. For example, a pixel in the nth image in the chain must be translated into the first image's coordinates by multiplying by the 3×3 matrix $H_1 H_2 H_3 ... H_n$. Any error in one of these homography matrices is propagated to all further images until the chain is broken. For some chains of images this can happen almost immediately due to erroneous correspondence matches and the resulting image mosaic is grossly misshapen.

Figure 3 shows the output of the AutoStitch software package which does homography-based image mosaicing. This plane is nearly a best-case scenerio with many features spread uniformly across it. Even so, the mosaicing produces errors that causes straight lines to appear as waves on the plane. This image was generated after careful hand tuning. Many planes that had fewer features simply failed. This led us to conclude that image mosaicing alone is not enough to reliably texture map our indoor 3D modeling dataset.

Figure 4. Using the graph-based localization refinement algorithm from [11] suffers from the problem of compounding errors.

3.3. Image-Based 3D Localization Refinement

Another approach is to refine the camera matrices using image correspondences to guide the process. Each camera matrix has 6 degrees of freedom. Previous work attempted to refine the camera matrix by solving a non-linear optimization problem [**?**]. This process is carried out at the same time as the laser based backpack localization and is therefore specific to the system in [**?**, **?**]. Unfortunately, this approach suffers from a similar error propagation problem shown in Figure 4. In this work we also refine the placement of images using image correspondences. However, we do so in two dimensions on the plane where as this previous work did so over all 6 degrees of freedom. Refining in two dimensions on the plane is less flexible in that it cannot fix certain projection errors. Emperically we find that it avoids the error propagation problem.

4. something

5. NOTES

4 Special Considerations due to BMS blabla sucky title The backpack modeling system described in Section 2 was designed to prioritize for texturing walls. As mentioned previously, the backpack contains two fisheye cameras, one aimed directly to the left of the backpack operator, and one aimed directly to the right. Thus, in order to texture the ceiling where the operator walked, images from both cameras must be combined. Even more problematic, the floor beneath the backpack is physically obscured by the operator's body, and thus must be textured by images taken from different locations. These problems are compounded by the fact that the operator's data acquisition path attempts to fully scan each wall, but not necessarily the entirety of a floor or ceiling. This means that large swaths of floors and ceilings are never walked directly over or under, and thus must be textured by sidelong images taken from further distances and at oblique angles. We strongly prefer head-on images because when images are projected

onto a planar surface at an extreme angle, the resultant texture ends up spanning a large portion of the plane, applying low-resolution textures to areas at a great distance from the camera location.

In short, these factors mean that wall planes are associated with a multitude of upright, uniform images, taken head-on and from a relatively close distance at regular intervals, while floors and ceilings are associated with images of all orientations and distances. This major difference suggests we employ two strategies, one tailored for each scenario.

4.1 Texturing Floors and Ceilings

4.2 Texturing Walls Since our candidate images for texturing a wall are all taken from similar distances and orientations relative to our wall, their individual ”qualities” are all fairly similar. We don’t have to worry as much about filtering out bad images; rather we seek to select some subset of our images which together form a good texture.

5 Results and Analysis but also occlusion projection thing. door projected onto wall is fine, door projected onto floor is bad.

6. EVERYTHING AFTER HERE IS TEMPLATE STUFF

7. Introduction

Please follow the steps outlined below when submitting your manuscript to the IEEE Computer Society Press. This style guide now has several important modifications (for example, you are no longer warned against the use of sticky tape to attach your artwork to the paper), so all authors should read this new version.

7.1. Language

All manuscripts must be in English.

7.2. Dual submission

By submitting a manuscript to 3DIMPVT, the authors assert that it has not been previously published in substantially similar form. Furthermore, no paper which contains significant overlap with the contributions of this paper either has been or will be submitted during the 3DIMPVT 2012 review period to **either a journal** or any conference or any workshop. **Papers violating this condition will be rejected**.

If there are papers that may appear to the reviewers to violate this condition, then it is your responsibility to: (1) cite these papers (preserving anonymity as described in Section 1.6 below), (2) argue in the body of your paper why your 3DIMPVT paper is non-trivially different from these concurrent submissions, and (3) include anonymized versions of those papers in the supplemental material.

7.3. Paper length

3DIMPVT papers may be between 6 pages and 8 pages. Overlength papers will simply not be reviewed. This includes papers where the margins and formatting are deemed to have been significantly altered from those laid down by this style guide. Note that this LaTeX guide already sets figure captions and references in a smaller font. The reason such papers will not be reviewed is that there is no provision for supervised revisions of manuscripts. The reviewing process cannot determine the suitability of the paper for presentation in eight pages if it is reviewed in eleven.

7.4. The ruler

The LaTeX style defines a printed ruler which should be present in the version submitted for review. The ruler is provided in order that reviewers may comment on particular lines in the paper without circumlocution. If you are preparing a document using a non-LaTeX document preparation system, please arrange for an equivalent ruler to appear on the final output pages. The presence or absence of the ruler should not change the appearance of any other content on the page. The camera ready copy should not contain a ruler. (LaTeX users may uncomment the `\cvprfinalcopy` command in the document preamble.) Reviewers: note that the ruler measurements do not align well with lines in the paper — this turns out to be very difficult to do well when the paper contains many figures and equations, and, when done, looks ugly. Just use fractional references (e.g. this line is 097.15), although in most cases one would expect that the approximate location will be adequate.

7.5. Mathematics

Please number all of your sections and displayed equations. It is important for readers to be able to refer to any particular equation. Just because you didn’t refer to it in the text doesn’t mean some future reader might not need to refer to it. It is cumbersome to have to use circumlocutions like “the equation second from the top of page 3 column 1”. (Note that the ruler will not be present in the final copy, so is not an alternative to equation numbers). All authors will benefit from reading Mermin’s description of how to write mathematics: http://www.cvpr.org/doc/mermin.pdf.

7.6. Blind review

Many authors misunderstand the concept of anonymizing for blind review. Blind review does not mean that one must remove citations to one’s own work—in fact it is often impossible to review a paper unless the previous citations are known and available.

Blind review means that you do not use the words “my”

3DIMPVT #****

or “our” when citing previous work. That is all. (But see below for techreports)

Saying “this builds on the work of Lucy Smith [1]” does not say that you are Lucy Smith, it says that you are building on her work. If you are Smith and Jones, do not say “as we show in [7]”, say “as Smith and Jones show in [7]” and at the end of the paper, include reference 7 as you would any other cited work.

An example of a bad paper just asking to be rejected:

> An analysis of the frobnicatable foo filter.
>
> In this paper we present a performance analysis of our previous paper [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.
>
> [1] Removed for blind review

An example of an acceptable paper:

> An analysis of the frobnicatable foo filter.
>
> In this paper we present a performance analysis of the paper of Smith *et al*. [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.
>
> [1] Smith, L and Jones, C. “The frobnicatable foo filter, a fundamental contribution to human knowledge”. Nature 381(12), 1-213.

If you are making a submission to another conference at the same time, which covers similar or overlapping material, you may need to refer to that submission in order to explain the differences, just as you would if you had previously published related work. In such cases, include the anonymized parallel submission [**?**] as additional material and cite it as

> [1] Authors. “The frobnicatable foo filter”, F&G 2012 Submission ID 324, Supplied as additional material `fg324.pdf`.

Finally, you may feel you need to tell the reader that more details can be found elsewhere, and refer them to a technical report. For conference submissions, the paper must stand on its own, and not *require* the reviewer to go to a techreport for further details. Thus, you may say in the body of the paper “further details may be found in [**?**]”. Then submit the techreport as additional material. Again, you may not assume the reviewers will read this material.

Sometimes your paper is about a problem which you tested using a tool which is widely known to be restricted to a single institution. For example, let’s say it’s 1969, you have solved a key problem on the Apollo lander, and you believe that the CVPR70 audience would like to hear about your solution. The work is a development of your celebrated 1968 paper entitled “Zero-g frobnication: How being the only people in the world with access to the Apollo lander source code makes us a wow at parties”, by Zeus *et al*.

Figure 5. Example of caption. It is set in Roman so that mathematics (always set in Roman: $B \sin A = A \sin B$) may be included without an ugly clash.

You can handle this paper like any other. Don’t write “We show how to improve our previous work [Anonymous, 1968]. This time we tested the algorithm on a lunar lander [name of lander removed for blind review]”. That would be silly, and would immediately identify the authors. Instead write the following:

> We describe a system for zero-g frobnication. This system is new because it handles the following cases: A, B. Previous systems [Zeus et al. 1968] didn’t handle case B properly. Ours handles it by including a foo term in the bar integral.
>
> ...
>
> The proposed system was integrated with the Apollo lunar lander, and went all the way to the moon, don’t you know. It displayed the following behaviours which show how well we solved cases A and B: ...

As you can see, the above text follows standard scientific convention, reads better than the first version, and does not explicitly name you as the authors. A reviewer might think it likely that the new paper was written by Zeus *et al*., but cannot make any decision based on that guess. He or she would have to be sure that no other authors could have been contracted to solve problem B.

FAQ: Are acknowledgements OK? No. Leave them for the final copy.

7.7. Miscellaneous

Compare the following:

`$conf_a$`	$conf_a$
`$\mathit{conf}_a$`	conf_a

See The TeXbook, p165.

The space after *e.g.*, meaning "for example", should not be a sentence-ending space. So *e.g.* is correct, *e.g.* is not. The provided `\eg` macro takes care of this.

When citing a multi-author paper, you may save space by using "et alia", shortened to "*et al.*" (not "*et. al.*" as "*et*" is a complete word.) However, use it only when there are three or more authors. Thus, the following is correct: " Frobnication has been trendy lately. It was introduced by Alpher [**?**], and subsequently developed by Alpher and Fotheringham-Smythe [**?**], and Alpher *et al.* [**?**]."

This is incorrect: "... subsequently developed by Alpher *et al.* [**?**] ..." because reference [**?**] has just two authors. If you use the `\etal` macro provided, then you need not worry about double periods when used at the end of a sentence as in Alpher *et al.*

For this citation style, keep multiple citations in numerical (not chronological) order, so prefer [**?**, **?**, **?**] to [**?**, **?**, **?**].

8. Formatting your paper

All text must be in a two-column format. The total allowable width of the text area is $6\frac{7}{8}$ inches (17.5 cm) wide by $8\frac{7}{8}$ inches (22.54 cm) high. Columns are to be $3\frac{1}{4}$ inches (8.25 cm) wide, with a $\frac{5}{16}$ inch (0.8 cm) space between them. The main title (on the first page) should begin 1.0 inch (2.54 cm) from the top edge of the page. The second and following pages should begin 1.0 inch (2.54 cm) from the top edge. On all pages, the bottom margin should be 1-1/8 inches (2.86 cm) from the bottom edge of the page for 8.5×11-inch paper; for A4 paper, approximately 1-5/8 inches (4.13 cm) from the bottom edge of the page.

8.1. Margins and page numbering

All printed material, including text, illustrations, and charts, must be kept within a print area 6-7/8 inches (17.5 cm) wide by 8-7/8 inches (22.54 cm) high.

8.2. Type-style and fonts

Wherever Times is specified, Times Roman may also be used. If neither is available on your word processor, please use the font closest in appearance to Times to which you have access.

MAIN TITLE. Center the title 1-3/8 inches (3.49 cm) from the top edge of the first page. The title should be in Times 14-point, boldface type. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave two blank lines after the title.

AUTHOR NAME(s) and AFFILIATION(s) are to be centered beneath the title and printed in Times 12-point, non-boldface type. This information is to be followed by two blank lines.

The ABSTRACT and MAIN TEXT are to be in a two-column format.

MAIN TEXT. Type main text in 10-point Times, single-spaced. Do NOT use double-spacing. All paragraphs should be indented 1 pica (approx. 1/6 inch or 0.422 cm). Make sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs.

Figure and table captions should be 9-point Roman type as in Figures 5 and 6. Short captions should be centred.

Callouts should be 9-point Helvetica, non-boldface type. Initially capitalize only the first word of section titles and first-, second-, and third-order headings.

FIRST-ORDER HEADINGS. (For example, **1. Introduction**) should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one blank line after.

SECOND-ORDER HEADINGS. (For example, **1.1. Database elements**) should be Times 11-point boldface, initially capitalized, flush left, with one blank line before, and one after. If you require a third-order heading (we discourage it), use 10-point Times, boldface, initially capitalized, flush left, preceded by one blank line, followed by a period and your text on the same line.

8.3. Footnotes

Please use footnotes[1] sparingly. Indeed, try to avoid footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence). If you wish to use a footnote, place it at the bottom of the column on the page on which it is referenced. Use Times 8-point type, single-spaced.

8.4. References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [**?**]. Where appropriate, include the name(s) of editors of referenced books.

8.5. Illustrations, graphs, and photographs

All graphics should be centered. Please ensure that any point you wish to make is resolvable in a printed copy of the paper. Resize fonts in figures to match the font in the body text, and choose line widths which render effectively in print. Many readers (and reviewers), even of an electronic

[1] This is what a footnote looks like. It often distracts the reader from the main flow of the argument.

Figure 6. Example of a short caption, which should be centered.

Method	Frobnability
Theirs	Frumpy
Yours	Frobbly
Ours	Makes one's heart Frob

Table 1. Results. Ours is better.

copy, will choose to print your paper in order to read it. You cannot insist that they do otherwise, and therefore must not assume that they can zoom in to see tiny details on a graphic.

When placing figures in LaTeX, it's almost always best to use `\includegraphics`, and to specify the figure width as a multiple of the line width as in the example below

```
\usepackage[dvips]{graphicx} ...
\includegraphics[width=0.8\linewidth]
                {myfile.eps}
```

8.6. Color

Color is valuable, and will be visible to readers of the electronic copy. However ensure that, when printed on a monochrome printer, no important information is lost by the conversion to grayscale.

9. Final copy

You must include your signed IEEE copyright release form when you submit your finished paper. We MUST have this form before your paper can be published in the proceedings.