VCIP '97, Conference on Visual Communications and Image Processing, SPIE Symposium on Electronic Imaging, Science and Technology, San Jose, CA, February 1997

3D-3D registration of free formed objects using shape and texture

Fernando C. M. Martins[†*], *Hirohisa Shiojiri*[‡], *and José M. F. Moura*[†]

[†] *Department of Electrical and Computer Engineering,*
Carnegie Mellon University
Pittsburgh, PA 15213–3890.

[‡] *NEC 2nd Development Department, 2nd Transmission Division*
1753 Shimonumabe, Nakahara-ku
Kawasaki, Kanagawa 211 Japan.

November 15, 1996

ABSTRACT

Automatic generation of textured object models from a sequence of range and color images requires two major tasks: measurement registration and measurement integration. *Measurement registration* is the estimation of the current position and orientation of the object in 3D space with respect to an arbitrary fixed reference, given the current measurement and the 3D object model under construction. *Measurement integration* is the updating of the 3D object model using the current registered measurement.

In this paper we present an iterative 3D-3D registration technique that uses both texture[1] and shape information available in the 3D object models and the 3D measurements. The proposed technique handles probabilistic models that are potentially incomplete before the measurement integration step. Measurements are acquired via a sensor characterized by a probabilistic sensor model. The object models are constructed automatically without user interaction. Each model is a compact uniform tessellation of 3D space, where each cell of the tessellation represents shape and texture in a probabilistic fashion. Free formed objects are supported and no prior knowledge about the object shape, texture or pose is assumed. Traditional registration methods consider only shape and geometric information. We consider texture information as an additional evidence by defining a generalized inter-cell distance measure that considers both the relative positioning of cells in space and the texture discrepancy between cells. Experimental results demonstrate the efficiency and robustness of the proposed method. The usefulness of texture in registration is highlighted in a comparison with results obtained considering only geometric information.

Keywords: 3D-3D registration, 3D pose estimation, 3D object modeling, Textured geometric models, Range and image sequence processing.

*Work of first author partially supported by CNPq - National Council for Scientific and Technologic Development, Brazil.

[1] Throughout this paper the term "Texture" refers to the light captured by the camera after reflecting on the given object surface. Also known as object radiance, "Texture" depends on object surface photometric properties and environment illumination.

1 INTRODUCTION

The problem of 3D-3D registration arises in several applications, particularly where accurate 3D geometric object models must be constructed from a sequence of non-registered range and color images. Examples of such applications are map construction for navigation, model based object recognition, computer assisted surgery, product inspection, and model based representations for video.

In general, the process of model construction requires two steps for each new measurement: *Registration* and *Integration*. *Measurement registration* is the problem of estimating the current position and orientation of the object in 3D space with respect to an arbitrarily fixed coordinate system, given the current measurement and a 3D geometric model. Registration is a problem dual to the problem of computing point correspondences between overlapping regions of the object model and the current measurement. Given a set of point correspondences, there is a closed form solution to the pose problem.[1,2] With precise correspondences it is possible to compute a mapping to juxtapose the new measurement to an existing object model. This duality highlights the importance of accurate registration for successful model generation from observations.

Registration is usually accomplished by minimizing a cost function based on discrepancy metrics between potential model-measurement correspondences. Traditional systems use only shape information and adopt Euclidean distances as measure of discrepancy. Our system employs a novel generalized distance that uses both shape and texture information.

Each new measurement must have some overlapping area with the model in order to allow the establishment of a large set of reliable correspondences. This assumption is satisfied if data is acquired such that motion is small between successive measurements. This is usually the case when data is acquired at video rates.

The Iterative Closest Point Algorithm (ICP), introduced simultaneously by several groups, solves the discrepancy minimization problem by hypothesizing correspondences to be the closest points iteratively.[3–6] An important issue not addressed by most ICP implementations is the handling of incomplete models, where measurement-model correspondences may not exist for all individual measurements. This is frequently the case in applications where registration is performed when models are still being constructed. The proposed algorithm handles incomplete models.

Some extensions of ICP reduce the cost of closest point search by operating only on selected features, or by creating search index trees. A fast implementation of the ICP algorithm capable of pose estimation at 10Hz has been reported.[7] Recently, ICP has been extended to employ a generalized distance function that considers surface normals as additional cues for increased reliability in pose estimation.[8]

Other registration and motion estimation methods are based on features and on factorization. These methods require precise feature extraction and the solution of large dimensional singular value decompositions, which are computationally expensive tasks. For surveys on registration methods please see references.[3,9,10]

Measurement integration, is the process of altering the model according to the contributions of the new registered measurement. The first task in this process is to map the registered measurement to a canonical fixed reference frame using the pose estimate computed in the registration step.

The actual integration procedure is heavily dependent on model structure. Distinct deterministic techniques have been proposed for the integration of range measurements into surface based models.[11–14] Stochastic integration techniques consider a probabilistic sensor model and are based on Bayesian updating or Kalman filtering.[15,16] In these techniques, redundant measurements help to reduce overall model entropy, while conflicting or ambiguous measurements are handled gracefully. We are primarily interested on representations for video using compact textured 3D models, i.e., geometric models that also carry texture information to allow a perceptually acceptable image reconstruction.

In this paper we propose and demonstrate the efficiency of an iterative 3D-3D registration technique based on shape and texture information, capable of handling potentially incomplete models. The paper is organized as follows: Section 2 describes the information used by the registration method. In section 3 we describe the proposed 3D-3D registration method. The algorithm description and some relevant implementation details are discussed in section 4. In section 5, the experimental results that illustrate the efficiency of the proposed method are presented. Section 6 concludes the paper.

2 AVAILABLE INPUT DATA

The proposed registration method computes a pose estimate for an object in the scene given a set of measurements and a 3D object model. In this section, we describe the input data required by the proposed registration method.

2.1 Depth and Intensity Measurements

Each measurement with a light stripe range sensor produces a depth map R and a co-registered image I. By co-registered we mean that the pair of measurements is taken with respect to the same reference, through the same camera, and for every intensity image pixel there is a corresponding depth measurement and vice versa.

A range measurement R is considered a set of points $\{\vec{r}\}$ in space. Similarly an image I is defined as an organized set of color or intensity measurements $\{\vec{t}\}$, where each measurement $\vec{t}$ can assume one of the colors listed in pallete $\mathcal{C}$. For color images the pallete is a set of tridimensional vectors, and for gray level images it is a set of scalars. A 3D probabilistic sensor model $p(\vec{r}|\vec{z})$ characterizes the uncertainties in data obtained through the range finder. This sensor model can be obtained experimentally.

Intensity (or color) images provide information about the texture already mapped to the object's shape. Texture information depends on surface material properties, viewing direction, and environment illumination.

2.2 3D Object Model Structure

The tridimensional non-parametric object model is voxel based. It is defined as a compact uniform tessellation of 3D space $\Gamma = \{C_i\}$, where each cell C_i represents multiple properties. Object shape is represented by cell occupancy $O(C_i)$ and object surface texture is represented by cell texture $T(C_i)$. As before, occupancy may assume the value of *occupied* or *empty*, while texture may assume one of the valid values in the color pallete $\mathcal{C}$.

Each cell C_i stores a probability distribution for occupancy, i.e., $p(O(C_i) = \textit{occupied}|\{\Psi_k\})$ that is obtained through the integration of multiple measurement grids Ψ_k. Holding probability distributions instead of current estimates is what makes Γ a useful representation for Bayesian integration of a sequence of measurements. Initial lack of knowledge is expressed by assigning equiprobable probability density functions.

Earlier work on sensor fusion for robot navigation and object modeling for robotic manipulation have successfully explored this stochastic model structure.[15] For details about incremental object model construction please refer to prior work on model-based video representations by Martins and Moura.[16]

3 PROPOSED POSE ESTIMATION TECHNIQUE

This section presents our method to compute the pose estimate given a set of measurements and a 3D object model. The process of 3D-3D registration requires the following two steps for each new measurement. *Measurement pre-processing* **:** The available depth and intensity measurements are organized into a uniform measurement grid. This pre-processing step greatly simplifies the other two tasks in video analysis by eliminating later concerns with camera pose, geometry, and sensor model. *Measurement Grid-Model registration* **:** This step computes the current object pose given both the current measurement grid and the 3D geometric model under construction.

3.1 Measurement Preprocessing - Construction of a Uniform Measurement Grid

The measurement grid Ψ is an auxiliary data structure that contains all the information and only the information available in a given pair of range R and co-registered intensity I measurements.

The creation of a uniform measurement grid is a preprocessing step that greatly simplifies the registration by eliminating later concerns with camera pose, camera geometry and sensor model. As the sensor output is a set of discrete measurements, each individual measurement $(\vec{r}, \vec{t})$ corresponds to a volumetric lattice in the compact 3D space. Under the assumption of a pinhole perspective camera, the size of each corresponding lattice depends on the depth, please see figure 1. Therefore, a pinhole perspective camera defines a non-uniform tessellation of 3D space due to perspective non-linearities.

Dealing directly with these non-uniform tessellations is inconvenient because the tessellation depends on camera geometry and relative object-camera position; and model construction would require computationally complex raycasting and resampling.

To avoid these difficulties, we collect the information available in the co-registered measurements (R_k, I_k) into a uniform measurement grid Ψ_k, as in figure 2, taking into account the camera geometry and the sensor model.

We define a measurement grid Ψ as a compact uniform tessellation of 3D space $\Psi = \{M_i\}$, where shape is represented by the probability distribution of occupancy $p(O(M_i) = \mathit{occupied}|(R, I))$ and texture is represented by cell texture $T(M_i)$. Occupancy is a binary property, such that a cell may be either *occupied* or *empty*. The color pallete $\mathcal{C}$. contains valid values for texture.

With camera calibration and sensor model assumed known, the range R and the image I are used to compute the probability of occupancy $p(O(M_i) = \mathit{occupied}|(R, I))$ and the mapped texture $T(M_i)$ for each cell M_i of the measurement grid Ψ.

The generation of the measurement grid Ψ is realized in two steps. First, the measurement data are used to create a warped temporary grid Υ. Then the grid Υ is unwarped and resampled to generate the uniform tessellation Ψ. The mapping between Υ and Ψ is defined such that the camera frustum (a pyramid of rectangular base) in the uniform grid Ψ is mapped to a rectangular orthogonal prism in Υ. As we assume a perspective pinhole camera model, in Ψ all rays that reach pixels in the retinal plane pass through the camera center forming the camera frustum. Through this warping, these rays become parallel in Υ and reach the retinal plane perpendicularly. This mapping is used to cancel the effects of perspective and provide a properly resampled grid Ψ constructed from measurements taken using a perspective camera.

This method of 3D model generation is a reversed version of the shear-warp volume rendering algorithm, where instead of synthesizing 2D views from 3D models, 3D volumetric representations are created from 2D measurements.[12]

Figure 1: The perspective pinhole camera model establishes a non-uniform tessellation of the 3D space.

Figure 2: The uniform measurement grid avoids dealing with perspective non-linearities.

3.2 Measurement Grid - Model Registration

The proposed method for 3D-3D registration is an extension of the ICP algorithm that uses texture and shape for pose estimation and deals with incomplete or inconsistent models.

To consider texture information, we define an intercell discrepancy metric $d(C_i, C_j)$ as:

$$d(C_i, C_j) = (1 - \lambda) \| \vec{r_i} - \vec{r_j} \|^2 + \lambda \| \vec{t_i} - \vec{t_j} \|^2 \tag{1}$$

where $\vec{r_i}$ is the position of cell C_i in 3D with respect to a canonical fixed coordinate system, and $\vec{t_i}$ is the color of cell C_i. The parameter λ controls the relative importance of shape and texture in the discrepancy criterion. The discrepancy metric $d(C_i, C_j)$ is a generalized distance between two textured cells in 3D space, where both positions $\vec{r_i}$ and $\vec{r_j}$ must be taken with respect to the same arbitrarily fixed reference.

Given an arbitrary pose $\vec{q}$, the generalized distance $D(\Gamma, \Psi_k, \vec{q})$ between an object model Γ and the k^{th} measurement grid Ψ_k is defined as:

$$D(\Gamma, \Psi_k, \vec{q}) = \sum_{\{(i,j)|(C_i, M_j) \in \mathcal{S}\}} d(C_i, \mathcal{M}_{\vec{q}}(M_j)) \tag{2}$$

where C_i is a cell from model Γ, M_j is the corresponding cell from measurement grid Ψ_k, $\mathcal{S}$ is the set of correspondences (C_i, M_j), and $\mathcal{M}_{\vec{q}}$ is the mapping derived from the arbitrary pose $\vec{q}$. The set of correspondences $\mathcal{S}$ may also depend on the pose $\vec{q}$.

To handle models that are potentially incomplete, we modify the definition of the generalized distance presented in equation 2 to consider only correspondences whose distances lie within a plausible range d_{max} as in equation 3.

$$D(\Gamma, \Psi_k, \vec{q}) = \sum_{\{(i,j)|(C_i, M_j) \in \mathcal{S}, d < d_{max}\}} d(C_i, \mathcal{M}_{\vec{q}}(M_j)) \tag{3}$$

This modification removes false correspondences and outliers from the computation of the generalized distance. As the motion is assumed small, correspondences based on shape cannot be too far apart. Texture variations due to differences in illumination are usually very small. Larger differences in texture, as specular reflection highlights, are localized spots that are considered outliers according to this improved criterion. Similar outlier removal techniques have been successfully applied to pose estimation based only on shape information.[6]

The pose estimate is then given by the argument $\vec{q_k}$ that minimizes the generalized distance:

$$\vec{q_k} = \arg\min_{\vec{q}}\{D(\Gamma, \Psi_k, \vec{q})\} \tag{4}$$

4 ALGORITHM

The proposed 3D-3D registration method leads to the minimization problem in equation 4. The ICP algorithm solves the minimization problem by iteratively alternating between the following assumptions.

First the algorithm assumes the object pose known in order to compute the model-measurement correspondences as the closest points. Then it assumes the set of point correspondences known in order to compute the closed form solution for pose.

Before the first measurement Ψ_0, the model Γ holds no information. The object is assumed to be initially in an arbitrary canonical pose $\vec{q_0}$ that will be used as a fixed canonical reference.

To compute the pose estimate $\vec{q_k}$ for the subsequent measurements Ψ_k, the minimization in equation 4 must be solved. We propose the following extension of the iterative ICP algorithm to solve the minimization problem:

begin

`Initial pose` $\vec{q}$:= `Initial guess` $\vec{q}_{k-1}$

while `Generalized distance` $D(\Gamma, \Psi_k, \vec{q}) > \xi$ `and Iteration counter` $K < K_{max}$

`Compute set of correspondences` $\mathcal{S}$ `as the set of points in model` Γ `closest to the registered measurements` $\mathcal{M}_{\vec{q}}(\Psi_k)$.

`Compute new pose estimate` $\vec{q}$ `as the closed form solution for pose given by the set of correspondences` $\mathcal{S}$.

end

end

4.1 Initial Guess for Pose Estimation

As the algorithm performs a local minimization, a good initial estimate is important to ensure reliable results. When processing a video sequence with small interframe motion, a good initial estimate for pose for a given frame F_k is the pose estimate computed for the previous frame $\vec{q}_{k-1}$.

4.2 Algorithm Parameters

Parameter λ controls the relative importance of shape and texture in the discrepancy criterion in equation 1. If initial estimates for pose are not precise and object texture presents fine details, the lack of spatial smoothness in texture may become a potential hazard to algorithm reliability. To overcome this problem, it is possible to adopt the following schedule for increasing λ in order to speedup convergence and increase the quality of

the final results. At the beginning, more weight is given to shape, i.e., $\lambda \approx 0$. As the generalized distance $D(\Gamma, \Psi_k, \vec{q})$ reduces, λ is progressively increased, with more strength given to texture in order to fine tune the pose. Alternative scheduling policies can be devised to suit distinct setups.

Parameter d_{max} is the threshold for outlier removal in equation 3. This parameter controls which measurements are considered for registration. A larger value for d_{max} entails that more measurements will be considered in the computation of the generalized distance.

Two parameters control the stopping criterion of the main iterative loop. Parameter ξ is the precision component in the criterion. It is usually a small number used as a threshold on the generalized distance to stop the iterations when reached. Parameter K_{max} is the iteration count threshold. It assures that the loop finishes, even if the precision bound ξ is not reached.

4.3 Four Dimensional Binary Tree

Binary index trees are useful structures to speedup read transactions. To reduce the computational cost of searching for the closest cell, we construct an auxiliary index tree for the model data.

For grey level textures, a four dimensional binary tree is adopted, where three of the dimensions are related to shape via the spatial coordinates $\vec{r}_i = (x_i, y_i, z_i)$ and the fourth dimension is associated with texture t_i . All variables are normalized to span the interval $[0, 1]$.

The tree is constructed as follows. All cells in the model are initially associated with the root node. For a given tree node, and corresponding data cluster, the means and variances of all variables (x_i, y_i, z_i, t_i) are computed. The variable with largest variance among the four variables is chosen as a node discriminant. The mean of the node discriminant is then used to partition the data in two clusters. This discriminant selection and partitioning procedure is recursively repeated until clusters are small enough for an efficient linear search. The only values stored in the tree node are the node discriminant variable name, the value of the discriminant's mean, and two pointers to nodes in the following layer.

For a given measurement cell $\mathcal{M}_{\vec{q}}(M_j)$, searching for a nearest model cell C_i becomes a process of navigation through the index tree, followed by a linear search on the cluster pointed by the leaf node. If the cell C_i lies near the border with other clusters, those neighboring cells are also included in the linear search to increase precision in the matching.

For color textures, we similarly construct a six dimensional binary tree with 3 dimensions related to shape and 3 dimensions related to texture color components (Y_i, I_i, Q_i) .

5 EXPERIMENTAL RESULTS

We use a raytracer to generate synthetic sequences with available ground truth for pose and controllable environment conditions. Real data sequences acquired using a laser stripe range finder are used to test the proposed method in a real indoors setup.

The first experiment uses *LuxoSr* , a synthetic sequence of a lamp performing a rotation around the horizontal axis. The complexity of shape, lack of distinctive surface texture, and self-occluding motion pattern lead to a challenging pose estimation problem.

Figure 3: Error in pose estimation for all frames of *LuxoSr*.

(a)

(b)

frame 3 frame 14 frame 16 frame 19 frame 34 frame 37

Figure 4: (a) Top row contains frames rendered positioning the object *LuxoSr* according to ground truth pose information. (b) Bottom row contains frames rendered positioning the object *LuxoSr* according to pose estimates obtained by our pose estimation algorithm.

To analyze the efficiency of the proposed method, we plot the generalized distance D obtained after convergence for each of the frames in the *LuxoSr* sequence. This plot is presented in figure 3.

To show how precise these results are, we construct two distinct video sequences using the object model constructed from observations and two distinct motion scripts. The first motion script is composed by the available ground truth pose information, while the second motion script is composed by pose estimates obtained using the method we propose.

Snapshots of the rendered sequences are presented in figure 4. Figure 4a contains selected frames extracted from the video sequence rendered using ground truth pose information and figure 4b contains the same six snapshots rendered using pose estimates computed using the proposed algorithm. The similarity between the of snapshots, illustrates the precision of the proposed registration method. According to figure 3, the worst pose estimation result in this experiment occurs for frame *19* . Comparing the snapshots for frame *19* in figure 4, we observe that the result obtained by the proposed registration method is very precise.

Additional experiments with the same object performing different motion scripts, including translation only and simultaneous translation and rotation generate estimates with similar accuracy.

The second experiment uses *Cylinder*, a synthetic sequence of a marble textured cylinder with semi-spherical caps rotating around the vertical axis. Snapshots of the object moving according to the ground truth pose information are presented in figure 6a.

The goal of this experiment is to evaluate the usefulness of texture in pose estimation. The *Cylinder* sequence was chosen because, for the described motion pattern, the shape of the object is a highly ambiguous cue for pose estimation. The ambiguity is such that pose estimation using the original ICP algorithm leads to results that are not reliable, as presented in figure 6c.

The quantitative contribution of texture information to the efficiency of the pose estimation method is presented in figure 5, where the generalized distance D upon convergence is plotted for each of the frames in the *Cylinder* sequence. The experiment is repeated for several values of the parameter λ, that controls the importance of texture in the pose estimation criterion. We observe that the average error in pose estimation decreases as we increase the importance of texture in the discrepancy metric by increasing the value of λ in equation 1.

For a qualitative evaluation of the results, snapshots of a sequence rendered using the pose estimates obtained using our proposed method are presented in figure 6b. A comparison with the ground truth snapshots from figure 6a, illustrates the precision of the proposed technique. Comparing the results obtained using our method presented in figure 6b and the outcome of the ICP algorithm presented in figure figure 6c, we observe that our method outperforms ICP, that only relies on shape information.

The final experiment uses *Mug*, a real data sequence of a mug rotating around its principal axis. The ambiguity of the shape is not as extreme as in the second experiment, and texture is not as uniform as in the first experiment. The light stripe range finder generates a depth map and a co-registered intensity image sequence. The output of the range finder is corrupted by noise and the mug handle generates several self-occlusion episodes.

Selected frames from the original video sequence are presented in figure 7, while figure 8 presents the same set of frames reconstructed using the object model constructed from observations and the pose estimates obtained using our proposed method. A comparison between the object position in corresponding snapshots reveals the precision of the proposed method.

6 CONCLUSIONS AND FUTURE WORK

A novel 3D-3D registration method that uses shape and texture has been proposed. The efficiency and robustness of the method have been confirmed via experiments with synthetic and real data. Significative performance enhancement is observed with respect to the original ICP algorithm.

To further improve on pose estimation, we currently investigate the use of stochastic occupancy distributions and sensor model knowledge to enhance precision and reduce processing time.

7 REFERENCES

[1] B. K. P. Horn, "Closed form solution of absolute orientations using unit quaternions," *Journal of the Optical Society of America*, vol. 4, no. 4, pp. 629–642, April 1987.

[2] O. D. Faugeras and M. Hebert, "The representation, recognition, and locating of 3D objects," *International*

Journal of Robotic Research, vol. 5, no. 3, pp. 27–52, Fall 1986.

[3] P. J. Besl and H. D. Mckay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–56, February 1992.

[4] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image and Vision Computing (UK)*, vol. 10, no. 3, pp. 145–55, April 1992.

[5] G. Champleboux, S. Lavallee, R.Szeliski, and L. Brunie, "From accurate range imaging sensor callibration to accurate model-based 3D object localization," in *Proceedings CVPR: Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 1986, pp. 435–7, Computer Society Press.

[6] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *International Journal of Computer Vision*, vol. 13, no. 2, pp. 119–52, October 1994.

[7] D. Simon, M. Hebert, and T. Kanade, "Real time 3D pose estimation using a high-speed range sensor," Tech. Rep. CMU-RI-TR-93-24, Robotics Institute - Carnegie Mellon University, November 1993.

[8] J. Feldmar, N. Ayache, and F. Betting, "3D-2D projective registration of free-form curves and surfaces," in *Proceedings of IEEE International Conference on Computer Vision*, Cambridge, MA, June 1995.

[9] B. Sabata and J. K. Aggarwal, "Estimation of motion from a pair of range images: a review," *CVGIP: Image Understanding*, vol. 54, no. 3, pp. 309–24, November 1991.

[10] L. G. Brown, "A survey of image registration techniques," *Computing Surveys*, vol. 24, no. 4, pp. 325–76, December 1992.

[11] F. Solina and R. Bajcsy, "Recovery of parametric models from range images: the case for superquadrics with global deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 2, pp. 131–47, February 1990.

[12] G. Turk and M. Levoy, "Zippered polygon meshes from range images," in *Proceedings of SIGGRAPH 21th Annual International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'94)*, July 1994, pp. 311–18.

[13] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, "Surface reconstruction from unorganized points," in *Proceedings of SIGGRAPH 19th Annual International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'92)*, 1992, pp. 71–77.

[14] M. Soucy and D. Laurendeau, "A general surface approach to the integration of a set of range views," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 4, pp. 344–58, April 1995.

[15] A. Elfes, "Occupancy grids: A stochastic spatial representation for active robot perception," in *Proceedings of the Sixth Conference on Uncertainty and AI*, Cambridge, MA, July 1990, AAAI.

[16] F. C. M. Martins and J. M. F. Moura, "3-D Video Compositing: towards a compact representation for video sequences," in *Proceedings of IEEE International Conference on Image Processing*, Washington, DC, October 1995.

Figure 5: Error in pose estimation for all frames of *Cylinder* for distinct values of λ. The average generalized distance decreases as λ increases the importance of texture in the discrepancy criteria.

Figure 6: (a) The top row presents selected frames rendered positioning object *Cylinder* using the ground truth pose reference. (b) The middle row contains frames rendered positioning the object *Cylinder* according to pose estimates computed using our proposed method. (c) The lower row has frames rendered positioning the object *Cylinder* according to pose estimates computed using the ICP algorithm.

Figure 7: Selected frames from original video sequence *Mug*.

Figure 8: Frames rendered using object models and pose estimates obtained by applying the proposed method to the video sequence *Mug*.