

## BUILDING INVENTORY INFORMATION EXTRACTION FROM REMOTE SENSING DATA AND STATISTICAL MODELS

Pooya Sarabandi<sup>1,2</sup> and Anne S. Kiremidjian<sup>2</sup>

<sup>1</sup>Risk Management Solutions, Inc., Newark, CA, USA <sup>2</sup>Department of Civil and Environmental Engineering, Stanford University, CA, USA Email: <u>psarabandi@stanford.edu</u>, <u>ask@stanford.edu</u>

## **ABSTRACT:**

In this paper, algorithms for extracting building attribute information from remotely sensed data are presented. In particular, a methodology for rapidly extracting spatial and structural information from a single highresolution satellite image, using rational polynomial coefficients (RPCs) as a camera replacement model is introduced. Geometric information defining satellite's sensor orientation is used in conjunction with the RPC projection model to generate an accurate digital elevation model (DEM). Additionally, a methodology for inferring engineering attributes of the built-environment, i.e. structural type and occupancy type of buildings, from 3-D building models is formulated. A dataset collected for Southern California, USA, is used to train multinomial logistic regression models and establish inference rules in order to predict the regional engineering parameters of the buildings. Classification error and prediction power of these models are then presented in the paper and an example of the marginal probability distribution computed for a sample building is shown.

**KEYWORDS:** Building Inventory, Remote Sensing, Rational Function Model, Height Extraction, 3-D Building Modeling, Statistical Modeling, Inference Rules, Multinomial Logistic Regression.

#### **1. INTRODUCTION**

When risk analysis is performed on a large region, one of the most daunting tasks is the compilation of regional building inventories. These inventories need to contain information on various building attributes such as geographic location (longitude and latitude), height or number of stories, footprint area, total square footage, structural class (construction type), usage or occupancy of buildings, age or year of construction, plan and elevation irregularities (re-entrant corners, setbacks and etc.), cladding type, roof shape, roof type and etc. Several of these attributes can directly be extracted from remotely sensed data -such as the data collected by aerial photography or satellite imagery. Height, footprint and other geometric features of buildings are among this group of attributes. There is however another set of attributes -that cannot directly be derived from remotely sensed data- which can be inferred using engineering or statistical rules. Structural type, occupancy type and age of structures are among the attributes of this group.

This paper introduces a methodological approach to use a single satellite image and dynamic measurement to generate an accurate digital elevation model (DEM). In the dynamic measurement mode, a pair of pixels - in the image- is selected such that they represent corners of a building at ground level and their corresponding rooftop points. The proposed methodology uses Rational Function Model (RFM) as a replacement for the rigorous (physical) satellite sensor model to establish a transformation between 3-dimensional object-coordinates of buildings on the ground (latitude, longitude and height) and their corresponding 2-dimensinal image coordinates (row and column). In addition to extracting geometric attributes of the built-environment from satellite images, a methodology for inferring structural type and occupancy type of buildings from other signatures and attributes of an urban area such as the ones that can be derived from imagery is then presented in this paper. Since the response variables to be modeled (outputs of the model), i.e. structural type and occupancy type, as well as some of the independent variables (inputs to the model) such irregularity of buildings or roof type of structures are categorical data, the statistical model to be used should incorporate both categorical and quantitative data. Therefore, "multinomial logistic regression model" is chosen in order to establish the inference rules.



## 2. GEOMETRIC INFORMATION EXTRACTION FROM SATELLITE IMAGES

This section introduces a methodological approach to use a single satellite image and dynamic measurement to generate an accurate digital elevation model (DEM). In dynamic measurement mode, a pair of pixels - on the image plane- is selected such that they represent the corner of a building at ground level and its corresponding rooftop point.

## 2.1. Fundamentals

Recent advances in high-resolution satellite imaging are extending application of commercial satellite imaging such as those acquired by IKONOS, QuickBird, OrbView and SPOT5 - to accurate 3-D building modeling and geospatial information extraction. To support real-time calculations and provide an easy-to-use sensor model, many commercial high-resolution satellite image providers use Rational Function Model (RFM) as a replacement for their rigorous (physical) sensor model. RFM is a generalization of polynomial models that can be used to describe the image-to-ground relationship. RFM uses ratio of two polynomial functions to define the transformation between 3-dimensional coordinates of an object on the ground (latitude, longitude and height) and its corresponding 2-dimensinal image coordinates (row and column). For a given image, RFM can be expressed as:

$$r_{n} = \frac{f_{1}(\phi_{n}, \lambda_{n}, h_{n})}{f_{2}(\phi_{n}, \lambda_{n}, h_{n})} , \ c_{n} = \frac{f_{3}(\phi_{n}, \lambda_{n}, h_{n})}{f_{4}(\phi_{n}, \lambda_{n}, h_{n})}$$
(1)

where  $r_n$  and  $c_n$  are the row and column indices of pixels in the image, respectively;  $\phi_n$ ,  $\lambda_n$  and  $h_n$  are geodetic latitude, geodetic longitude and height above the ellipsoid, respectively.

Polynomials 
$$f_i$$
 (*i*=1,2,3,4) in Equation 1, have the general form of:  $f = \sum_{i=0}^{3} \sum_{j=0}^{3} \sum_{k=0}^{3} a_{ijk} \phi^i \lambda^j h^k$  (2)

The ratio of first-order terms in RFM usually compensates for distortions caused by optical projection, secondorder terms can be used to correct for earth curvature, atmospheric refraction and lens distortion while thirdorder terms can model other sources of noise and distortions [1]. To determine longitude, latitude and height of a structure, one needs to measure image coordinates of corner of a building at ground level and its corresponding rooftop coordinates. For each conjugate pair obtained in this dynamic measurement, the following set of equations can be obtained:

$$r_{ground} = \frac{f_1(\phi, \lambda, h_1)}{f_2(\phi, \lambda, h_1)} , c_{ground} = \frac{f_3(\phi, \lambda, h_1)}{f_4(\phi, \lambda, h_1)} , r_{roof} = \frac{f_1(\phi, \lambda, h_2)}{f_2(\phi, \lambda, h_2)} , c_{roof} = \frac{f_3(\phi, \lambda, h_2)}{f_4(\phi, \lambda, h_2)}$$
(3)

where  $r_{ground}$ ,  $c_{ground}$ ,  $r_{roof}$  and  $c_{roof}$  are the measured (normalized) image coordinates of a ground-point and its corresponding rooftop-point on the image (conjugate pair), respectively;  $\phi$ ,  $\lambda$ ,  $h_1$  and  $h_2$  are the unknown (normalized) object space coordinates.

#### 2.2. Image Acquisition Geometry and Height Metrology

Approximate image acquisition geometry and satellite orientation can be described by sensor's elevation and azimuth angles. Sensor's elevation angle is the angle form the horizon up to the satellite [2]. The projection of sensor's line of sight to the area-of-interest (AOI) onto the horizontal plane measured clockwise defines the sensor's azimuth [2] as shown in Figure 1.

By knowing a sensor's collection azimuth ( $\beta$ ) and measuring the image coordinates for the corner of a building at ground level, ( $r_{ground}$ ,  $c_{ground}$ ), and its corresponding rooftop-point coordinates, ( $r_{roof}$ ,  $c_{roof}$ ), it is possible to calculate height of a building through trigonometric relationship as described in Equation 4 and shown in Figure 2.

$$H = \frac{H^*}{\cos(\beta\beta)}, \text{ where } H^* = GSD \times \sqrt{(r_{ground} - r_{roof})^2 + (c_{ground} - c_{roof})^2}$$
(4)

where *GSD* is ground sample distance at the viewing angle  $\beta$ . *H* is the physical height of a building and *H*<sup>\*</sup> is the measured height of a building on the image plane.





Figure 1. Image Acquisition Geometry

Figure 2. Relationship between real height of a building And the measured height on the image plane

## 2.3. 3-D Reconstruction Algorithm

A system of homogeneous nonlinear over-determined equations can be obtained by adding the geometric constraint for height, derived in Equation 4, to the set of equations introduced by Equation 3. The unknown variables are geodetic longitude, geodetic latitude, height of ground-point above the geoid and height of rooftop-point above the geoid,  $(\phi, \lambda, h_1, h_2)$ , as presented in Equation 5:

$$\begin{cases} \frac{f_{1}(\phi,\lambda,h_{1})}{f_{2}(\phi,\lambda,h_{1})} - r_{g} = 0 \\ \frac{f_{3}(\phi,\lambda,h_{1})}{f_{4}(\phi,\lambda,h_{1})} - c_{g} = 0 \\ \frac{f_{1}(\phi,\lambda,h_{2})}{f_{2}(\phi,\lambda,h_{2})} - r_{r} = 0 \\ \frac{f_{3}(\phi,\lambda,h_{2})}{f_{4}(\phi,\lambda,h_{2})} - c_{r} = 0 \\ \frac{f_{3}(\phi,\lambda,h_{2})}{f_{4}(\phi,\lambda,h_{2})} - c_{r} = 0 \\ (h_{2} - h_{1}) - H = 0 \end{cases}$$
(5)

The above system of nonlinear equations can be solved using the Trust-Region Dogleg Method [3] & [4]. To do so, a linear system of equations is solved to find the search direction, and trust-region techniques [5] are used to improve the robustness of the algorithm when the starting point is far from the solution or in cases where Jacobian of design matrix (Equation 5) is singular. The starting point  $\mathbf{x}_o = (\phi^*, \lambda^*, h_1^*, h_2^*)$ , used in the iterative solution of Equation 5, can be obtained by linearizing Equation 3 considering only the first-order terms in the numerator and denominator as expressed in Equations 6 and 7. Singular Value Decomposition (SVD) of matrix *A* can be used if *A* becomes singular.

$$r_{g} = \frac{a_{1} + a_{2}\lambda + a_{3}\phi + a_{4}h_{1}}{b_{1} + b_{2}\lambda + b_{3}\phi + b_{4}h_{1}} + \varepsilon_{1}, \quad c_{g} = \frac{c_{1} + c_{2}\lambda + c_{3}\phi + c_{4}h_{1}}{d_{1} + d_{2}\lambda + d_{3}\phi + d_{4}h_{1}} + \varepsilon_{2}$$

$$r_{r} = \frac{a_{1} + a_{2}\lambda + a_{3}\phi + a_{4}h_{2}}{b_{r} + b_{r}\lambda + b_{r}\phi + b_{r}h_{r}} + \varepsilon_{3}, \quad c_{r} = \frac{c_{1} + c_{2}\lambda + c_{3}\phi + c_{4}h_{2}}{d_{r} + d_{r}\lambda + d_{r}\phi + d_{r}h_{r}} + \varepsilon_{4}$$
(6)

$$A \cdot \mathbf{x}_{o} = b \quad \text{where} \quad A = \begin{bmatrix} r_{g}b_{2} - a_{2} & r_{g}b_{3} - a_{3} & r_{g}b_{4} - a_{4} & 0 \\ c_{g}d_{2} - d_{2} & c_{g}d_{3} - d_{3} & c_{g}d_{4} - d_{4} & 0 \\ r_{r}b_{2} - a_{2} & r_{r}b_{3} - a_{3} & 0 & r_{r}b_{4} - a_{4} \\ c_{r}d_{2} - d_{2} & c_{r}d_{3} - d_{3} & 0 & c_{r}d_{4} - d_{4} \end{bmatrix}, \quad \mathbf{x}_{o} = \begin{bmatrix} \phi^{*} \\ \lambda^{*} \\ h_{1}^{*} \\ h_{2}^{*} \end{bmatrix}, \quad b = \begin{bmatrix} a_{1} - r_{g}b_{1} \\ c_{1} - c_{g}d_{1} \\ a_{1} - r_{r}b_{1} \\ c_{1} - c_{r}d_{1} \end{bmatrix}$$
(7)



## 2.4. Measurement Error

The amount of error introduced by an operator in the process of selecting ground-points and rooftop-points affects the accuracy of differential-height estimation in Equation 4 and, therefore, the overall accuracy in determining longitude, latitude and height. In this section, only the direct effect of selection-error on height is addressed. Figure 3 shows the relationship between rooftop selection-error and its corresponding error in height. Considering two independent measurements in determining the location of ground-point and rooftop-point, the differential-height error can be calculated as shown below:



Figure 3. Pixel selection and its corresponding height estimation error

# 3. APPLICATION OF STATISTICAL INFERENCING TECHNIQUES TO BUILDING INVENTORY COMPILATION

In order to infer the structural- and occupancy-type of buildings from geometric and spatial attributes of the built-environment, a statistical framework which incorporates both quantitative and qualitative variables should be utilized. In this section, application of *multinomial logistic regression models* in inferring categorical attributes of urban areas is investigated. Models developed in this section are then used to establish a set of inference rules using training datasets.

#### 3.1. Fundamentals

*Multinomial logistic regression models*, use multinomial probability distribution to model probability of a response variable, from the  $i^{th}$  observation, falling into the  $k^{th}$  category given a set of explanatory variables as shown below:

$$P(y_i = k \mid \mathbf{x}_i) = \pi_k(\mathbf{x}_i) = \pi_{ik}$$
  
 $i = 1: N \text{ and } k = 1: K - 1$ 
(9)

where  $(\mathbf{x}_i, y_i)$  is the *i*<sup>th</sup> observation such that  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$  is a vector of *p* explanatory variables and  $y_i$  is the corresponding response variable.  $\pi_{ik}$  is the probability of *i*<sup>th</sup> response variable falling in the *k*<sup>th</sup> category.

It can be seen that for the *i*<sup>th</sup> observation, the response variable with *K* categories can be treated as a multinomial variable with probabilities  $\{\pi_{i1}, \pi_{i2}, ..., \pi_{iK}\}$  and the constraint that  $\sum_{k=1}^{K} \pi_{ik} = 1$ . To impose the constraint that fitted probabilities on the *K* categories should sum to one, one of the categories should arbitrary be selected as the *base category* or the *control group*. This category can be the first, the last or any other. Choosing the last category as the baseline category, the *log-odds* or "logarithm of the ratio between *logit* model of the *k*<sup>th</sup> category and the baseline category" for *p* explanatory variables can be expressed as shown in Equation 10.

$$\ln(\frac{\pi_{ik}}{\pi_{iK}}) = \sum_{j} \beta_{jk} x_{ij} = \alpha_k + \beta_{1k} x_{i1} + \dots + \beta_{pk} x_{ip}$$
(10)



(12)

where  $\alpha_k$  and  $\beta_{jk}$ 's are the logistic regression coefficients of the log-odds of the  $k^{th}$  category relative to the base category.

Using Equation 10, probability of  $i^{th}$  observation falling in the  $k^{th}$  category can then be expressed as below:

$$\pi_{ik} = \frac{\exp(\sum_{j} \beta_{jk} x_{ij})}{1 + \sum_{k=1}^{K-1} \exp(\sum_{i} \beta_{jk} x_{ij})} ; i = 1:N, j = i:p$$
(11)

and therefore  $\pi_{iK} = 1 - \sum_{k=1}^{K-1} \pi_{ik}$ 

where  $\pi_{ik}$  is the probability of  $i^{th}$  observation falling in the  $k^{th}$  category.

Parameter estimation for log-odds of explanatory variables is done by maximizing the expectation of loglikelihood of each variable. For each log-odds, estimated parameters ( $\hat{\beta}$ ) as well as their standard error ( $\sigma_{\hat{\beta}}$ ) can be calculated. The significance of each parameter in the model can be assessed using the *Wlad statistics*. The *Z*-value of Wald statistics for each parameter can be calculated by computing the ratio of estimated parameters and their standard error term as shown in Equation 13.

$$Z = \frac{\hat{\beta}}{\sigma_{\hat{\beta}}} \sim N(0,1) \tag{13}$$

where Z is the Wald statistics of the estimated parameter  $\hat{\beta}$ . The standard error of parameter  $\hat{\beta}$  is shown by  $\sigma_{\hat{\beta}}$ .

The computed *Z*-value has a normal distribution and can be used to judge the significance of the coefficient. It can be shown that for large sample sizes,  $Z^2$  has a chi-square distribution with one degree of freedom. To judge the overall suitability and parsimony of a model, the *Akaike Information Criterion (AIC)* is used. In the context of logistic models presented in this paper, the *AIC* can be defined as the sum of residual deviance of the model and the number of regression coefficients as shown in Equation 14.

$$AIC = -2L(\hat{\boldsymbol{\beta}}) + 2n \tag{14}$$

where  $L(\hat{\beta})$  is the maximum log-likelihood of the fitted model and *n* denotes total number of variables in the model.

Smaller values of *AIC* indicate a better fit to the data. *AIC* can also be used as a comparison tool when it comes to model selection (as shown later in this paper in Table 2).

#### 3.2. Datasets and Model Development

Detailed inventory data of eighteen census tracts within the Orange County (Southern California, USA) is used as the training dataset in this paper. The inventory database of Orange County is collected at the building level with attributes extracted from remotely sensed data -as discussed in section 2- as well as some ancillary attributes extracted from tax assessor database of the county. In this dataset, height, square footage, configuration irregularities as well as rooftop types of 1947 buildings are extracted from QuickBird imagery of the area, while structural type, occupancy type and year construction of buildings are extracted from tax assessor databases.

In order to identify all the possible models that can be created using this dataset, the most primitive combination of attributes, also known as the *baseline model*, should first be identified. Additional explanatory variables then will be added to the baseline model one at a time. The most basic model to be considered is the one associated with four explanatory variables; i.e. the height, square footage, irregularity and rooftop type of buildings. This choice of variables is mainly because of the fact that height, footprint area of structures, irregularity and rooftop type are the ones which can directly be extracted from remotely sensed data when one constructs a 3-D city model of the area. Depending upon availability of ancillary data, year of construction (age of buildings),



occupancy type or structural type can be included in the model in the later steps. Table 1 summarizes different models which can be created using the training dataset.

Model	Response	Explanatory Variable Included in the Model							
ID	Variable	Variable #1	Variable #2	Variable #3	Variable #4	Variable#5	Variable#6		
Model I*	Str. Type	Height (ft)	Area (ft <sup>2</sup> )	Configuration	Roof	-	-		
Model II	Str. Type	Height (ft)	Area (ft <sup>2</sup> )	Configuration	Roof	Occ. Type	-		
Model III	Str. Type	Height (ft)	Area $(ft^2)$	Configuration	Roof	Occ. Type	Age		
Model IV*	Occ. Type	Height (ft)	Area (ft <sup>2</sup> )	Configuration	Roof	_	-		
Model V	Occ. Type	Height (ft)	Area (ft <sup>2</sup> )	Configuration	Roof	Str. Type	-		
Model VI	Occ. Type	Height (ft)	Area $(ft^2)$	Configuration	Roof	Str. Type	Age		

 Table 1 Summary of models created using the Southern CA training dataset

\* Baseline Model

## 3.3. Measurement Error

The training dataset is used to compute the parameters of multinomial logistic models, fitted to each set of variables defined in Table 1. In order to calculate the overall classification error of a multinomial logistic model, the corresponding classification table for the response variable using prediction rules defined by the model should be calculated. The diagonal elements of this table represent the number of correctly classified observations. Classification error can then be calculated by computing the ratio between sum of the diagonal elements of the table and total number of elements in the table as shown in Equation 15.

$$\varepsilon = 1 - \frac{sum[diag(T)]}{sum[T]}$$
(15)

where  $\varepsilon$  is the classification error or misclassification rate, in table *T* and *diag(.)* refers to the diagonal elements of the table.

Table 2 summarizes the *AIC*, the degrees of freedom *df*, for estimating parameters of each model and the overall classification error of each of the models, respectively.

Model	Response	Explanatory Variable					AIC	df	Classification	
ID	Variable	#1	#2	#3	#4	#5	#6			Error
*	~ -			~ ~						
Model I	Str. Type	Н	Α	Config.	Roof	-	-	2663	35	21.98%
Model II	Str. Type	Н	Α	Config.	Roof	Occ.	-	1827	50	17.36%
Model III	Str. Type	Н	А	Config.	Roof	Occ.	Age	1771	55	16.18%
Model IV*	Occ. Type	Н	А	Config.	Roof	-	-	2984	21	29.74%
Model V	Occ. Type	Н	А	Config.	Roof	Str.	-	2062	36	18.90%
Model VI	Occ. Type	Н	А	Config.	Roof	Str.	Age	2064	39	18.90%

Table 2 Summary of AIC, degrees of freedom (df) and the overall classification error of models used in building the multinomial logistic regression models from the training dataset

\* Baseline model

#### 3.4. Results and Examples

The computed multinomial logistic model in each case (in Table 2) can be used for classification purposes by defining a set of decision rules. For instance, for an input *attribute vector* -consisted of *p* independent values- to the model, i.e.  $\mathbf{x} = (x_1, x_2, ..., x_p)$ , probability of the response falling into each of the *K* categories can be computed as  $\boldsymbol{\Pi} = \{\pi_1, \pi_2, ..., \pi_K\}$  using Equations11 and 12. The category corresponding to the highest probability in  $\boldsymbol{\Pi}$  can be selected as the class to which the input attribute vector belongs. Furthermore, a minimum probability threshold can be chosen such that if the highest probability in  $\boldsymbol{\Pi}$  is below that threshold,



the classification results in an "unclassified" status. In cases in which the probability difference between two classes is not significant, a tie assignment between class-membership will result and therefore, a set of rules should be defined in order to assign the response variable to a desirable class. It should be noted that in many cases decision rules depend on the nature of the input variables to the model as well the resulted response variable, and hence they differ from one problem to the other. Therefore, careful consideration should be given while compiling the decision rules for a specific problem.

Table 3 shows the result of parameter estimation for a model which uses height classes (low: 1-3 stories, medium: 4-7 stories and high: 15+ stories) and total square footage of a building to infer the construction type. Construction classes to be inferred from the mode are: Concrete (C), Steel (S), Concrete/Steel (C/S), Reinforced Masonry (RM), Un-Reinforced Masonry (URM) and Wood (W). The *base-category* (also known as *control group*) for explanatory variable in this table is the structural class type "*C*", i.e. the concrete class.

Tuble 5 Log odds parameters of model 1 (nom Tuble 1)							
Log-odds		Intercept	$\mathrm{H}^{*}$	Н	Н	Average	
			(High)	(Medium)	(Low)	Area	
C/S	Coefficients	-2.216	-	-0.324	-0.572	-1.3E-05	
	Std. Error	5.2E-11	-	5.3E-12	4.3E-11	2.2E-06	
RM	Coefficients	-0.574	-	0.337	-0.355	-1.2E-05	
	Std. Error	1.7E-11	-	2.7E-12	1.3E-11	8.3E-07	
S	Coefficients	-1.154	-	-0.130	-1.502	8.4E-06	
	Std. Error	5.3E-12	-	2.2E-12	2.6E-12	4.1E-07	
URM	Coefficients	-0.401	-	0.382	-0.388	-1.7E-05	
	Std. Error	2.5E-11	-	3.5E-12	2.0E-11	9.7E-07	
W	Coefficients	-8.096	-	11.408	9.099	-4.9E-05	
	Std. Error	3.6E-11	-	5.9E-12	3.1E-11	8.7E-07	
* _							

Table 3 Log-odds parameters of *Model I* (from Table 1)

Base-category Variable

In order to assign a class to an independent observation, the probability vector  $\Pi$  should first be computed. The category corresponding to the highest probability in  $\Pi$  is then assigned to the observation. As an example let's assume the structural type of a low-rise building with average square footage of 2,176  $ft^2$  in Southern California is to be predicted. Using the estimated parameters for the corresponding model, shown in Table 3, and using structural class *C*, i.e. concrete, as the base-category for the response variable, the log-odds ratios can be calculated as shown in Equation 16.

$$\begin{cases} \ln(\frac{\pi_{C/S}}{\pi_{C}}) = -2.216 - 0.324x_{1} - 0.572x_{2} - 0.000013x_{3} \\ \ln(\frac{\pi_{RM}}{\pi_{C}}) = -0.574 + 0.337x_{1} - 0.355x_{2} - 0.000012x_{3} \\ \ln(\frac{\pi_{S}}{\pi_{C}}) = -1.154 - 0.130x_{1} - 1.502x_{2} + 0.0000084x_{3} \\ \ln(\frac{\pi_{URM}}{\pi_{C}}) = -0.401 + 0.382x_{1} - 0.388x_{2} - 0.000017x_{3} \\ \ln(\frac{\pi_{W}}{\pi_{C}}) = -8.906 + 11408x_{1} + 9.099x_{2} - 0.000049x_{3} \end{cases}$$
(16)

where  $x_1$  and  $x_2$  are dichotomous dummy variables corresponding to height of the structure as defined in Table 4.  $x_3$ , is a quantitative variable corresponding to average square footage of the structure in  $ft^2$ .

Table 4	Dummy	variables $x_1$	and $x_2$ as	indicators of	f height in	Equation	16
						1	

		<u> </u>
Height (High-rise)	0	0
Height (Medium-rise)	1	0
Height (Low-rise)	0	1



Therefore, the marginal mass-probability density of the observation falling into different response categories can be computed, using  $x_1 = 0$ ,  $x_2 = 1$  and  $x_3 = 2176$ , as:

$$\Pi = \{\pi_c = 0.227, \ \pi_{c/s} = 0.014, \ \pi_{RM} = 0.087, \ \pi_s = 0.016, \ \pi_{URM} = 0.099, \ \pi_W = 0.556\}$$

This means that there is 55.6% chance that the construction class of the building-of-interest is wood; there is 22.7% chance that it is concrete; there is about 10% chance that the building is unreinforced masonry and etc. It can be seen that the last category, i.e. W, in the probability vector  $\Pi$  has the highest value (55.6%) and therefore, in absence of any other information such as age or occupancy type, the predicted structural type of a low-rise building in Southern California with an average square footage of 2,176  $ft^2$  is "wood frame". This class prediction is in agreement with realizations of the training dataset.

## 3. CONCLUSIONS

A cost effective methodology for extracting spatial information of the built-environment, i.e. height, footprint, shape configuration and etc., using single optical satellite images is presented in this paper. This information can be used to rapidly compile building inventory information for hard-to-reach and remote areas for which minimum or no data is available.

Furthermore, the multinomial logistic regression model presented in this paper provides the means to compute quantitative measures of marginal occupancy and building class-membership probabilities for categorical attributes associated with structures. This information can be used to augment probabilistic risk assessment models by providing a measure of uncertainty for construction and occupancy classes of the inventory-at-risk for a particular region, when the full distribution of construction and occupancy is not available.

## ACKNOWLEDGEMENT

This research was supported by the National Science Foundation (NSF) Grant EEC-9701568, the UPS Foundation Grant from Stanford University, the Multidisciplinary Center for Earthquake Engineering Research Grant EEC-9701471, and ImageCat Inc. The authors would like to gratefully acknowledge their supports.

The material presented in this paper is part of the doctoral dissertation of the first author [6]. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the authors, and do not necessarily reflect those of NSF, MCEER, Stanford University or ImageCat Inc.

## REFERENCES

- [1] Grodecki, J., Dial G. and Lutes J., (2004), "Mathematical Model for 3D Feature Extraction from Multiple Satellite Images Described by RPCs," *ASPRS conference proceedings*, Denver Colorado.
- [2] Grodecki, J., Dial G. and Lutes J., (2004), "Error Propagation in Block Adjustment of High-Resolution Satellite Images," *ASPRS conference proceedings*, May 2004, Anchorage, Alaska.
- [3] Powell, M.J.D., (1978), "A Fast Algorithm for Nonlinearly Constrained Optimization Calculations," Numerical Analysis, (G.A.Watson ed.), Lecture Notes in Mathematics, Springer Verlag, Vol. 630.
- [4] Powell, M.J.D., (1970), "A Fortran Subroutine for Solving Systems of Nonlinear Algebraic Equations," Numerical Methods for Nonlinear Algebraic Equations, (P. Rabinowitz, ed.), **Ch.7**.
- [5] Moré, J.J. and D.C. Sorensen, (1983), "Computing a Trust Region Step," *SIAM Journal on Scientific and Statistical Computing*, Vol. 3, pp 553-572.
- [6] Sarabandi P., (2007), "Development of Algorithms for Building Inventory Compilation through Remote Sensing and Statistical Inferencing," Ph.D. Dissertation, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA.