IMPROVEMENTS TO CONCRETE COLUMN DETECTION IN LIVE VIDEO

Zhenhua Zhu, Ibrahima J. Ndiour, Ioannis Brilakis and Patricio A. Vela
Georgia Institute of Technology, Atlanta, USA
{zhzhu, ndiour, brilakis, pvela}@gatech.edu

Abstract

The automated detection of structural elements (e.g. concrete columns) in visual data is useful in many construction and maintenance applications. The research in this area is under initial investigation. The authors previously presented a concrete column detection method that utilized boundary and color information as detection cues. However, the method is sensitive to parameter selection, which reduces its ability to robustly detect concrete columns in live videos. Compared against the previous method, the new method presented in this paper reduces the reliance of parameter settings mainly in three aspects. First, edges are located using color information. Secondly, the orientation information of edge points is considered in constructing column boundaries. Thirdly, an artificial neural network for concrete material classification is developed to replace concrete sample matching. The method is tested using live videos, and results are compared with the results obtained with the previous method to demonstrate the new method improvements.

KEYWORDS: robust detection, concrete columns, live videos.

INTRODUCTION

Many building construction and maintenance applications will be facilitated, if structural elements (e.g. concrete columns, beams and walls) can be automatically detected from construction site images and videos. For example, project progress will be monitored to tell whether the project is ahead or behind schedule by comparing the detected as-built structural elements with project's 4D as-designed models (Golparva-Fard et al., 2009). Construction operations can be analyzed to indicate whether they have high or low productivity by knowing how many structural elements have been built within a certain period (Gong and Caldas, 2009). Traditional routine infrastructure inspection practices can be enhanced, when the defects are automatically located on structural element surfaces, and further their impacts are evaluated with quantitative measurements (Zhu and Brilakis, 2010).

Although automated structural element detection is useful, the research in this area is under initial investigation. Most existing methods are not adequate enough, as they relied on either boundary or color/texture information to detect structural elements. The sole reliance on boundary information makes the methods fail in differentiating structural elements with other objects that have similar shapes (e.g. concrete columns vs. wooden table legs). Similarly, although color and texture are good indicators for object recognition, the sole use of color and texture makes the methods unable to differentiate two structural elements in a building, when these two structural elements have the same material and connect each other. The authors

previously presented a concrete column detection method that combined boundary and color information as detection cues (Zhu and Brilakis, 2010).

Compared against the previous method, this paper presents a new concrete column detection method that aims to reducing the reliance of parameter settings mainly in three aspects. First, edge maps are produced using RGB (Red, Green and Blue) values instead of intensity values. Color-based edge detectors can locate edges that were not found by intensity-based edge detectors. Secondly, the orientation information of edge points is considered in the Hough transform to construct column boundaries. Based on the orientation information, each edge points only votes for one line once. This not only eliminates the number of falsely detected lines, but also significantly reduces the computation load in the Hough transform. Thirdly, machine learning techniques are introduced to generalize concrete material classification knowledge. An artificial neural network for concrete material classification is designed to replace concrete sample matching.

The method is tested using live videos indoors and outdoors. The performance of the method is measured using the criteria suggested in the signal detection theory (Wickens, 2002). The ground truth is identified manually first, and then the test results are compared with manual detection ones to find the true positive (TP), false positive (FP) and false negative (FN). The detection precision and recall are calculated as TP / (TP+FP), and TP / (TP+FN). The results are also compared with the results obtained with the previous method to demonstrate the new method improvements.

BACKGROUND

Detecting structural elements is always the first step to automate construction applications, such as progress monitoring (Wu and Kim, 2004; Golpard-Vard et al., 2009), productivity measurement (Chi et al. 2009; Gong and Caldas, 2009), and construction safety improvement (Teizer, 2008). Given a template, the structural element detection in general is regarded as the problem of locating the structural elements that "looks" similar to the template (Ge et al. 2008). Depending on the different template types, most existing methods are classified as color/texture-based and shape-based. Although scale/affine-invariant features can also use to visually characterize a specific object in multiple images, they are not appropriate for structural element category detection. This is because the extracted "features" only reflect the individual characteristics of each structural element not the salient features of the whole structural element category (Zhu and Brilakis, 2010).

Color/Texture-based Structural Element Detection

Color/Texture-based methods make use of structural elements' material information to perform detection. Neto et al. (2002) observed that construction materials, such as concrete and steel, do not have a constant color value but a range of nuance of the value due to the variance of light conditions or the nature of materials. Based on this observation, their method starts from identifying structural elements' boundaries by tracing the image pixels that satisfy two conditions: 1) the color values of these pixels fit in the color value range; and 2) the color values of at least one of their neighbors does not fit in the color value range. When one boundary has been identified, all internal pixels in the boundary are checked. Those whose color values do not fit in structural elements' special color value range are removed, and the remaining pixels represent structural element regions.

Brilakis and Soibelman (2008) created a method of detecting linear structural elements using the concept of material "signature". In their method, an image is first cropped into regions. The signature of each region is represented by a vector of the mean and the standard deviation of region's color and texture values. This signature is compared with the signatures of material samples by measuring their Euclidean distances. If the distance of the region's signature to the signature of one material sample is smaller than a pre-defined threshold, the region is assumed to contain that material. For example, if the signature of one image region is matched with that of a concrete sample, the material contained in the region is assumed to be concrete. When all material regions are identified, their maximum cluster dimensions (MCD) and maximum dimensions along the perpendicular axis of MCD (PMCD) are calculated. The region is regarded as a column (or beam) based on three assumptions: 1) the region is linear, if its MCD is significantly larger than PMCD; 2) if the region is linear, then the tangent of the MCD edge points represents its direction on the image plane; and 3) the linear region is a column (or beam), if the computed direction is within 45° from the vertical (or horizontal) image axis.

Color/texture-based methods rely on the retrieved material information to infer the shape information of structural elements. This may lead to their failure in detecting structural elements of most frame structures, where one structural element (e.g. a concrete column) is connected to another structural element with the same material (e.g. a concrete beam) (Zhu and Brilakis, 2010). It was also found that the methods performed poorly in detecting thin and stick like components, as their color and texture are easily corrupted (David, 2005).

Shape-based Structural Element Detection

Shape-based detection methods make use of structural elements' boundary information to locate the elements. This boundary information preserves the important object's structure properties (e.g. discontinuities in object surface orientation and changes in object surface material), and are therefore useful in detecting thin and stick-like objects that are difficult to be detected using color/textures-based methods.

So far, several shape-based structural element detection methods have been created. Wu and Kim (2004) used the Canny edge detector to extract structural component regions from image background with gaps close and hole fill mechanisms. Lukins and Trucco (2007) developed a concrete column classifier using the Haar features (e.g. edge features, line features and centre-surround features) and the Adaboost algorithm (Viola and Jones, 2001). The classifier was trained on a set of one hundred concrete column samples, and the test results indicated that the identification of columns is not entirely reliable, as some scaffolds are also identified as concrete columns (Ibrahim and Kaka, 2008). Wu et al. (2010) recently proposed a concrete column recognition method using morphology-based 3D CAD model filtering techniques. In their method, the results from the Canny edge detector and Watershed Transformation are fused to retrieve image edge information. Then, a 3D CAD-based image mask is generated by aligning a 3D CAD model to the image with the manual measurements of camera location and orientation. The mask filters the retrieved edge information to get desired columns.

In general, the sole reliance on the shape information is not always robust in object detection. The algorithms based on edge statistical analysis only (i.e. high edge magnitude and variance) may fail in complex and cluttered scenes; and also, the failure may be produced, when the

boundaries of the object of interest are not clear due to damages or dirt, or other objects that have similar shapes are also presented (Zhu and Brilakis, 2010).

Previous Work in Concrete Column Detection

The authors recently combined both color/texture and boundary information as concrete column detection cues. For each concrete column, it was found that (1) the shape of one concrete column surface is dominated by a pair of long vertical lines and (2) the texture and color patterns on the surface are uniform. Based on these two findings, a concrete column detection method was proposed (Figure 1).
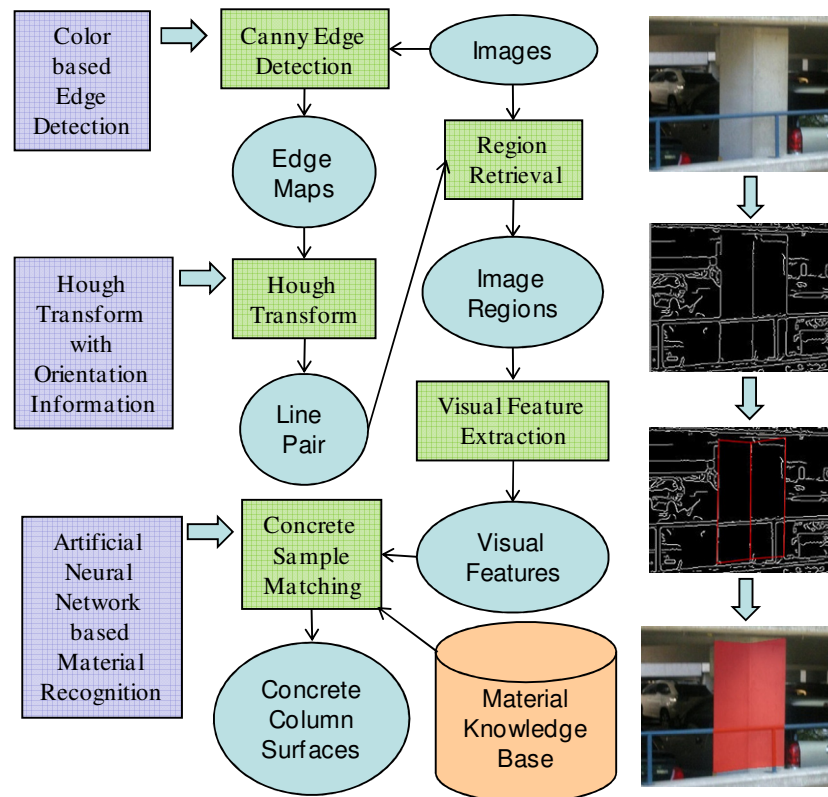

Figure 1: Previous Method Overview and Improvements

The method starts from the Canny edge detector to produce image's edge map, where edge points are shaded white and non-edge points are shaded black. The map does not contain any line information imposed on the edge points. In order to retrieve long vertical lines from the edge points, the Hough transform is applied. The transform maps the edge points' Cartesian coordinates (x, y) into a radius-and-angle (ρ, θ) parameterization space and records the voting number for any possible pair of ρ and θ to satisfy Eq. 1 at each edge point. The places having high voting number along the line of $\theta = 0$ in the parameterization space are located, and the corresponding radius values (ρ) are used to formulate long vertical lines.

$$\rho = x\cos(\theta) + y\sin(\theta) \tag{Eq. 1}$$

Each long vertical line is compared with its neighbouring ones. Two lines are regarded as a pair only when they are close to each other and have similar length. The comparison is performed iteratively, till no pair can be found. A bounding rectangle for each pair is constructed, and its width/height ratio is checked. If the ratio is larger than 1, it means the rectangle's width is larger than its height, which is not common for most concrete columns in practical buildings. Therefore, the pairs like this are discarded. For the remaining pairs of vertical lines, the image region contained in each pair is retrieved. The region's material signature is calculated using the method presented by Brilakis et al. (2006), and compared with the signatures of pre-defined concrete samples. If its material is determined to be concrete, the corresponding image region is believed to be a concrete column's surface.

Most concrete columns in images can be successfully located with the method, when the optimal parameters are experimentally selected. The parameters include the ones for edge detection, line pair extraction, and concrete sample matching. The sensitivity of the method to these parameters reduces the ability of the method to robustly detect concrete columns in live videos where the manual selection of optimal parameters is infeasible.

METHOD IMPROVEMENTS

In order to robustly detect concrete columns in live videos, the method presented in this paper reduces the reliance of parameter settings from three aspects (Figure 1). Compared against the previous method, the new method first locates edge points using their color information, which permits a more complete representation of images (Liu et al. 2007). Secondly, the orientation information of edge points is considered in the Hough transform to find vertical lines. At each edge point, not all possible pairs of ρ and θ that satisfy Eq. 1 are recorded. Instead, θ is fixed by edge point's gradient direction. This not only removes falsely detected lines, but also greatly reduces the computation burden. Thirdly, machine learning techniques are adopted to classify concrete materials. An artificial neural network based concrete material classifier is created to replace concrete sample matching.

Column Boundary Retrieval

In the Canny edge detector, the edges detected from a gray image are controlled by two hysteresis thresholds: one for initial strong edge segments locating and the other for edge linking. Previous research results indicated that these two thresholds significantly affect edge detection results. Therefore, a color-based edge detector is used.

Consider a color image as a mapping from an x-y plane to a RGB space. The Jacobian matrix (J_c) of the mapping is represented in Eq. 2, which indicates the color change induced by moving any infinitesimal step (dx, dy) in the image plane. The Euclidean squared magnitude of the change can then be calculated as Eq. 3, where the largest eigenvalue of the matrix Q is the gradient magnitude and the eigenvector corresponding to the largest eigenvalue is the gradient direction. When the precise gradient magnitude and direction of each pixel (x, y) are retrieved, those with local maximum gradient magnitude in their gradient direction are marked as potential edge points (Liu et al. 2007). These points are further checked and the isolated potential edge points are treated as noise and removed from edge point collection.

$$J_c = \begin{bmatrix} \partial R/\partial x & \partial R/\partial y \\ \partial G/\partial x & \partial G/\partial y \\ \partial B/\partial x & \partial B/\partial y \end{bmatrix} \quad \text{(Eq.2)}$$

$$M^2 = (dx,dy)Q(dx,dy)^T$$
$$where \; Q = J_c^T J_c \quad \text{(Eq.3)}$$

Not all the retrieved edge points are considered to construct long vertical lines. Instead, only edge points with horizontal gradient direction are used to vote in the Hough Transform. Moreover, at each edge point, not all possible pairs (ρ, θ) that satisfy Eq. 1 are enumerated and recorded. θ is fixed based on the edge point's gradient direction obtained in the previous edge detection stage. Therefore, each edge point only votes once.

Once lines are detected using the Hough-based approach described previously, there remains the need to associate line pairs together in order to find potential columns and other vertical structures in the image. This association is completed using geometric criteria based on line proximity, both horizontally and vertically. For a given reference line described by the coordinates $(xref_1, yref_1)$ and $(xref_2, yref_2)$ of its extremities and a line candidate given by (x_1, y_1) and (x_2, y_2), we define the distance based on Eq. 4. For all candidate lines, the distance to the reference line is computed. Subsequently, the candidate line with minimum distance is paired to the reference line. In our experiments, the value 0.6 is used.

$$d = \alpha \cdot \left(\|x_{ref1} - x_1\| + \|x_{ref2} - x_2\| \right) + (1-\alpha) \cdot \left(\|y_{ref1} - y_1\| + \|y_{ref2} - y_2\| \right) \quad \text{(Eq. 4)}$$

Concrete Material Classifier

The regions between long vertical line pairs are extracted, and the visual features of the regions are calculated. These visual features include the ratios of the normalized red to the normalized green and the normalized blue to the normalized green. Instead of using statistical concrete sample matching, an artificial neural network is created to classify whether the materials contained in the regions belong to concrete or not. The network is composed of three layers (one input layer, one hidden layer and one output layer). In the input layer, the network takes, and transfers them to the hidden layer, where one material classification boundary is created on each neuron-node. Then, these boundaries built on the neuron-nodes of the hidden layer are combined together using an AND operation in the output layer. A classification region in the feature space is formed.

The network is trained through supervised learning. The visual features of training samples are first extracted and their output values of the network are manually specified as 1 for concrete samples and -1 for non-concrete samples. During the training process, the network automatically adjusts the neurons' weights to minimize identification errors. The training is repeated thousands of times. Although it is possible to have more repetitions to further reduce error rates, the long training will result in an overtraining problem, which makes the network only memorize the peculiar patterns of the training samples.

In cases where the appearance of the concrete columns cannot be described by the artificial neural network, thresholding in the RGB colorspace is performed to achieve concrete material classification. Specifically, the mean color intensity of the material sample is defined as well as the standard deviations tolerated along each channel (R, G, B). From these quantities, a parallelepiped is defined in the RGB colorspace. A pixel whose color intensity falls inside the parallelepiped is then classified as concrete while being classified as no-

concrete when the color intensity falls outside. This is a simple, fast and very efficient binary classification method that can be used to perform concrete material classification when the appearance model of the concrete samples remains unimodal.

IMPLEMENTATION AND RESULTS

The aforementioned improvements were implemented into a prototype developed using Microsoft Visual C++. OpenCV (Intel® Open Source Computer Vision Library) was used as the prototype's main image processing library. The prototype run on a Laptop (HP Compaq 2210b with Intel® Core™2 Duo Processor T7100 and 1G Memory) connected to a regular web camera (Logitech Pro 9000) (Figure 2). The video resolution is fixed at 960x720. Figure 3 shows concrete column detection results in a series of video frames.

Figure 2: Camera Set-up

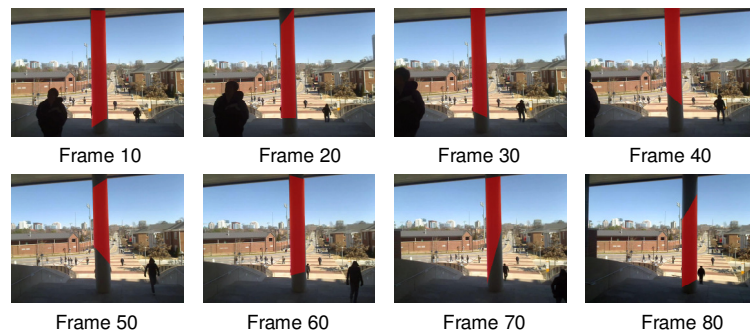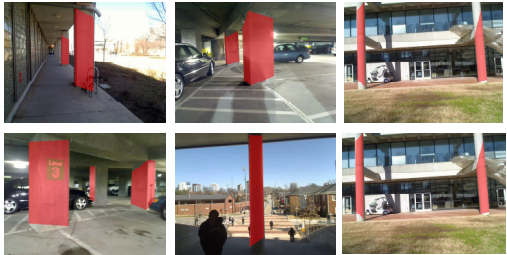| Frame 10 | Frame 20 | Frame 30 | Frame 40 |
| Frame 50 | Frame 60 | Frame 70 | Frame 80 |

Figure 3: Concrete Column Detection Results from a Video

True Positive (TP), False Positive (FP) and False Negative (FN) are figured out based on the results. In each video frame, TP is the number of correctly detected concrete columns; FP is the number of incorrectly detected concrete columns; and FN is the number of actual concrete columns that are not detected by the method. Therefore, the total number of actual concrete columns in each video frame can be represented as the sum of TP and FN (TP + FN), and the total number of detected concrete columns can be represented as the sum of TP and FP (TP+FP). The detection precision of the method is calculated as TP / (TP+FP), which indicates the detection exactness or fidelity. High (low) detection precision means many (few) detected concrete columns are actual concrete columns. The detection recall of the method is calculated as TP / (TP+FN), which measures the detection completeness. High (low) detection recall means many (few) actual concrete columns are correctly detected. Table 1 illustrates the detection precision and recall ratios from five videos, where sample concrete column detection results are included.

In order to demonstrate the new method improvements, the concrete column detection results obtained by the new method are compared with the results obtained with the previous method frame by frame (Zhu and Brilakis, 2010). Figure 4 shows both methods' concrete column detection results from the video #1 in Table 1; and the corresponding TPs, FPs and FNs for the first one hundred video frames are illustrated in Figure 5.

Table 1: Detection precision and recall from five videos with samples included

Video	# of Video Frames	Precision [%]	Recall [%]
#1	250	98.85	91.06
#2	184	98.84	63.33
#3	56	80.00	87.27
#4	100	80.95	62.33
#5	100	93.71	74.87
Total	690	92.55	76.22

CONCLUSIONS

The detection of structural elements from images/videos is useful for facilitating many building construction and maintenance applications including project progress monitoring, construction productivity analysis, and automated infrastructure visual inspection. The research in this area is under initial investigation. Most existing methods relied on either structural elements' shape information or their color/texture information to perform detection. The sole reliance on shape information makes the methods unable to differentiate structural elements with other objects that have similar shapes. Similarly, the sole use of color and texture makes the methods fail in identifying two structural elements in a building, when these two structural elements have the same material and connect each other.

The authors previously proposed a method of combining the boundary and color/texture information as concrete column detection cues. Most concrete columns in images can be located when the optimal parameters of the method are experimentally set. However, it was found that the detection results were easily influenced by parameter settings. This makes the method unable to robustly detect concrete columns in live videos where the manual adjustment of parameters for each video frame is impossible. The method presented in this improves the previous one mainly from three aspects. First, edge points are located using the precise gradient magnitude and direction obtained from their color values. Secondly, the gradient direction information at each edge point is used in the Hough transform to construct column boundaries. Thirdly, an artificial neural network for concrete material classification is designed to replace traditional concrete sample matching.

The method was tested using live videos indoors and outdoors. Preliminary test results indicated that three improvements can significantly reduce the method's sensitivity to parameter setting for detecting concrete column in videos. Compared with the previous methods, the overall detection precision (represented as TP / (TP+FP)) and recall (represented as TP / (TP+FN)) of the improved method can reach 92.55% and 76.22%.

Figure 4: Detection Results for Both Methods from the Video #1 in Table 1

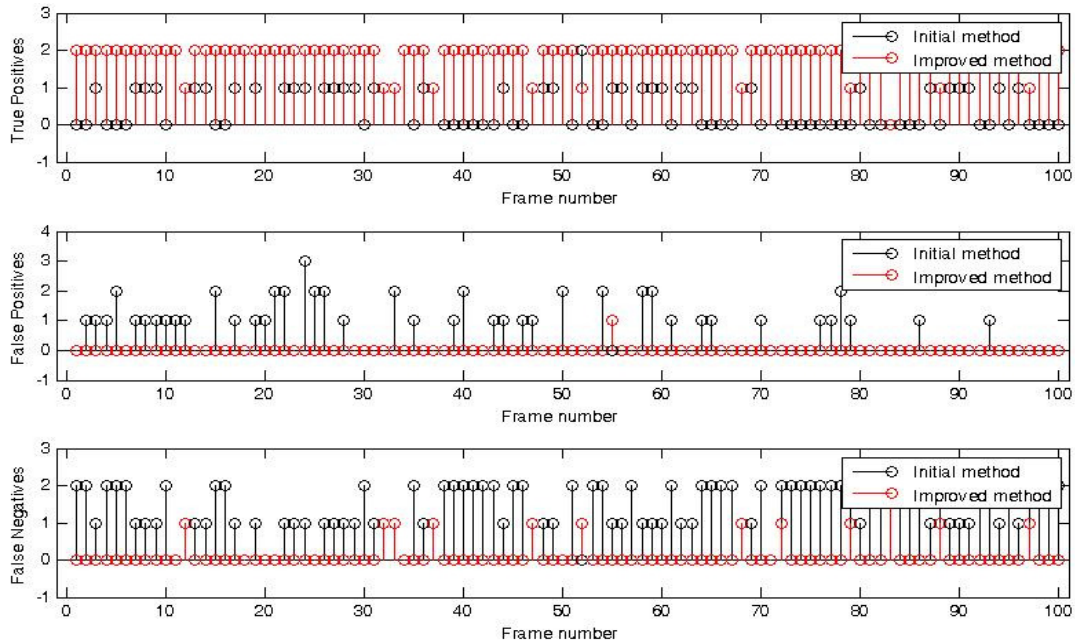Figure 5: True Positives, False Positives and False Negatives from the first 100 video frames

REFERENCES

Brilakis, I. and Soibelman, L., (2008) "Shape-Based Retrieval of Construction Site Photographs", J. of Comp. in Civ. Engr., 22(1): 14 – 20

Brilakis, I., Soibelman, L., and Shinagawa, Y., (2006) "Construction Site Image Retrieval Based on Material Cluster Recognition", J. of Adv. Engr. Info., 20(4): 443 – 452

Chi, S., Caldas, C., and Kim, D., (2009) "A Methodology for Object Identification and Tracking in Construction based on Spatial Modeling and Image Matching Techniques", J. Of Computer-Aided Civ. and Infr. Engr., 24(3): 199 - 211.

David, P. and DeMenthone, D., (2005) "Object Recognition in High Clutter Images using Line Features", Proc. of the 10th IEEE Int. Conf. on Computer Vision, 2(7-21): 1581-1588.

Ge, F., Liu, T., Wang, S. and Stahl, J., (2008) "Template-Based Object Detection through Partial Shape Matching and Boundary Verification", J. of Sig. Processing, 4(1): 148 – 157

Golparvar-Fard M., Savarese S. and Peña-Mora F., (2009) "Interactive Visual Construction Progress Monitoring with 4D Augmented Reality Model", In: Proc., 2009 CRC, pp. 41-50.

Gong, J. and Caldas, C., (2008), "Construction Site Vision Workbench: A Software Framework for Real-Time Process Analysis of Cyclic Construction Operations", In: Proc. of the 2009 ASCE International Workshop on Computing in Civil Engineering, Austin, TX

Ibrahim, Y. And Kaka, A., (2008), "Review of Photographic/Imaging Applications in Construction", The Built & Human Environment Review, Vol. 1, pp. 99-117

Liu, Y., Ikenaga, T., and Goto, S., (2007) "An MRF model-based approach to the detection of rectangular shape objects in color images", J. of Sig. Processing 87(2007): 2649-2658

Lukins, T., and Trucco, E., (2007) "Towards Automated Visual Assessment of Progress in Construction Projects", Proc of the British Machine Vision Conference, Warwick UK.

Neto, J. and Arditi, D., (2002) "Using Colors to Detect Structural Components in Digital Pictures", Computer Aided Civil and Infrastructure Engineering, 17(2002): 61-76

Teizer, J., (2008) "3D Range Image Sensing for Active Safety in Construction", J. of Information Technology in Constr., Sensors in Constr. & Infra. Mgmt., Issue 13, pp. 103-117

Viola, P. and Jones, M. (2001), "Rapid object detection using a boosted cascade of simple features", IEEE Conference on Computer Vision and Pattern Recog.. Vol. 1, pp. 511-518

Wickens, T., (2002) "Elementary Signal Detection Theory", ISBN 0195092503

Wu, Y., and Kim, H., (2004) "Digital Imaging in Assessment of Construction Project Progress", In: Proc. Of the 21th ISARC, pp. 537 – 542.

Zhu, Z. and Brilakis, I. (2010) "Concrete Column Recognition in Images and Videos", Journal of Computing in Civil Engineering, American Society of Civil Engineers, (in print)