

GEOMETRY AND TEXTURE MEASURES FOR INTERACTIVE VIRTUALIZED REALITY INDOOR MODELER

Kalaivani Thangamani^{a,*}, Ryosuke Ichikari^a, Takashi Okuma^a, Tomoya Ishikawa^b, Takeshi Kurata^a

^a Service SEnsing, Assimilation, and Modeling Research Group, Human Informatics Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Umezono, Tsukuba 305-8568, Japan - (thangamani.kalaivani, r.ichikari, takashi-okuma, t.kurata)@aist.go.jp

^b Site Sensing Inc., Japan - t.ishikawa@site-sensing.com

KEY WORDS: Virtualized reality indoor models, Texture distortion, Inpainting, Shape from Texture

ABSTRACT:

This paper discusses the algorithm to detect the distorted textures in the virtualized reality indoor models and automatically generate the necessary 3D planes to hold the undistorted textures. Virtualized reality (VR) interactive indoor modeler, our previous contribution enables the user to interactively create their desired indoor VR model from a single 2D image. The interactive modeler uses the projective texture mapping for mapping the textures over the manually created 3D planes. If the user has not created the necessary 3D planes, then the texture that belong to various objects are projected to the available 3D planes, which leads to the presence of distorted textures. In this paper, those distorted textures are detected automatically by the suitable principles from the shape from texture research. The texture distortion features such as the slant, tilt and the curvature parameters are calculated from the 2D image by means of affine transformation measured between the neighboring texture patches within the single image. This kind of affine transform calculation from a single image is useful in the case of deficient multiple view images. The usage of superpixels in clustering the textures corresponding to different objects, reduces the modeling labor cost. A standby database also stores the repeated basic textures that are found in the indoor model, and provides texture choices for the distorted floor, wall and other regions. Finally, this paper documents the prototype implementation and experiments with the automatic 3D plane creation and distortion detection with the above mentioned principles in the virtualized reality indoor environment.

1. INTRODUCTION

1.1 Virtualized reality indoor modeler

Virtualized Reality (VR) modeling generates photo-realistic 3D models from real-world clues such as photographs from the real scenes. These models enable the viewers to feel the virtual environment from any viewpoint. While virtual reality is simply the artificially created CAD models, the virtualized reality starts with the real scene and virtualizes it (Kanade1995). Generating photo-realistic indoor models from 2D digital photos and overcoming the textural constraints such as the texture distortion are discussed in this paper.

Interactive virtualized reality indoor modeler which was our previous contribution (Ishikawa2011), allows the user to create the VR model from a single 2D photo from the ordinary digital camera. The interactive VR indoor modeling tool starts the process by finding the vanishing point in the 2D image. The vanishing point in the 2D image is used to find the camera parameters and hence fixing the world coordinates for the VR model. The height of the ground, in the world space is interactively set by the user. The planes are added interactively in the world space by the user; Projective texture mapping is used to map the textures over the created 3D planes. The indoor models created from the single photos are integrated to generate the complete indoor environment. Figure 2 shows the local indoor model and the complete VR model for the Ganko restaurant chain, Tokyo.

The VR model faces the occlusion and the distortion which are explained in detail in the coming sections. There are often untextured regions / occluded regions on some of the 3D planes since it is not easy to take a set of photos so that every region of the

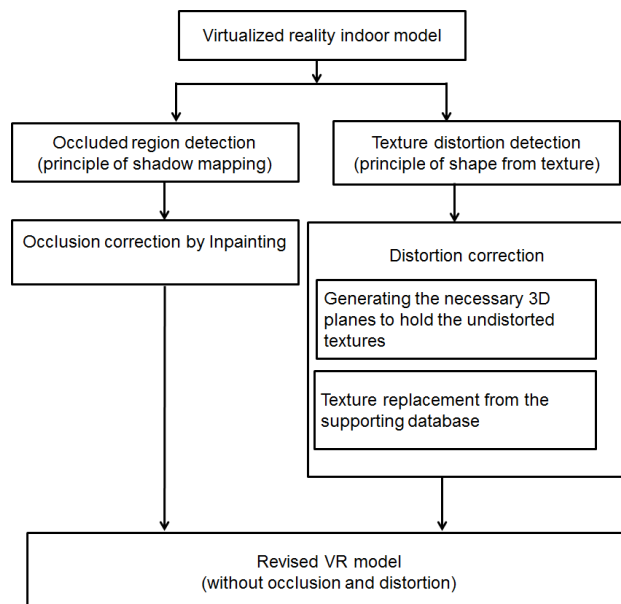


Figure 1: Virtualized reality indoor model and occlusion and distortion handling

*Corresponding author

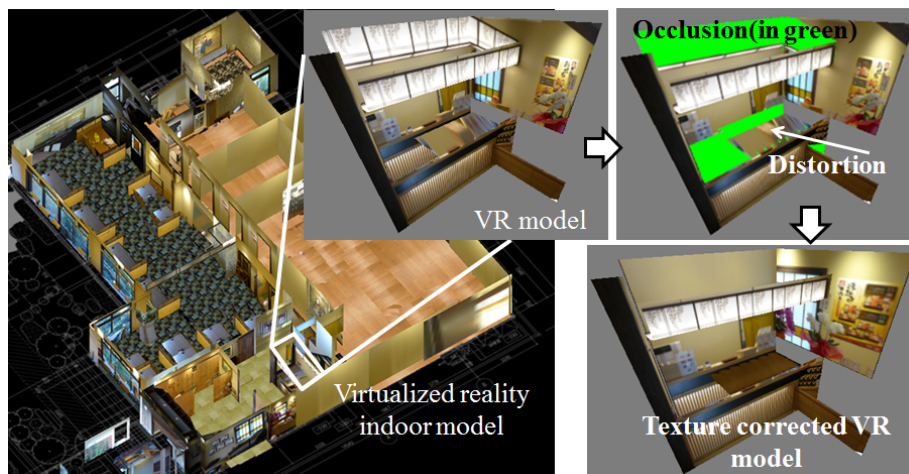


Figure 2: Virtualized reality indoor model and occlusion and distortion handling

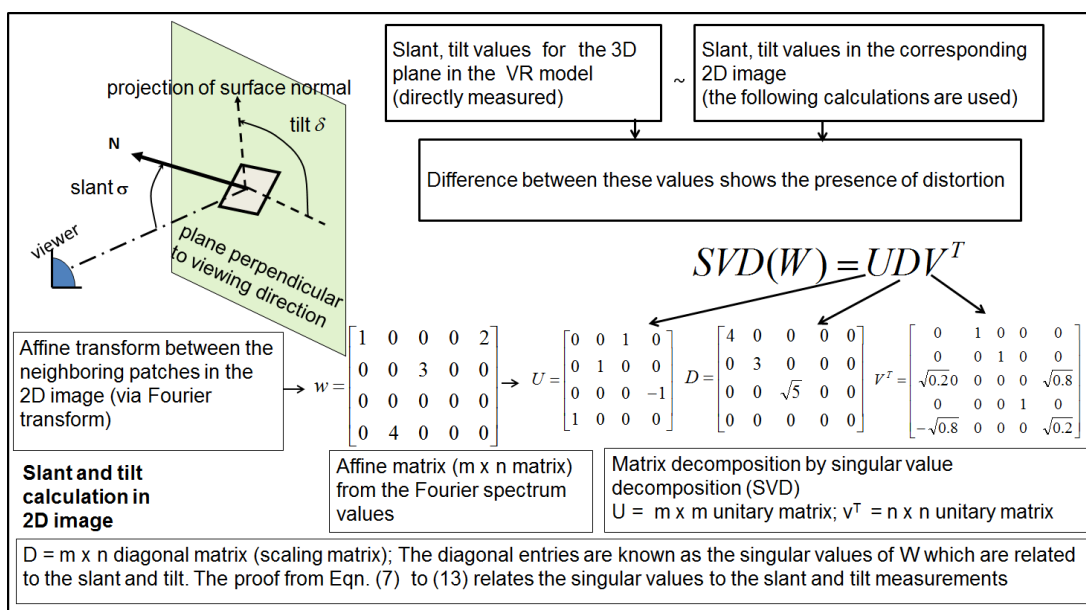


Figure 3: Distortion detection algorithm

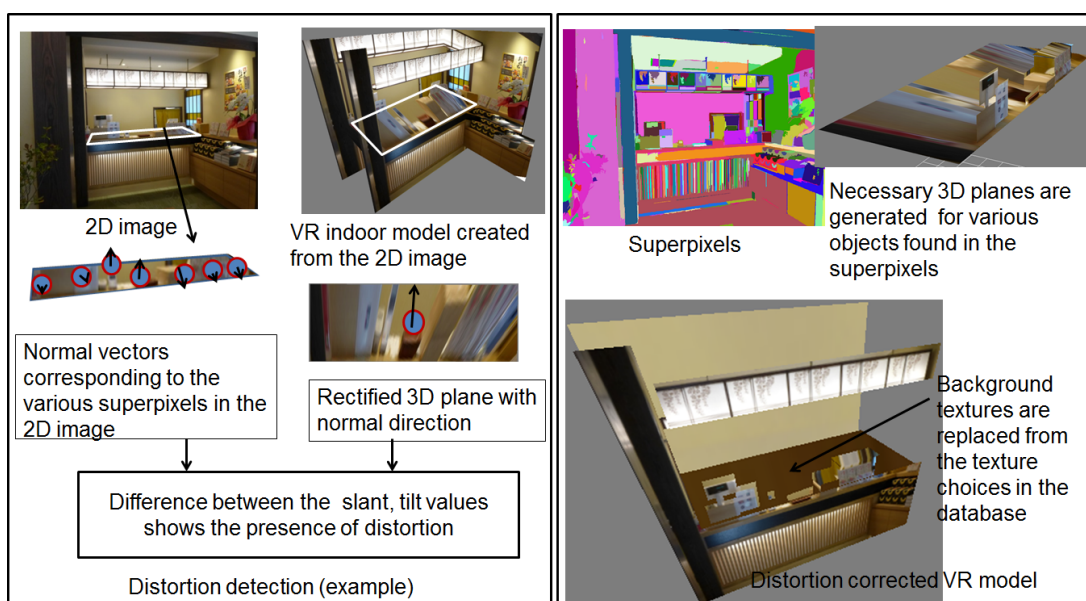


Figure 4: Virtualized reality indoor model and distortion handling

3D model is included at least in one of those photos. The occluded region is shaded in green in the Fig. 2 and these regions are textured with the Inpainting techniques (Thangamani2012). We have successfully tested the Inpainting algorithms to handle the occlusion in the VR indoor models.

The other problem in the VR model is the distorted textures. The users enjoy creating their indoor models but it is tedious for them to create every single object in the 2D photo. The lack of the necessary 3D planes causes the projected textures to get mapped on the available 3D planes. This action leads to the presence of distorted textures. Texture distortion are also caused by the image resolution used in the input photo. This paper discusses about the distortion caused by the lack of the necessary planes in the VR indoor model.

1.2 Literature review

The problem of texture distortion was referred in some of the 3D modeling works. Texture distortion was found in the works of Hoiem et al.(Hoiem2005), in their developed 3D model, from a single image. They separated the segmented pixel clusters in the single image, in to three categories namely ground, wall and the sky and map the textures to the horizontal, vertical and the sky planes. They maintained a learning system and presented the automatic 3D modeling. Due to the limited 3D planes, the mapped textures faced the problem of distortion.

While photographing the scene, the textures in the real world are rotated, translated, scaled and mapped as the 2D image. To back project the 2D textures once again to the 3D space, needs oneself to travel back the rotation, translation and scaling processes. One should able to decompose the mapped 2D texture matrix to the rotation, translation and scaling matrices and find the distortion parameters. This is the fundamental concept in the shape from texture research. The term "distortion parameters" used by the shape from texture research, refer the basic parameters that hold the relation between the 3D to 2D transformation. The works by Gibson(Gibson1950) is the starting point for this shape from texture, and structure from motion research. Later works by Garding (Garding1992),(Garding1993) shows the theoretical and experimental ways in realizing the structures from the texture. They measure the shape parameters by measuring the texture gradients in the image. Then the work by Malik (Malik1997) travels much closer to the shape from texture and calculate the distortion parameters by measuring the affine transform from the single image. This paper considers the work by Malik (Malik1997) as the basis for distortion detection in our modeling scenario.

Fourier transform was used by (Malik1997) for finding the affine transform between the neighboring patches within the single image. Fourier transform maps the pixel values in the frequency domain. Any raw data observed against time is called as the data in time domain, and for any analysis, the time domain has to be converted into the frequency domain. Frequency domain, states the number of times any particular action occurred; the frequency count. Fourier transform is one such time domain to frequency domain converter. The method of applying Fourier transform to find the affine matrix by (Lucchese1996) between the image pairs holds the good start for the Fourier transform usage. The similar textured region in the neighboring patches in the image, are considered for fixing the neighboring patches in the affine calculation. The idea of superpixels are useful for segmenting the similar pixel clusters. Superpixels (Ren2003) correspond to small, nearly-uniform segmented regions in the image. They are defined to be local, coherent and preserve most of the structure necessary for segmentation at the scale of interest. Automatic

pop-up by (Hoiem2005) used superpixels as one of their criterion in categorizing the pixel clusters in their 3D modeling works. The efficient graph based segmentation (Felzenswalb2004) is used for segmenting the image regions into superpixels. This graph based method segments the region based on the intensity, color, motion, location and other local attributes.

2. CONTRIBUTION AND OVERVIEW

This paper proposes the texture distortion algorithm derived from the shape from texture principles, for detecting the distorted textures, in our previous VR modeling (Ishikawa2011) system. The flowchart in Fig.1 shows the overall process flow for checking the geometry and texture in the VR indoor model generated from the modeling tool (Ishikawa2011). Once the user designs the VR model, it is analysed for the presence of occlusion and distortion. Occluded textures are found out by the concept of shadow mapping and handled by the Inpainting techniques. Inpainting process is not discussed in this paper since our successful results with Inpainting are discussed in detail in our previous works (Thangamani2012). Distortion is detected by the principles from shape from texture. The distortion features such as the slant, tilt and the curvature parameters define the 3D to 2D geometry and texture transformation. The slant and tilt are measured for the 3D planes in the VR model and the same are compared with their 2D superpixels. The degree of variation in the values, gives the degree of distorted textures. We are not considering the curvature parameters since our modeling system uses only the planar region for the VR modeling. The detected distortion parameters, together with the superpixel, are used for generating the necessary 3D planes for the various objects found in the 2D image. These newly generated planes hold the undistorted textures which are meant for them. The distorted background textures are replaced by the textures in the local database.

Figure 3 shows the definition for the slant and tilt, and explains the distortion detection algorithm in detailed steps. The slant angle σ is defined to be the angle between the surface normal N and the viewing direction p . Tilt δ is referred as an angle in positive x axis, during the parallel projection of the surface normal. These parameters are directly measured in the VR model, but certain techniques are to be followed in calculating for the 2D image. The superpixels corresponding to the 3D model plane are subjected to Fourier transform and the response are used to create the affine matrix. This affine matrix is decomposed by singular value decomposition. The singular values of the affine matrix are related to the slant and tilt parameters. The proofs in the next section explains the relation in detail. Once the distortion parameters are found out, they are used for generating the new 3D planes to hold the undistorted textures. The superpixels are used to define the various objects in the scene and helps in the automatic 3D planar region generation. There is also a supporting texture data base, formed by categorizing the available textures in the VR indoor model, based on their GLCM (Haralick1973) energy levels. This database suggests the possible texture choices for replacing the distorted background textures.

The example in fig. 4, shows the distortion detection and distortion handling in one such VR model. The 3D planes in the VR model is taken one after the other and their slant and tilt are measured (the sketch for explaining the slant and tilt in Fig. 3 shows the clue for the direct measurement). The superpixels corresponding to the rectified plane textures are subjected to the affine matrix calculation. The normal vectors are shown for the 2D region and the rectified 3D plane. The various superpixels in the 2D region hold the normals in different directions. But

these textures are mapped to the single plane whose normal is also shown for comparison. The difference between the distortion parameters shows the presence of distortion. The detected slant and tilt values are used for generating the necessary 3D planes to hold the undistorted textures. The distorted background textures are replaced by the textures from the local database.

3. SHAPE FROM TEXTURE- REVIEW

There are totally 5 parameters which are said to be the distortion parameters; namely, the slant σ , the tilt δ and the 3 curvature parameters (normal curvature along the slant direction, normal curvature along the tilt direction, geodesic torsion). These are the parameters that hold the relation between the 3D and 2D mapping. The slant angle σ is defined to be the angle between the surface normal N and the viewing direction p . Tilt δ is referred as an angle in positive x axis, during the parallel projection of the surface normal.

3.1 Relationship between the texture distortion map and 3D shape

The following principles and formulae are derived from the works by (Malik1994). The reader is referred to access (Malik1997), (Garding1992), (Garding1993), for the complete proof. The texture distortion in an particular direction on the image plane is measured as an affine transformation between a pair of image patches. The following derivation are written for the spherical projection but the VR modeling uses the planar projection.

The relation between the slant and tilt are discussed as follows. When a smooth surface S is mapped by central projection to a unit sphere \sum centered at a focal point, the back projection map F from \sum to S is defined as $F(p) = r(p) = r(p)p$ where p is a unit vector from the focal point to a point on the image sphere, and $r(p)$ is the distance along the visual ray from the focal point through p to the corresponding point $r = F(p)$ on the surface S .

The slant angle σ is defined to be the angle between the surface normal N and the viewing direction p , so that $\cos\sigma = N.p$.

$F * (p)$ (differential of $F(p)$) can be expressed as

$$F * (p) = \begin{bmatrix} r/\cos\sigma & 0 \\ 0 & r \end{bmatrix} = \begin{bmatrix} 1/m_p & 0 \\ 0 & 1/M_p \end{bmatrix} \quad (1)$$

where r is the distance to the object from the center of the viewing sphere, and σ is the slant angle. We see that $m_p = \cos\sigma/r$ is the scaling of the texture pattern in the minor axis i.e., in the tilt direction and $M_p = 1/r$ is the scaling in the major axis i.e., in the orthogonal direction. The shape of the surface is captured in the shape operator, which measures how the surface normal N changes as one moves in various directions in the tangent space of the surface.

The shape operator, in the (T, B) basis as

$$\begin{bmatrix} -\nabla_T N \\ -\nabla_B N \end{bmatrix} (p) = \begin{bmatrix} k_t & \tau \\ \tau & k_b \end{bmatrix} \begin{bmatrix} T \\ B \end{bmatrix} \quad (2)$$

where k_t is the normal curvature in the T (tilt) direction, k_b the normal curvature in the B (slant) direction and τ is the geodesic torsion.

Our aim is to find the matrix A , which represents the affine transformation between the spherically projected texture between the two nearby pixel points. This matrix will be a function of the local parameters and shape parameters. The orientation parameters are σ , the slant of the surface and t , the direction of tilt of the surface. The shape parameters are rk_t , rk_b , and $r\tau$. The variable r is the distance from the center of the viewing sphere to the given point on the surface.

Malik et al., derives the A as follows,

$$A = Rot(\delta_t) \cdot \begin{bmatrix} k_m \cos\delta_T & k_m \sin\delta_T \cos\sigma \\ k_M \sin\delta_T / \cos\sigma & k_M \cos\delta_T \end{bmatrix} \quad (3)$$

where

$$k_m = 1 + \nabla m / m \circ \begin{bmatrix} \nabla t \\ \nabla b \end{bmatrix} \quad (4)$$

$$k_M = 1 + \nabla M / M \circ \begin{bmatrix} \nabla t \\ \nabla b \end{bmatrix} \quad (5)$$

We find the affine transformation between the neighboring patches in the superpixels. The affine transformation will depend on the direction and magnitude of the vector displacement between the two patches in the superpixel. Affine transforms are estimated in a number of directions to recover the surface orientation and shape parameters. Shape recovery via affine transformation is the useful suggestion by Malik et al., (Malik1997) as it successfully overcomes the traditional formulation of applying texture gradients to find the shape parameters. The actual affine transformation will not be in terms of the (t, b) basis. Instead, the affine matrix is related as, $\hat{A} = UAU^{-1}$. Here comes the necessity of the Singular Value Decomposition (SVD) to realize the individual matrices. The rotation matrix, U , is given by

$$Rot(\theta_t) = \begin{bmatrix} \cos\theta_t & -\sin\theta_t \\ \sin\theta_t & \cos\theta_t \end{bmatrix} \quad (6)$$

where θ_t is the tilt angle at the point of interest.

3.2 Shape recovery by SVD

There are five unknowns, the slant σ , the tilt direction θ_t and the three shape parameters $(rk_t, rk_b, r\tau)$. Each estimation of an affine transform in an image direction yields four non-linear equations. Two directions are sufficient for solving the equations and finding the distortion parameters. Malik et al.,(Malik1997) proposed two shape recovery algorithms. The first is the linear algorithm based on singular value decomposition (SVD) of the affine matrices. The second algorithm is based on the least squares formulation of an error criterion.

Singular value decomposition (SVD) is adapted for this work. The simple logic of the SVD shows that the 2D image is the combination of the translation, rotation and scaling. If the resultant matrix is decomposed, then the diagonal matrix represents the scaling in the vertical and the horizontal axis (minor and major axes). These singular values can be processed further to find the slant and the tilt angle for the particular patch of interest. If the patches in the superpixels are subjected for their slant and tilt test, the orientation of the corresponding objects are found out.

The following derivation part relates the singular values to the minor and major scaling axes. Malik et al., justify this by noting that the singular values S_1, S_2 of matrix \hat{A}_i are related to the eigen values λ_1, λ_2 of the matrix $\hat{A}_i^T \hat{A}_i$ by the relationship $\lambda_1 = S_1^2$ and $\lambda_2 = S_2^2$. Then expressions are computed for $\text{trace}(\hat{A}_i^T \hat{A}_i) = \text{trace}(A_i^T A_i)$ and $\text{det}(\hat{A}_i^T \hat{A}_i) = \text{det}(A_i^T A_i)$. Dropping the subscripts and the superscripts i for a better view,

$$S_1^2 + S_2^2 = \text{trace}(A^T A) \quad (7)$$

$$= k_m^2 \cos^2 \delta_T + k_M^2 \sin^2 \delta_T / \cos^2 \sigma + k_m^2 \sin^2 \delta_T \cos^2 \sigma + k_M^2 \cos^2 \delta_T \quad (8)$$

and

$$s_1^2 s_2^2 = \text{det}(A_i^T A_i) = k_m^2 k_M^2 \quad (9)$$

$$(s_1 + s_2)^2 = \text{trace}(A^T A) + 2\sqrt{\text{det}(A^T A)} \quad (10)$$

$$= k_m^2 \cos^2 \delta_T + k_M^2 \sin^2 \delta_T / \cos^2 \sigma + k_m^2 \sin^2 \delta_T \cos^2 \sigma + k_M^2 \cos^2 \delta_T + 2k_m k_M (\cos^2 \delta_T + \sin^2 \delta_T) \quad (11)$$

$$= (k_m + k_M)^2 - k_m^2 \sin^2 \delta_T (1 - \cos^2 \sigma) + k_M^2 \sin^2 \delta_T ((1/\cos^2 \sigma) - 1) \quad (12)$$

$$= (k_m + k_M)^2 - (k_m \sin \delta_T \sin \sigma)^2 + (k_M \sin \delta_T \tan \sigma)^2 \quad (13)$$

This suggests that we can approximate $s_1 + s_2 \approx k_m + k_M$. So we can estimate k_m and k_M as the singular values of \bar{A} . Then the slant and tilt direction are found from the k_m and k_M values. This assumption is true for any planar surface and for the curved surfaces.

4. ADAPTING THE SHAPE FROM TEXTURE PRINCIPLE IN THE VR MODELING

The proofs stated in the previous sections are aimed at the spherical projection but our VR modeler uses the planar projection. The conversion between these two are necessary for realizing the application.

The 2D image is segmented into superpixels and they are subjected to the affine transformation. The size of the superpixel should be reasonable so that it could afford for the least patch size, set up for the Fourier transform process. The VR modeling system uses its own technique (Ishikawa2011) to fix the orientation for the 3D plane generation. These techniques are collaborated with the distortion calculation measures, so that the calculated slant and the tilt values after the SVD process can be used for deciding the orientation of the 3D plane. The modeling tool suggests the user to click over the object boundary to create a new plane, but these steps are made automatic by upgrading the modeling system to use the superpixel boundary for the creation of the new plane.

5. TEXTURE DATABASE

In the indoor VR model, there are many repeated textures which are meant for the floor, wall and other objects. The undistorted texture samples can be stored and used for the places of need.

Designing a texture database to maintain the textures in the entire model is really useful. Kholgade et al. (Kholgade2014) uses the textures from the internet database to fill the occluded textures in their 3D model. We have already introduced the GLCM (Haralick1970) based texture categorization for fast inpainting process (Thangamani2012) in the VR model. The same can be extended to categorize and sort the repeated textures in the VR model. The texture patches are classified and stored in the hash table for quick search during the distortion handling processes. The hash function relates the incoming texture patch to the address of the particular hash table bin. During the query step, the necessary patch is retrieved by a single/ minimum search. The texture information is adequately specified by a set of gray tone spatial dependence matrices otherwise known as Gray level co-occurrence matrix (Haralick1970) which are computed for various angular relationships ($0^\circ, 45^\circ, 90^\circ$ or 135°) and distances between neighbouring resolution pairs on the image. More details can be found in (Thangamani2012) for categorizing the texture patches and storing in the hash table.

6. RESULTS AND DISCUSSION

This paper detects the texture distortion and removes the distorted textures by generating the necessary planar regions so that every plane holds the undistorted texture. The principles from shape from texture, usage of superpixels together with the features of the modeling tool enabled the automatic 3D plane generation. The database formed from the available textures are also referred for good choices for the floor and the wall textures. The proposed prototype system is tested in the virtualized reality model and the preliminary results are stated in the table 1. Qualitative analysis in table 1, compares the VR local model before and after texture treatment (occlusion and distortion detection). The individual 3D planes in the VR model are shown separately. The planes that are occluded are inpainted and the distorted ones are replaced by the similar textures in the groups. In some cases, the occluded textures are also replaced by the similar textures in the groups. The textures which do not need any change are preserved as such for the final model. There are also shortcomings, found in our way of calculating the affine transformation from a single image, due to the deficient patch size in some of the superpixels. Our algorithm should be extended and optimized to handle the variety of input targets and modeling conditions.

7. CONCLUSION AND FUTURE WORKS

This paper proposed the algorithm for detecting the distorted textures in the 3D planes of the virtualized reality indoor model. Out of the five distortion parameters, only the slant and tilt parameters are experimented for automatic creation of the planar regions in our VR tool. Extending the trials to use all the five distortion parameters to design the non-planar regions and mesh surfaces would be of our future work. In that case, we need to adapt the suitable texture mapping method; the simple projective texture mapping is not enough to handle the regions that are out of the view zone. The techniques such as the PPTM (Parallel Projective Texture Mapping) are of good choice for projecting the textures over the non-planar regions.

Currently, our database is designed to hold the textures from the indoor model and these textures are used as the choices for replacing the distorted ones. This database can be extended to use the texture reference from the internet such as the SNS (Social Networking Services) textures. The textures from the internet can also be categorized by the GLCM filters and stored in the hash bins for our texture reference.

Table 1: Qualitative analysis

2D image used	VR local model (before texture treatment)	Depth map	VR local model with occluded (in green) and distorted region	Rectified planes in the local model			VR local model (after texture treatment)
				Before and after inpainting	Before and after distortion handling	That did not undergo any texture correction	
						All the textures underwent texture correction	
						All the textures underwent texture correction	

ACKNOWLEDGEMENTS

This research is partly supported by Grant-in-Aid for Young Scientists (B) No.26870905.

REFERENCES

Kanade, T., 1995. Virtualized Reality: Concepts and Early Results. *IEEE Workshop on the Representatives of Visual Scenes*.

Ishikawa, T., Thangamani, K., Kourogi, M., Gee, A.P., Mayol-Cuevas, W., Hyun, J., Kurata, T., 2011. Interactive 3-D indoor modeler for virtualizing service fields. *Int. J. Virtual Reality, SI: Mixed and Augmented Reality*, 19, ISSN. 1359-4338, pp. 1-21.

Thangamani, K., Ishikawa, T., Makita, K., Kurata, T., 2012. Fast and texture-structure preservable inpainting and its application for creating virtualized-reality indoor models. *Int. J. Computer science and informatics*, 1(3), pp. 85-101.

Hoiem, D., Fros, A.A., Hebert, M., 2005. Automatic photo pop-up. *ACM Transactions on Graphics*, 24(3), pp. 577-584.

Gibson, J., 1950. The perception of the visual world. *Boston: Houghton Mifflin*

Garding, J., 1992. Shape from texture for smooth curved surfaces in perspective projection. *J. of mathematical imaging and vision*, 2(4), pp. 327-350.

Garding, J., 1993. Shape from texture and contour by weak isotropy. *J. of Artificial intelligence*, 64(2), pp. 243-297.

Malik, J., and Rosenholtz, R., 1997. Computing local surface orientation and shape from texture for curved surfaces. *International journal of computer vision*, 23(2), pp. 149-168.

Lucchese, L., 1996. Estimation of affine transformations between images pairs via Fourier transform. *Proceedings of International conference on image processing*, 3, pp. 715-718.

Ren, X., and Malik, J., 2003. Learning a classification model for segmentation. *Proceedings of the 9th International conference on computer vision*, 1, pp. 10-17.

Felzenswalb, P.F., and Huttenlocher, D.P., 2004. Efficient graph-based image segmentation. *International Journal of computer vision*, 59(2), pp. 167-181.

Haralick, R.M., 1973. Textural features for image classification. *IEEE T. Systems, Man and cybernetics*, 3(6), pp. 610-621.

Kholgade, N., 2014. 3D object manipulation in a single photograph using stock 3D models. *SIGGRAPH 2014*.