Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections

Andrey Dimitrov \textsuperscript{a,*}, Mani Golparvar-Fard \textsuperscript{b,c,1}

\textsuperscript{a} Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY, United States
\textsuperscript{b} Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, United States
\textsuperscript{c} Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, United States

**Abstract**

Automatically monitoring construction progress or generating Building Information Models using site images collections – beyond point cloud data – requires semantic information such as construction materials and inter-connectivity to be recognized for building elements. In the case of materials such information can only be derived from appearance-based data contained in 2D imagery. Currently, the state-of-the-art texture recognition algorithms which are often used for recognizing materials are very promising (reaching over 95\% average accuracy), yet they have mainly been tested in strictly controlled conditions and often do not perform well with images collected from construction sites (dropping to 70\% accuracy and lower). In addition, there is no benchmark that validates their performance under real-world construction site conditions. To overcome these limitations, we propose a new vision-based method for material classification from single images taken under unknown viewpoint and site illumination conditions. In the proposed algorithm, material appearance is modeled by a joint probability distribution of responses from a filter bank and principal Hue-Saturation-Value color values and classified using a multiple one-vs.-all kernel Support Vector Machine classifier. Classification performance is compared with the state-of-the-art algorithms both in computer vision and AEC communities. For experimental studies, a new database containing 20 typical construction materials with more than 150 images per category is assembled and used for validation. Overall, for material classification an average accuracy of 97.1\% for 200 × 200 pixel image patches are reported. In cases where image patches are smaller, our method can synthetically generate additional pixels and maintain a competitive accuracy to those reported above (90.8\% for 30 × 30 pixel patches). The results show the promise of the applicability of the proposed method and expose the limitations of the state-of-the-art classification algorithms under real world conditions. It further defines a new benchmark that could be used to measure the performance of future algorithms.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Material classification is an important part of any vision-based system for automated construction progress monitoring or generation of semantically-rich as-built 3D models. Beyond 3D geometrical information in form of point cloud models, such tasks require additional semantic information such as construction materials and interconnectivity to be recognized for building elements. In the case of materials such information can mainly be derived from appearance-based data contained in 2D imagery.

Current state-of-the-art methods for automatically monitoring construction progress or generating building information models (BIM) primarily focus on using laser scanners or image-based 3D reconstruction methods to generate 3D point cloud models \cite{1-13}. These methods can generate the required geometrical information for producing BIM \cite{13}. Nonetheless, for generating semantically-rich models, beyond geometrical information, material and spatial relationship/interconnectivity among elements (e.g., a beam is supported by a column) needs to be extracted from the collected data. In this paper, we focus primarily on the problem of automated material classification. Automated material classification (term used interchangeably with recognition in computer vision literature) not only helps with deriving appearance-based data required for construction progress monitoring purposes, but it also helps with segmentation of elements for
2. Automated construction progress monitoring and building information modeling

3D modeling of the as-built environment is used by the AEC industry in a variety of engineering analysis scenarios. Significant applications include progress monitoring of construction sites, quality control of fabrication and on-site assembly, energy performance assessment, and structural integrity evaluation. The modeling process mainly consists of three sequential steps [11,13]: data collection, modeling, and analysis. In current practice, these steps are performed manually by surveyors, designers, and engineers. Such manual tasks can be time-consuming, expensive, and prone to errors. While the analysis stage is fairly quick, taking several hours to complete, data collection and modeling can be the bottleneck of the process. The data collection can spread over a few days, nonetheless the modeling stage can span over multiple weeks or even months. Additionally, modeling often tends to be specific to certain analysis, making the application of one model to multiple analyses very challenging. Thus, the applicability of as-built modeling has been traditionally restricted to high latency analysis, where results need not be updated frequently. In fast changing environments such as construction sites, due to the difficulty in rapidly updating 3D models, model-based assessment methods for purposes such as progress or quality monitoring have had very limited applications. In consequence, there is a need for a fast, low-cost, reliable, and automated method for as-built modeling. This method should quickly generate and update accurate and complete semantically-rich models in a master format that is translatable to any engineering scenario and can be widely applied across all construction projects [11,13]. The development of such a 3D modeling method also needs to be accompanied with a low-cost and fast data collection process that is able to generate accurate and complete data sets for modeling purposes.

Over the past decade, research on as-built modeling has primarily focused on data collection and modeling techniques that can generate 3D geometrical information. Cheok et al. [14] is among the early works that demonstrated how LADAR (Laser Distance and Ranging) can be used as an as-built 3D modeling tool for construction monitoring purposes. This led to more systematic studies of structured light data collection techniques [13,15,16]. Several research studies have focused on showing that Terrestrial Laser Scanning (TLS) techniques provide sufficient accuracy for use in construction dimensional surveying and Quality Assurance/Quality Control (QA/QC) [8,7,16–18]. Another line of research focuses on automated registration of 3D laser scanning point clouds with BIM (e.g., [6,13,19]) for the purpose of controlling the quality of the data collection. Turkan et al. [4,5] and Bosche et al. [7–9] have also focused on automated recognition of 3D CAD objects in site laser scans for project performance control purposes.

While collecting data for as-built modeling using laser scanners can generate very accurate and complete geometrical models, in most cases the approach suffers from the lack of a semantic understanding of the scene; in particular the ability to account for construction materials. Indeed, there is a need for distinguishing different materials (e.g., formwork vs. finished concrete or concrete masonry units vs. facade bricks) during the 3D modeling process. Nonetheless, the laser scanning point clouds in their XYZ form do not provide the data necessary for such analysis. The recent addition of mounted cameras to laser scanners has allowed RGB color information to be associated with each 3D point. Recent works by Kim et al. [20] and Son et al. [21] proposed using color to identify concrete elements and construction equipment. However, this approach has not yet been shown to be applicable for the simultaneous classification of multiple classes of construction materials.

Over the past few years, cheap and high-resolution digital cameras, extensive data storage capacities, and the availability of Internet connections on construction sites have enabled capturing and transmitting information about construction performance on a truly massive scale. Photographic data has become the favored documentation medium, as it contains rich information (geometry and appearance), and data is collected quickly and inexpensively. In the meantime, exciting research progress has been made on techniques that can assist with 3D point cloud modeling of construction sites or individual building elements using digital imagery. For example, Golparvar-Fard et al. [1,22] presented a new algorithm based on Structure-from-Motion (SfM) coupled with Multi-View Stereo (MVS) [23–25] and Voxel Coloring [26] for generating dense 3D point cloud models from unordered imagery. Fathi et al. [27–29] also presented a method for generating 3D point cloud models of infrastructure using video streams. Several research projects such as [30–35] have also focused on evaluating the accuracy of as-built 3D modeling from photos and laser scanners. In majority of the experiments conducted using laser scanners, compared to the image-based reconstruction methods, a higher density and completeness of the point cloud models has been observed. Despite the recent advancements, the generation of semantically-rich Solid Geometric Models (SGM) compatible with the Industry Foundation Classes (IFC) format still poses three main open research challenges:

1. **Segmentation**: robust techniques in recognition are required to segment a 3D point cloud dataset based on geometric and appearance information into distinct subsets;
2. **Object placement**: techniques in geometrical modeling and recognition are needed to populate the scene with distinct BIM objects based on the segmented subsets;
3. **Inter-object relationships**: techniques in recognition are required to identify the physical relationships between BIM objects in the scene.

Ongoing research on image-based 3D modeling techniques can benefit from appearance information during the production of semantically-rich SGMs and progress monitoring; in particular:

1. **For producing surface and solid geometry**: by segmenting point clouds based on appearance information so each subset can be separately used for extracting geometric information (see Fig. 1),
2. For associating semantic information to geometry: by recognizing construction materials to semantically label each generated 3D element (see Fig. 2),

3. For producing semantically-rich SGMs: by assembling adjacent 3D elements of the same material into complete objects, and

4. For progress monitoring: by recognizing construction materials to detect the correct stage of the construction operation (see Fig. 3).

3. Previous work

3.1. Material and texture recognition in computer vision

The need for extracting semantic information from a scene has yielded two main computer vision research thrusts: object recognition and material recognition. Although object recognition methods have made significant advances in recent years [36], they tend to rely on material-invariant features and overlook material specificity. Although the visual characteristics of an object is to some degree a function of its material category, different classes of objects can be made of the same material and a given class of objects can be made from different materials [36]. Particularly, in the context of construction projects, shape and material of objects tend to have a weak correlation. Similar shapes can be made up of different construction materials; e.g., a foundation element as well as formwork section can both be shaped as rectangular boxes. Also objects of a same material can have high shape variability; e.g., (1) a concrete column can be rectangular as well as circular; (2) a column, a beam, or a slab can all be made of concrete.

Material classification for automation in construction is closely related to, but different from material and texture classification in computer vision in interesting ways:

(1) In computer vision, most methods assume there is no strong prior available for material recognition, and thus these methods should be able to handle classification on surfaces with certain randomness in orientation and periodicity in both their texture and geometry. In contrast, for the purpose of progress monitoring in construction, model-based methods such as [1,4,5] can take advantage of an expectation of materials present in the scene;
Several works in computer vision such as [37] define texture in terms of dimensions such as periodicity, orientedness, and randomness. As addressed by Liu et al. [36], surfaces made of different materials can have similar texture patterns and thus existing computer vision methods for texture recognition such as [38,39] may not be ideal for material classification.

In the context of material classification, a variety of feature-based strategies have emerged and proven applicable to more loosely defined scenes. They are usually coupled with machine learning techniques for training and testing purposes. Popular features include filter responses [38,40,41] and image patches [42], and have been coupled with a Nearest Neighbor (NN) learning [39,43]. Although they have shown promising results reaching over 95% accuracy, yet they do not perform well with images collected from construction sites. The authors conducted exhaustive experiments (see Section 6.1) and in most cases accuracy dropped to 70% and lower using real-world construction imagery. Brostow et al. [44] have proposed to incorporate SM point cloud information to appearance features for recognition and segmentation, however only accuracies of below 70% have been reported from real-world data. Caputo et al. [45] have shown that Support Vector Machine (SVM) learning can improve on NN, yet its performance has not increased the flexibility of multi-modal image retrieval methods and can further improve detection of elements by differentiating them into nonlinear and linear materials of certain directionality (e.g., steel columns, concrete beams). Zhu and Brilakis [50] further improved the method for automated concrete detection in image data. These methods clearly established the applicability of image-based texture recognition for material classifications. Kim et al. [51] developed a scanning technique that utilizes laser scanner point cloud data to automatically identify the sizes of stone aggregates.

Nonetheless, to support automated generation of BIM and/or generate models for progress and quality monitoring purposes, the state-of-the-art approaches for material classification both in computer vision and AEC communities still need to be improved in the following ways:

1. There is a need for testing various types of features and machine learning classification methods to form a robust automated material classification technique that maintains a high level of accuracy in uncontrolled construction environments. In Section 6, the performance of the state of the art algorithms are compared for uncontrolled construction environments.
2. There is a need for systematic data collections and comprehensive datasets of material classification in uncontrolled construction environments that can be used for benchmarking material recognition algorithms.

In the following sections, a new method and a comprehensive dataset for material classification are introduced. The performance of our proposed method is also compared against the state-of-the-art methods in both computer vision and AEC communities.
4. Proposed material recognition approach

Our proposed image-based material recognition algorithm is based on a statistical distribution of filter responses over the images in form of primitives such as edges, spots, waves, and Hue-Saturation-Value (HSV) color values, and is divided into two stages: learning and classification. The statistical distribution of filter responses has been shown to be a good descriptor for texture recognition [39,45]. The main reason for choosing color values similar to others such as [45,47,48] as it is validated in Section 6.3 is to improve robustness of the proposed method for recognition of materials when images are taken far from objects and texture information is less apparent. In addition, the HSV color values are chosen to improve the robustness to changes in brightness which can be dominant in construction imagery. Fig. 4 shows how the brightness and color gradients of a concrete foundation wall can change when images are taken far from the elements.

The main technical contributions of this work are (1) a Bag of Words (BoW) pipeline for forming statistical distributions of both filter responses and HSV color values; (2) a multiple binary SVM classifier that can robustly learn and infer construction material categories; and (3) a construction material library and validation process similar to [39], first and second derivatives of Gaussian at 6 orientation and 3 scales (total of 36), 8 Laplacian of Gaussian filters, and 4 Gaussians. The scales of filters in this filter bank range between σ = 1 to 10. These filters are shown in Fig. 6.

Next, following a bag-of-words model, we cluster the filter responses to generate codebooks of material appearances using the k-means clustering method [52]. The cluster centers (codewords) will be Textons which are visual atoms of the visual perception of materials [38,53]. Thus, the texture of the materials will be defined by modeling the texton frequencies learned from the training images. For all images captured under various illumination and viewpoints for all different materials in the training dataset, the filter responses are concatenated. Next, a fixed number of texton cluster centers are computed using the k-means clustering algorithm where k is the number of visual words (the size is discussed in Section 6.1). After quantizing individual features into k visual words, where each feature is assigned a membership to the closest cluster, the distribution of visual words per image is calculated for all of the different features by assigning each pixel in the image to the nearest visual word index and forming the histogram over the frequency of the visual words.

The histogram of texton frequencies is then used to form the codebooks and the material models that are corresponding to our training images. In our work, HSV colors are also incorporated to leverage the color information for a robust material classification. It was hypothesized that since HSV instead of standard R, G, and B color channels can separately model the brightness, it can provide a higher degree of invariance to varying illumination conditions on the sites or the time images are captured. Similar to the texture recognition, for all images in the training datasets, the normalized HSV color values are aggregated and clustered into HSV textons using the (k-means clustering algorithm). Next, a histogram of color textons, i.e., the frequency with which each color texton occurs in the images, forms the codebook corresponding to each training image. The texture and color histograms are then concatenated to generate a unique codebook for each image. Fig. 6 shows the details of the training process.

4.1. Generating codebooks of material appearances

In the learning stage of our algorithm, training images are initially converted to gray-scale and their color intensities are normalized to have zero mean and unit standard deviations. This will make the distribution of the color values less dependent on the illumination which varies significantly on a construction site. Next, these images are convolved with a comprehensive filter bank [38] to generate filter responses using following equation:

\[ I(x, y) * f(x, y) = \sum_{k=0}^{N} \sum_{l=0}^{M} I(k, l)f(x-k, y-l) \]

where \( I(x, y) \) is the normalized gray-scale intensity at \( (x, y) \), \( f(x, y) \) represents intensity of filter \( f \) at \( (x, y) \), and \( (n_1, n_2) \) and \( (m_1, m_2) \) represent the range of the filter bank. The filter bank consists of 48 filters similar to [39], first and second derivatives of Gaussian at 6 orientation and 3 scales (total of 36), 8 Laplacian of Gaussian filters, and 4 Gaussians. The scales of filters in this filter bank range between \( \sigma = 1 \) to 10. These filters are shown in Fig. 6.

Next, following a bag-of-words model, we cluster the filter responses to generate codebooks of material appearances using the k-means clustering method [52]. The cluster centers (codewords) will be Textons which are visual atoms of the visual perception of materials [38,53]. Thus, the texture of the materials will be defined by modeling the texton frequencies learned from the training images. For all images captured under various illumination and viewpoints for all different materials in the training dataset, the filter responses are concatenated. Next, a fixed number of texton cluster centers are computed using the k-means clustering algorithm where k is the number of visual words (the size is discussed in Section 6.1). After quantizing individual features into k visual words, where each feature is assigned a membership to the closest cluster, the distribution of visual words per image is calculated for all of the different features by assigning each pixel in the image to the nearest visual word index and forming the histogram over the frequency of the visual words.

The histogram of texton frequencies is then used to form the codebooks and the material models that are corresponding to our training images. In our work, HSV colors are also incorporated to leverage the color information for a robust material classification. It was hypothesized that since HSV instead of standard R, G, and B color channels can separately model the brightness, it can provide a higher degree of invariance to varying illumination conditions on the sites or the time images are captured. Similar to the texture recognition, for all images in the training datasets, the normalized HSV color values are aggregated and clustered into HSV textons using the (k-means clustering algorithm). Next, a histogram of color textons, i.e., the frequency with which each color texton occurs in the images, forms the codebook corresponding to each training image. The texture and color histograms are then concatenated to generate a unique codebook for each image. Fig. 6 shows the details of the training process.

4.2. Discriminative machine learning models of material categories

Our visual representation based on codebooks for construction materials is leveraged by training discriminative machine learning models of material categories. We train binary C-Support Vector Machine (SVM) classifiers [54] with linear, \( \chi^2 \), and radial basis function (RBF) kernels for each material category independently. SVM classifiers are discriminative binary classifiers that optimize a decision boundary between two classes. The SVM classification solves the following optimization problem:

\[
\min_w \frac{1}{2} ||w||^2 + C \sum_{i} \zeta_i
\]

subject to \( y_i(w^T x_i + b) \geq 1 - \zeta_i, \quad i = 1, 2, \ldots, m \) (3)

\[ \zeta_i \geq 0, \quad i = 1, 2, \ldots, m \] (4)

Here, \( w \) is the normal vector to the hyperplane that separates the images of a material class from others, \( b \) is the offset of the hyperplane from the origin, \( \zeta_i \) is the slack variables for measuring the degree of misclassification of the \( m \)-dimensional histogram data \( x \), \( C > 0 \) is the tradeoff between regularization and constraint violation, and finally \( y_i \) is the outcome of the classification. After solving, the SVM classifier predicts 1 if \( w^T x + b > 0 \) and −1 otherwise. The decision boundary is given by the line \( w^T x + b = 0 \) that separates
Fig. 5. An overview of the proposed learning and inference model for material recognition.

Fig. 6. The training process in our proposed material recognition method.
the images from a material class from others in a $m$-dimensional space ($m$: the size of Texton and HSV cluster centers).

Note that the above definition applies to a linear classifier. We also use non-linear kernels due to their suitability for classification of histograms. As validated in Section 6.1, the histograms cannot be separated linearly in a reasonable way, and thus $\chi^2$ (5) and RBF (6) kernels are used to map the data into a high dimensional feature space.

\[ k(x_i, x_j) = 1 - \frac{1}{n} \sum_{i=1}^{n} (x_i - x_j)^2 \]  
\[ k(x_i, x_j) = e^{-\gamma|x-x|^2} \]  

where $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$ denotes the inner product in the Hilbert space and corresponds to applying the algorithm to the mapped material histogram data point $\phi(x_i)$. In this dual formulation, we maximize the following:

\[ W(x) = \sum_{i=1}^{m} x_i - \frac{1}{2} \sum_{y} y_i x_i y_i k(x_i, x_j) \]  
\[ \text{subject to } 0 \leq x_i \leq C \text{ and } \sum x_i y_i = 0 \]

In this formulation, the decision function will be $\text{sign}(h(x))$, where:

\[ h(x) = \sum_{i=1}^{q} y_i k(x, x_i) + b \]  

where $q$ refers to the number of kernel computations needed to classify a joint texton and color material histogram with the kernelized SVM. For solving these equations, we use LibSVM [55] and $\chi^2$ kernel module of [56].

In order to extend the binary classification decision of each SVM classifier to multiple classes, we adopt the one-versus-all multi-class classification scheme where a classifier for each material is defined. When training the SVM classifier that corresponds to each material class, we set all the examples from that class as positive and the examples from all other classes as negatives. The result of the training process is one binary SVM classifier per material of interest. Given a novel testing image for which we need to identify the material, we apply all binary classifiers and select the material class corresponding to the classifier with the highest classification score.

5. Experiments and validation metrics

5.1. Data collection and experimental setup

Due to the lack of existing datasets for benchmarking performance of material recognition algorithms in real-world conditions, it was necessary to create a new comprehensive Construction Materials Library (CML) that enables recognition for a large quantity of material types recorded in a variety of as-built contexts. This dataset is for both training and testing purposes so that it can be released to the community for further development and validation of new algorithms. For this purpose, we collected a library of 3000 material samples, each being an uncompressed color image of resolution 200 $\times$ 200 pixels. These images were recorded from seven different scenes (i.e., 5 construction sites and 2 existing buildings).

In order to create a comprehensive dataset, varying degrees of viewpoint, scale, and illumination where accounted for during the collection period span of seven months. The following describes the dataset:


2. Sub-categories: Each course material definition is broken down to several representative sub-types, depending of the material. For example, concrete can be present in a scene as smooth finish, course finish, good condition, or deteriorated. Additionally, each material exhibits a specific appearance under different lighting conditions. Two to four lighting conditions per material where documented.

3. Poses: Scale and orientation variability is documented for each sub-category. For a digital resolution of 200 $\times$ 200 pixels, images representing an approximate physical length of 12”, 24” and 48” were gathered. Three to five view orientations in respect to the surface normal were also considered. Thus, a total of 9–15 poses make up each material sub-category. One sample for the 24” and 48” scales and two non-overlapping samples for the 12” scale were extracted from each pose. The entire dataset is made public at: http://raamac.cee.illinois.edu/materialclassification. Fig. 7 shows examples of the materials in the CML.

5.2. Performance evaluation measures

To quantify and benchmark the performance of the material classification algorithm, we plot the Precision–Recall curve and use a confusion matrix to represent average classification accuracies. These metrics are both set-based measures; i.e., they evaluate the quality of an unordered set of data entries. Here, we define precision and recall as follows:

\[ \text{precision} = \frac{TP}{TP + FP} \]  
\[ \text{recall} = \frac{TP}{TP + FN} \]

where in $TP$ is the number of True Positives, $FN$ is the number of False Negatives and $FP$ is the number of False Positives. For instance, if a brick image is correctly recognized under the brick class, it will be a $TP$; if a concrete image is incorrectly recognized as brick, it will be a $FP$ for the brick class. When a brick image is not recognized under the brick class, then the instance is a $FN$. The particular rule used to interpolate precision at recall level $i$ is to use the maximum precision obtained from the material class for any recall level great than or equal to $i$. For each recall level, the precision is calculated, and then the values are connected and plotted in form of a curve.

The performance of the one-vs.-all multi-class material classifiers is analyzed using a confusion matrix. The confusion matrix returns the average accuracy per material class. The average accuracy of the material classification is calculated using the following formula:

\[ \text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \]

A confusion matrix shows for each pair of material classes ($C_1, C_2$), how many material images from $C_1$ were assigned to $C_2$. Each column of the confusion matrix represents the predicted material class and each row represents the actual material class. The detected $TP$ and $FP$ are compared and the percentage of the correctly predicted classes with respect to the actual class is created and represented in each row.

6. Experimental results

Experiments to benchmark the performance of the proposed material classification method were conducted using the CML
and represent two different histograms and subsequent testing.

The results show that non-linear SVM generally performs better than $\chi^2$ histogram matching and that linear SVM is the worst. RBF and $\chi^2$ kernels show similar performance when used on texture-based features. However, the latter performs significantly better on color based features. For all testing algorithms, the combination of LM and HVS features yields the best results, suggesting the applicability of this feature type for material recognition.

To assess the computation cost of classification versus accuracy, several clustering granularities were tested. Granularity defines the number codewords (hence of bins) in each training and testing histogram. In theory, given $N$ bins, computation of each histogram is an $O(N^3)$ operation and classification with SVM is an $O(N^2)$ operation. This means that, as the number of bins increase, the computation time increases quadratically. It is thus important to limit the number of bins for practical purposes. Table 3 shows a summary of the results. In our implementations, the ranges shown in this table translated to classification times of 0.2 ms/image for 3 bins to 9 ms/image for 20 bins. The results show that the computational time stays in near real-time performance zone. The benchmarking on computational time was conducted on an Intel Core i7-2860QM @ 2.50 GHz with 16 GB RAM running Windows 7SP1 and Matlab 11b.

All features achieve best performance around 15 textons per material. The texture based features hover around in the 70% range with LM showing the highest average precision. However, color based features perform even better, ranging between 80% and 93%, with HSV outperforming RGB in all cases. A significant improvement is shown when LM features are combined with

### Table 2

<table>
<thead>
<tr>
<th>Feature</th>
<th>SVM (Linear) (%)</th>
<th>SVM ($\chi^2$) (%)</th>
<th>SVM (RBF) (%)</th>
<th>$\chi^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM</td>
<td>70.2</td>
<td>80.6</td>
<td>80.1</td>
<td>78.4</td>
</tr>
<tr>
<td>rLM</td>
<td>69.9</td>
<td>75.6</td>
<td>76.3</td>
<td>74.2</td>
</tr>
<tr>
<td>IP</td>
<td>59.1</td>
<td>63.8</td>
<td>64.6</td>
<td>63.7</td>
</tr>
<tr>
<td>HSV</td>
<td>73.3</td>
<td>92.7</td>
<td>89.2</td>
<td>91.1</td>
</tr>
<tr>
<td>RGB</td>
<td>71.1</td>
<td>92.6</td>
<td>87.1</td>
<td>88.0</td>
</tr>
<tr>
<td>LM + RGB</td>
<td>83.1</td>
<td>95.2</td>
<td>92.5</td>
<td>91.2</td>
</tr>
<tr>
<td>LM + HVS</td>
<td>87.1</td>
<td><strong>97.1</strong></td>
<td>93.7</td>
<td>92.1</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Feature</th>
<th>3 (%)</th>
<th>4 (%)</th>
<th>5 (%)</th>
<th>7 (%)</th>
<th>10 (%)</th>
<th>15 (%)</th>
<th>20 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM</td>
<td>66.9</td>
<td>70.4</td>
<td>70.0</td>
<td>73.8</td>
<td>74.9</td>
<td>80.6</td>
<td>77.7</td>
</tr>
<tr>
<td>rLM</td>
<td>62.9</td>
<td>66.9</td>
<td>68.0</td>
<td>73.5</td>
<td>73.7</td>
<td>75.6</td>
<td>73.4</td>
</tr>
<tr>
<td>IP</td>
<td>55.3</td>
<td>59.9</td>
<td>61.0</td>
<td>61.4</td>
<td>63.0</td>
<td>63.8</td>
<td>62.9</td>
</tr>
<tr>
<td>HSV</td>
<td>87.7</td>
<td>92.2</td>
<td>90.8</td>
<td>92.1</td>
<td>91.8</td>
<td>92.7</td>
<td>92.4</td>
</tr>
<tr>
<td>RGB</td>
<td>81.5</td>
<td>83.9</td>
<td>84.6</td>
<td>87.9</td>
<td>91.6</td>
<td>92.6</td>
<td>91.3</td>
</tr>
<tr>
<td>LM15 + HVS</td>
<td>94.00</td>
<td>96.4</td>
<td>96.9</td>
<td>95.7</td>
<td>96.5</td>
<td><strong>97.1</strong></td>
<td>96.0</td>
</tr>
</tbody>
</table>
HSV features. The results of between 94% and 97% remain steady across the range of granularities, with a best performance of 97.1% from combined LM features at 15 bins and HSV features at 15 bins.

We can note that the performance of the LM filter bank for the CML dataset is well under the one reported on the CURet dataset. The average of 78.4% classification rate is much lower than the 95%+ rate as reported in the literature [43,39]. This implies that this approach by itself is not robust enough for use in the context of as-built modeling for on-site photographs. The state-of-the-art AEC algorithm [48] of Brilakis et al. was implemented and tested against the CML benchmark, and an average of 52.8% was reported (see the precision-recall graph in Fig. 8). Here, we trained their classification method of [48] on the same training dataset as our classifier.

Fig. 9 shows the confusion matrix on the performance of our proposed material classification method. As observed in this figure, most material classifications have very high average accuracies (above 90%). Fig. 10 shows several examples of cases where the classification of images were incorrect. In several categories such as cement and concrete classes, there is still 2–7% confusion among them. For these categories, human observation only based on the 200 × 200 pixel images in the library may not result in right classification either, which is primarily related to the absence of context information. In Section 7, we discuss how multiple observations (several images of the same material) or shape/context information can improve the classification results.

6.2. Effect of image quality

On-site data can come from a variety of sources. Photographs can be captured by many devices ranging from high end SLR cameras to smart phones. In addition, images can be altered (compressed) during transfer or archiving. This can produce a data set that is both inconsistent and poor in terms of image quality. Thus it is important to validate our method against a variety of image quality conditions and determine which combination of features and training parameters works best.

The JPEG format was chosen for testing image quality due to its wide use and range of compression settings. By adjusting these settings, common sources of photographic data where replicated, as shown in Fig. 11. The entire image set of the Construction Materials Library was compressed to different degrees and tested. The results in Table 4 show that texture features become unreliable when the quality of images decreases. On the other hand, color based features retain a reasonable accuracy level even for highly compressed images. Combined texture and HSV color features produce the best results for all quality levels.

6.3. Effect of image size and synthesis

Each on-site photographs can depict a collection of objects of different sizes and distances from the camera. Thus the available number and distribution of pixels that represent a single material cannot be predicted. Material classification first requires the selection of a region (set of pixels) of unknown size and shape within the image. Due to the probabilistic nature of the proposed method, a reduction in the available pixels used to generate features can result in a decrease in accuracy. To this end, we test several region sizes to expose the practical limitations of the system. In addition, we test image synthesis techniques to test if the available real pixels can be combined with synthetically generated pixel in order to overcome the above mentioned limitation.

The original data contains images of 200 × 200 pixels in size. The 49 × 49 pixel size of the filters used to generate textural features reduces the effective area to 152 × 152 pixels. For an input image size of 100 × 100 pixels, this area is reduced by a factor of 4, and for 50 × 50 pixels, features can be extracted for only four pixels. To test the effects of size reduction, the images of the CML were progressively subsampled with an finite impulse response (FIR) filter [57] to retain consistency in respect to anti-aliasing.

To improve classification results, the smaller sized versions of the CML where expanded by two synthesis techniques: (1) A naive reflection/rotation (RR) method was used to expand the images with pixels similar to the ones around the edges; and (2) An image quilting (IQ) algorithm [58] was used to generate a larger sized image from samples of the original. The effect of these techniques on a sample of brick is shown in Fig. 12.

The results in Table 5 and Fig. 13 show several important results. First, we note that, in all experiments, the use of joint texture and color codebooks outperforms texture or color alone. Furthermore, it is shown to be viable for images as small as 30 × 30 pixels in size, maintaining an average precision of above 90.8%. Second, for smaller image sizes, expansion by synthesis increases precision significantly, with the RR method outperforming image quilting. Third, with the reduction of image size, texture based classification deteriorates drastically while color based classification suffers only slightly. This is consistent with the intuition that textures become less apparent when described by fewer pixels while the mean color of the image remains constant. Other techniques for image synthesis have been reported in literature, however the search for the optimal synthesis algorithm for material classification is left for future research.

7. Discussion on the proposed method and research challenges

The authors introduce a robust material classification method and the first comprehensive dataset for semantically-rich as-built 3D modeling purposes. The average accuracies of 97.1% across all materials promise the applicability of the proposed method. For detecting material in smaller image patches, we also presented
Fig. 9. Confusion matrix representing the cross validated average accuracies of the proposed method for material classification.

<table>
<thead>
<tr>
<th></th>
<th>Asphalt</th>
<th>Brick</th>
<th>Cement - Granular</th>
<th>Cement - Smooth</th>
<th>Concrete - Cast</th>
<th>Concrete - Precast</th>
<th>Foliage</th>
<th>Form Work</th>
<th>Grass</th>
<th>Gravel</th>
<th>Marble</th>
<th>Metal - Grills</th>
<th>Paving</th>
<th>Soil - Compact</th>
<th>Soil - Vegetation</th>
<th>Soil - Loose</th>
<th>Soil - Mulch</th>
<th>Stone - Granular</th>
<th>Stone - Limestone</th>
<th>Wood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asphalt</td>
<td>94.7</td>
<td>1.7</td>
<td>0.6</td>
<td>0.3</td>
<td>1.1</td>
<td>0.8</td>
<td>0.3</td>
<td>100</td>
<td>100</td>
<td>0.5</td>
<td>1.0</td>
<td>0.3</td>
<td>0.9</td>
<td>1.2</td>
<td>0.7</td>
<td>0.9</td>
<td>0.3</td>
<td>0.3</td>
<td>96.2</td>
<td>1.7</td>
</tr>
<tr>
<td>Brick</td>
<td>99.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>2.7</td>
<td>0.5</td>
<td>0.2</td>
<td>100</td>
<td>100</td>
<td>96.2</td>
<td>1.7</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>96.2</td>
<td>1.7</td>
</tr>
<tr>
<td>Cement - Granular</td>
<td>93.6</td>
<td>1.8</td>
<td>3.1</td>
<td>0.6</td>
<td>1.2</td>
<td>0.6</td>
<td>0.2</td>
<td>100</td>
<td>100</td>
<td>1.0</td>
<td>2.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>97.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Cement - Smooth</td>
<td>92.6</td>
<td>2.9</td>
<td>3.2</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
<td>100</td>
<td>100</td>
<td>1.4</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>95.8</td>
<td>1.0</td>
</tr>
<tr>
<td>Concrete - Cast</td>
<td>92.1</td>
<td>2.9</td>
<td>3.2</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
<td>100</td>
<td>100</td>
<td>1.4</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>95.8</td>
<td>1.0</td>
</tr>
<tr>
<td>Concrete - Precast</td>
<td>91.9</td>
<td>2.9</td>
<td>3.2</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
<td>100</td>
<td>100</td>
<td>1.4</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>95.8</td>
<td>1.0</td>
</tr>
<tr>
<td>Foliage</td>
<td>99.7</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>2.7</td>
<td>0.5</td>
<td>0.2</td>
<td>100</td>
<td>100</td>
<td>96.2</td>
<td>1.7</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>96.2</td>
<td>1.7</td>
</tr>
<tr>
<td>Form Work</td>
<td>1.3</td>
<td>1.0</td>
<td>2.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>100</td>
<td>1.2</td>
<td>1.2</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>95.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Grass</td>
<td>0.5</td>
<td>1.0</td>
<td>2.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>100</td>
<td>1.2</td>
<td>1.2</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>95.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Gravel</td>
<td>1.4</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>100</td>
<td>100</td>
<td>1.4</td>
<td>1.4</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>93.2</td>
<td>1.8</td>
</tr>
<tr>
<td>Marble</td>
<td>0.4</td>
<td>0.4</td>
<td>1.8</td>
<td>1.5</td>
<td>0.7</td>
<td>1.1</td>
<td>1.1</td>
<td>100</td>
<td>100</td>
<td>0.4</td>
<td>0.4</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>94.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Metal - Grills</td>
<td>1.6</td>
<td>2.1</td>
<td>3.2</td>
<td>2.1</td>
<td>3.2</td>
<td>3.2</td>
<td>3.2</td>
<td>100</td>
<td>100</td>
<td>0.4</td>
<td>0.4</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>97.6</td>
<td>0.3</td>
</tr>
<tr>
<td>Paving</td>
<td>0.4</td>
<td>1.0</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>100</td>
<td>0.4</td>
<td>0.4</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>97.6</td>
<td>0.3</td>
</tr>
<tr>
<td>Soil - Compact</td>
<td>0.4</td>
<td>1.0</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>100</td>
<td>100</td>
<td>0.4</td>
<td>0.4</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>97.6</td>
<td>0.3</td>
</tr>
<tr>
<td>Soil - Vegetation</td>
<td>1.4</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>100</td>
<td>100</td>
<td>1.4</td>
<td>1.4</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>94.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Soil - Loose</td>
<td>1.4</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>100</td>
<td>100</td>
<td>1.4</td>
<td>1.4</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>94.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Soil - Mulch</td>
<td>1.4</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>100</td>
<td>100</td>
<td>1.4</td>
<td>1.4</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>94.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Stone - Granular</td>
<td>1.4</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>100</td>
<td>100</td>
<td>1.4</td>
<td>1.4</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>94.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Stone - Limestone</td>
<td>1.4</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>100</td>
<td>100</td>
<td>1.4</td>
<td>1.4</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>94.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Wood</td>
<td>1.4</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>100</td>
<td>100</td>
<td>1.4</td>
<td>1.4</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>94.1</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Fig. 10. Several examples of incorrect material classifications.

BMP 117kb | JPEG 100 37.5kb | JPEG 75 8.2kb | JPEG 50 5.6kb | JPEG 25 3.8kb | JPEG 10 2.3kb

SLR Cameras | Mobile Phones | Typical Web Uploads | Very Low Quality Sources

Fig. 11. A range of compression settings were used to simulate an array of possible image sources.
the RR method and illustrated that our method can stay maintain accuracies higher than 90%.

Such results also indicate the robustness of the method to dynamic changes of illumination, viewpoint, camera resolution, and scale which is typical for unordered construction image collections. The method was also shown to remain effective for very low quality images. While this paper presented the initial steps towards processing unordered image collection for the purpose of material recognition, several critical challenges remain. Some of the open research problems include:

1. **The impact of depth textures**: The effect of texture maps generated from depth values of a reconstructed point cloud from the observed material.

2. **Expanded image sources**: Additional effects on quality, such as images originating from video frames or scenes with motion.

3. **Robust image synthesis methods**: To better classify material for very small image patches (e.g., when time-lapse images are used for detecting progress of concrete columns that are very far from the camera. In these cases, the number of pixels associated with the element could be very minimal).

4. **Multiple images**: The effect of leveraging multiple images registered to reconstructed point cloud models – typically generated through Structure-from-Motion algorithms – through various voting/scoring methods;

### Table 4
Effect of Image Compression.

<table>
<thead>
<tr>
<th>Feature</th>
<th>BMP (%)</th>
<th>JPEG100 (%)</th>
<th>JPEG90 (%)</th>
<th>JPEG75 (%)</th>
<th>JPEG50</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM15</td>
<td>80.6</td>
<td>74.6</td>
<td>74.8</td>
<td>39.9</td>
<td>16.6</td>
</tr>
<tr>
<td>HSV15</td>
<td>92.7</td>
<td>91.9</td>
<td>90.8</td>
<td>91.5</td>
<td>85.9</td>
</tr>
<tr>
<td>LM15 + HSV15</td>
<td><strong>97.1</strong></td>
<td>96.0</td>
<td>96.6</td>
<td>96.4</td>
<td>92.7</td>
</tr>
</tbody>
</table>

Fig. 12. Synthesizing the input images to increase area.

### Table 5
Effect of Image Size on Accuracy of Material Classification.

<table>
<thead>
<tr>
<th>Size (n x n pixels)</th>
<th>Feature</th>
<th>Original</th>
<th>RR</th>
<th>IQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>LM15</td>
<td>80.6%</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>HSV15</td>
<td>92.7%</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>LM15 + HSV15</td>
<td><strong>97.1%</strong></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>150</td>
<td>LM15</td>
<td>69.7%</td>
<td>76.4%</td>
<td>63.7%</td>
</tr>
<tr>
<td></td>
<td>HSV15</td>
<td>91.5%</td>
<td>92.1%</td>
<td>89.7%</td>
</tr>
<tr>
<td></td>
<td>LM15 + HSV15</td>
<td>96.1%</td>
<td>96.6%</td>
<td>93.7%</td>
</tr>
<tr>
<td>100</td>
<td>LM15</td>
<td>56.0%</td>
<td>66.4%</td>
<td>53.3%</td>
</tr>
<tr>
<td></td>
<td>HSV15</td>
<td>93.0%</td>
<td>91.7%</td>
<td>85.2%</td>
</tr>
<tr>
<td></td>
<td>LM15 + HSV15</td>
<td>94.3%</td>
<td>96.4%</td>
<td>93.6%</td>
</tr>
<tr>
<td>75</td>
<td>LM15</td>
<td>39.9%</td>
<td>59.6%</td>
<td>47.9%</td>
</tr>
<tr>
<td></td>
<td>HSV15</td>
<td>91.5%</td>
<td>91.6%</td>
<td>88.5%</td>
</tr>
<tr>
<td></td>
<td>LM15 + HSV15</td>
<td>92.1%</td>
<td>94.5%</td>
<td>92.2%</td>
</tr>
<tr>
<td>50</td>
<td>LM15</td>
<td>16.6%</td>
<td>53.9%</td>
<td>37.1%</td>
</tr>
<tr>
<td></td>
<td>HSV15</td>
<td>88.4%</td>
<td>87.3%</td>
<td>87.3%</td>
</tr>
<tr>
<td></td>
<td>LM15 + HSV15</td>
<td>92.0%</td>
<td>93.4%</td>
<td>86.8%</td>
</tr>
<tr>
<td>30</td>
<td>LM15</td>
<td>N/A</td>
<td>30.1%</td>
<td>30.3%</td>
</tr>
<tr>
<td></td>
<td>HSV15</td>
<td>84.8%</td>
<td>87.6%</td>
<td>84.2%</td>
</tr>
<tr>
<td></td>
<td>LM15 + HSV15</td>
<td>N/A</td>
<td>90.8%</td>
<td>85.7%</td>
</tr>
</tbody>
</table>

Fig. 13. The effect of image size on classification precision for images expanded synthetically with RR (except for original images).
5. Leveraging shape/context prior: For the purposes of as-built modeling, the effect of 3D shape/context information extracted from a point cloud reconstruction (e.g., this material belongs to a wall as opposed to a slab) on classification accuracy.

6. Leveraging 4D BIM: For the purposes of model-based progress monitoring, the impact of material expectation (e.g., for monitoring construction of a concrete wall, the expected materials are rebar cage, formwork, and concrete) can reduce the set of possible classes and could increase accuracy.

8. Conclusion and future work

This paper presented a robust material classification method for semantically-rich as-built 3D modeling and construction monitoring purposes. A Construction Materials Laboratory was formulated to train and test the proposed method. This real-world dataset exposed the limitations of current material classification methods and validated the performance of the proposed algorithm. From our testing, an average classification rate of 97.1% was achieved. The effects of various parameters such as the choice of the filter bank, the number of Textons and color bins, the size of the histograms, the impact of camera parameters, and various classifiers and kernels were considered. Furthermore, we have shown that, by leveraging image synthesis techniques, our method can produce accurate results for images as small as 30 × 30 pixels in size. The addition of HSV color channels as a feature for training results in a more robust classification algorithm, and can be implemented as critical process in an automated as-built 3D modeling system. In case of images with larger resolutions, our algorithm can act as a sampling method by classifying materials for a given number of 200 × 200 pixel image patches. Alternatively, larger resolution images could also be subsampled to a smaller resolution to decrease computation time.

The presented method has not yet addressed the challenge of segmentation. Given a photograph taken on site, a region belonging to a single material needs to be identified. Material classification can then be performed on the pixels in that region. In the presence of BIM, these regions can be derived by back-projecting BIM element surfaces with a prior expectation of material onto site images. In the presence of a segmented point cloud, the regions can be derived from back-projecting points and in-painting them. Our ongoing research will explore the use of the outlined method as a technique for random sample consensus within non-uniform regions in a single image as well as across multiple viewpoints. Further applications of the proposed material classification for segmentation of 3D point cloud models and model-based progress monitoring can be studied. In addition, material classification performance could be improved by exploring the incorporation of 3D geometrical information. The geometric surface properties of a given surface could supply additional descriptive features that can be integrated along with the aforementioned image based features.

Acknowledgments

The support of Turner Construction, Holder Construction, Gilbane Construction, and Skanska with data collection and members of the RAAMAC lab with collecting and annotating construction material images is acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the companies mentioned above.

References


