A MODIFIED SHADING MODEL METHOD FOR BUILDING DETECTION

V. Zeljkovic[1], A. Dorado [2,3,] and E. Izquierdo[3]*

[1]Faculty of Technical Sciences
D. Obradoviæa 6, 21000 Novi Sad
Serbia and Montenegro
vesnaz@uzzpro.sr.gov.yu

[2]School of Engineering
Pontificia Universidad Javeriana
Calle 18 #118-250 Cali, Colombia
andres.dorado@elec.qmul.ac.uk

[3]Electronic Engineering Department
Queen Mary, University of London
London E1 4NS, United Kingdom
ebroul.izquierdo@elec.qmul.ac.uk

ABSTRACT

A technique for detection of building images in real-world video sequences is presented. It combines fuzzy rule-based classification with a modified shading model method for changing region detection in outdoor environment. Proposed technique uses information extracted from video features to improve precision in classification results. It has been tested on sequences under various lighting conditions. Satisfactory and promising results have been achieved.

1. INTRODUCTION

A technique for classification and detection of building images is presented. It exploits information in video primitives to focus classification on features extracted from pixels belonging to static regions.

The technique is based on two simple observations: In static regions the possibility of finding features that match the pattern of "building" is higher because buildings are rigid and static objects; and misclassification can be reduced removing pixels belonging to changing regions. These regions can contain objects semantically different to building imagery but with feature distribution highly similar.

Existing approaches for classification of building images use a Bayesian framework to exploit image features by perceptual grouping [1], binary Bayesian hierarchical classifiers [2], or perform building semantic extraction using support vector machines [3].

Following, Section 2 presents the modified shading model. Section 3 gives a description of the building detection process. In Section 4 presents selected experimental results. Section 5 concludes the paper.

2. CHANGING REGION DETECTION

Existing algorithms for detection of changing regions in a video sequence usually do not resolve illumination changes inherent to exterior conditions [4]-[6].

Given an image sequence I consisting of N video frames, structural changes in a scene are detected using a window W of fixed size and position situated in I and a sliding mask A_i which performs scanning of W in each frame.

2.1. Shading Model Method

In [5] Skifstad and Jain use the ratio of pixel intensities in A_i between two frames to estimate the pixel variance σ_i^2 as follows:

$$\sigma_i^2 = \frac{1}{card\{A_i\}} \sum_{m \in A_i} \left(\frac{B_m}{C_m} - \mu_{A_i} \right)^2, \; i = 1..n, \qquad (1)$$

where B_m is a reference frame called *background* which does not contain changing regions, C_m is the current frame; B_m and C_m are pixel intensities within A_i, and μ_{A_i} is the mean of the pixel intensity ratio within A_i.

If $\sigma_i^2 \geq \varepsilon$ the center of A_i is marked as changing region, where ε is a suitable threshold.

Experiments have shown that for fast and large illumination changes this approach fails, i.e. some pixels are falsely assigned to changing regions.

2.3. Modified Shading Model Method

To overcome this shortcoming an improved and simplified version of the shading model method for changing region detection proposed in [7] is applied. This version is invariant to extreme illumination changes at pixel level. The pixel variance ${}^I\sigma_i^2$ is defined as

$${}^I\sigma_i^2 = \frac{1}{card\{A_i\}} \sum_{m \in A_i} \left(\frac{B_m}{C_m} K_i - median\{A_i\} \right)^2, \qquad (2)$$

where $median\{A_i\}$ estimates median of all pixel belonging to A_i and K_i is an adaptive coefficient to avoid falsely assignment of pixel to changing region for fast and large illumination changes. K_i is defined as

* The research leading to this paper was done within the framework of the Network of Excellence SCHEMA.

$$K_i = \frac{\sum_{m \in A_i} C_m}{\sum_{m \in A_i} C_{1m}} = \frac{\mu_i}{\mu_1}, \tag{3}$$

where C_{1m} is pixel intensity for the first frame in the sequence. If ${}^{I}\sigma_i^2 \geq \varepsilon$ the center of A_i is marked as changing region.

Changing region detection uses I_b to apply a filter on each pixel of current image I_m as follows

$$\hat{I}_m(x,y) = \begin{cases} 255 & \text{if } I_m^b(x,y) = 255 \\ I_m(x,y) & \text{otherwise} \end{cases}, \tag{4}$$

where $\hat{I}_m$ contains static regions only. In this way, changing regions are located and used to obtain a binary image I_b that contains white and black pixels to represent changing and static regions, respectively.

2.3. Comparative Analysis with Other Methods

While the shading model method in [5] is robust to illumination changes up to a certain point (roughly 10% of change) the coefficient k_i ensures sensitivity suppression far beyond mentioned level (up to 50%).

The modified shading model method for changing region detection was applied for analyzing video sequences containing about 49 monochrome images of 640×480 pixels with 256 gray levels (8 bits). The video rate is 25 frames/s and a window W of 5×60 pixels.

A sliding window A_i of 3×3 pixels was used for averaging. The larger windows have caused significantly greater execution times with negligible improvement. The optimal threshold was experimentally found to be $\varepsilon = 0.1$.

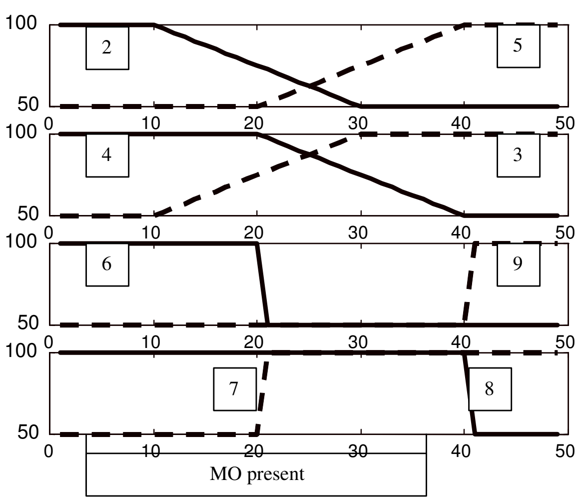


Fig. 1. Illumination changes detection (in %). The horizontal axe represents the ordinal number of the processed frame, in the range 1-49. The vertical axe shows the luminance changes.

In order to evaluate the modified method, several experiments with various illumination changes were performed. Selected results are depicted in Figure 1. For example, the dashed line 3 in the second diagram shows a luminance increase from 50% to 100% of the original. The changing region is detected between frame #11 and frame #30. Besides, a comparison with representative methods was performed. The obtained results are summarized in Table 1. Figure 2 illustrates the behavior of the mentioned methods under various illumination conditions for a selected experiment.

Table 1. Comparison between representative methods:
I – Modified shading model
II - Morphological edge detection [4]
III - Selective background updating [4]
IV - Change detection [6]
V – Shading model [5]
True positive (TP) and false positive (FP) correspond to percentage of correct detections and false alarms in W, respectively. P stands for Precision.

Result[%]	I	II	III	IV	V
TP	100	96.97	100	100	90.90
FP	0	81.25	75	81.25	56.25
P	100	54.41	57.14	55.17	61.77

The modified method always detected changing regions and there were no false alarms.

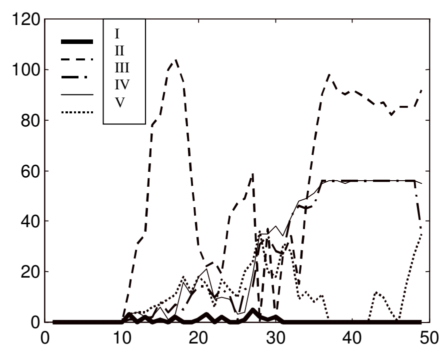


Fig. 2. Results for a selected experiment under different illumination conditions. The horizontal axe represents the ordinal number of the processed frame, in the range 1-49. The vertical axe shows the absolute difference of number of detected changing regions pixels in W between the normal sequence and sequence with the observed luminance change.

3. BUILDING DETECTION PROCESS

3.1. Building Image Classification

This work addresses the problem of classification of low level features extracted from video sequences for detecting and classifying building images.

Let $\mathbf{x} = (x_{11}, x_{12}, \ldots, x_{1\mathrm{N}}, x_{2\mathrm{N}}, \ldots, x_{\mathrm{MN}})$ be an image or video frame, $\mathrm{f} = \{\mathbf{f}^{(1)}, \ldots, \mathbf{f}^{(n)}\}$ be feature sets where f is a function of the image $\mathbf{x}$, $E^t = (e_1^t, \ldots, e_5^t)$ be a pattern extracted from a feature vector $\mathbf{f}^{(t)}$, and $Y = \{\mathrm{B}, \mathrm{NB}\}$ be a class set (B: Building, NB: Non Building). The building image classification problem is stated as learn a function

$$\mathrm{g}(\mathbf{x}) = E^t \mapsto Y, \tag{5}$$

where Y are symbols identifying classes and representing semantic interpretations of pattern E^t.

$\mathrm{g}(\mathbf{x})$ can be decomposed into K single-class specialized classifiers

$$\mathrm{g}_j(\mathbf{x}) = E^t \mapsto y_j, 1 \le j \le K \tag{6}$$

Subsequently, a fuzzy model is extracted from feature set $\mathbf{f}^{(t)}$ in order to approximate each function $\mathrm{g}_j(\mathbf{x})$ by a set $R_j = \{R_{j1}, \ldots, R_{j\mathrm{C}}\}$ of C if-then rules which has the following structure:

$$^{w}R_{jk} : \text{If } e_i^t \text{ is } A_i^k \text{ then } Y \text{ is } y_j \tag{7}$$

e_i^t is associated with a specific type of edge and A_i^k is a linguistic label used to transform values from a continuous to a discrete domain. Therefore, $\mathrm{g}_j(\mathbf{x})$ is summarized by a rule base R as follows:

$$\mathrm{g}(\mathbf{x}) \approx R = \bigvee_{j=1,k=1}^{\mathrm{M,C}} {}^{w}R_{jk} \tag{8}$$

where $w \in [0,1]$ is the weight of rule R_{jk}.

After several experiments, it was identified that most of the misclassifications of building images in real-world videos were due to objects with a similar edge distribution of buildings but semantically different. Most of them were moving non-rigid objects. Therefore, the technique was improved integrating a pre-processing method to detect changing regions.

3.2. Classification and Detection Process

The building image classifier uses a set of if-then rules and a fuzzy reasoning method. The rules have a number of antecedents and a consequent according to Equation 7.

The fuzzy reasoning method consists of three stages: The *Fuzzification stage* transforms feature values from a continuous to a discrete domain using membership functions. The *Inference stage* combines fuzzy sets and uses fuzzy rules to determine the class for these features. This combination requires a T-Norm operator. The importance of the rules is adjusted assigning weights to each one. If a class appears in more than one rule a T-Conorm operator is required. Finally, in the *Defuzzification stage* the result is transformed from fuzzy to real domain. Figure 3 summarizes the whole building classification and detection process.

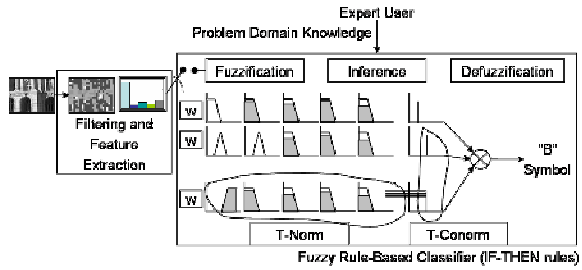


Fig. 3. Classification and Detection Process.

Figure 4 depicts the building image classification and detection process. The video frame is partitioned into 16 sub-images which are classified using the fuzzy classification process mentioned above. According to the number of sub-images classified as building, the whole image is or is not classified as "Building".

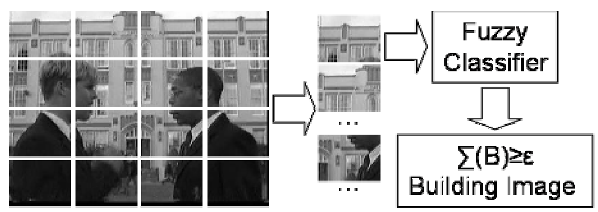


Fig. 4. Building Image Classification and Detection Process.

3.3. Improved Building Detection

The process results are improved using a pre-processing step to filter, select and extract suitable low level features. Changing regions are not considered to avoid misclassification as is shown in Figure 5.

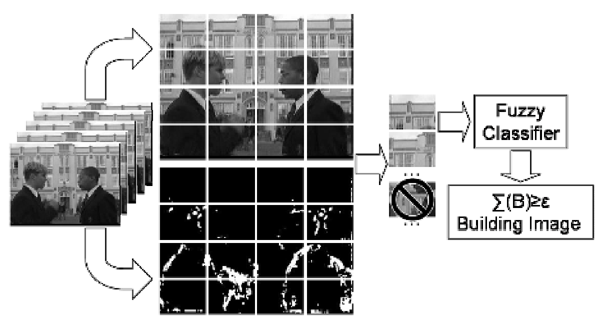
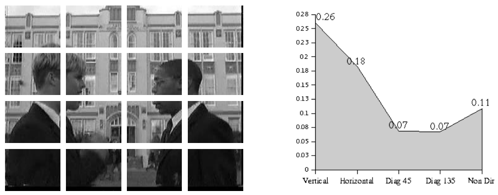


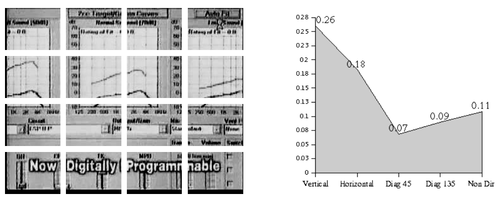
Fig. 5. Improved Building Image Classification and Detection Process.

As an example, frames 50389 and 10647 showed in Figures 6.a and 7.a, respectively, are classified as "building image". Looking at Figures 6.b and 7.b edge histograms of sub-images contributing in classification results have similar distribution. This shortcoming is sorted out using proposed method.



a b

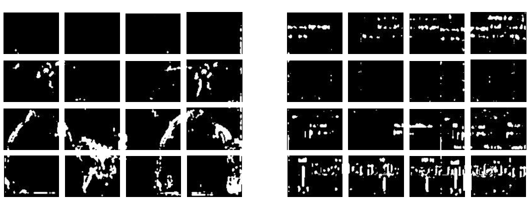
Fig. 6. Example of building image. a. is the video frame 50389. b. is the edge histogram of sub-image at row 1 and column 3.



a B

Fig. 7. Example of building misclassification. a. is the video frame 50389. b. is the edge histogram of sub-image at row 4 and column 3.

As is shown in Figures 8.a and 8.b changing regions are marked in the binary image with white pixels and the rest of it, i.e. the static background or foreground is marked with black pixels. This information was extracted from the corresponding video sequences.



a b

Fig. 8. Changing region detection using shading model method. Figure a and b depict binary images for video frames 50389 and 10647, respectively.

4. EXPERIMENTAL RESULTS

Building image classification was applied over 3000 frames from a set of videos randomly selected from TRECVID video repository. Features were extracted from key- and additional- frames at the shot level using an edge-based descriptor. A statistical performance evaluation based on the amount of imagery correctly classified (True Positive, TP) or misclassified (False Positive, FP) was conducted. The results are shown in Table 2. Precision is significatively improved combining changing region detection (CRD) based on shading model method with building image classification and detection (BICD).

Table 2. Precision results (P) [%]

Method	**TP**	**FP**	**P**
EBIC	77.70	21.96	77.97
BICD+CRD	89.54	10.12	89.84

5. CONCLUSIONS

A technique for improving classification of building images is presented. It exploits additional information extracted from a sequence of images and applies a fast, simple and effective shading model method for detecting changing regions.

The modified shading model method has been tested on real video sequences under various lighting conditions. It is insensitive to illumination changes, while the other methods seriously deteriorate. It always succeeds to detect changing regions under luminance variation.

Filtering out features extracted from pixels belonging to changing regions, misclassification is reduced and precision is significatively increased.

6. REFERENCES

[1] Q. Iqbal and J.K. Aggarwal, "Applying perceptual grouping to content-based image retrieval: Building image," *IEEE Int'l Conf. on Computer Vision and Pattern Recognition (CVPR99)*, vol. 1, pp. 42-48, Jun 1999.

[2] A. Vailaya, M.A.T. Figueiredo, A.K. Jain, and H-J. Zhang, "Image classification for content-based indexing," *IEEE Trans. on Image Processing*, vol. 10, no. 1, pp. 117-130, 2001.

[3] Y-N. Wang, L-B. Chen, and B-G. Hu, "Semantic extraction of the building images using support vector machines," in *Proc. IEEE Int'l Conf. on Machine Learning and Cybernetics,* Beijing, Nov 2002, vol. 3, pp. 1608-1613.

[4] M. Fathy and M.Y. Siyal, "An image detection technique based on morphological edge detection and background differencing for real-time traffic analysis," *Pattern Recognition Letters,* vol. 16, pp. 1321-1330, 1995.

[5] K. Skifstad and R. Jain, "Illumination independent change detection for real world image sequences," *Computer Vision, Graphics, and Image Processing,* vol. 46, pp. 387-399, 1989.

[6] G.L. Foresti and C. Regazzoni, "A change detection method for multiple object localization in real scenes," in *Proc. 20th Int'l Conf. on Industrial Electronics, Control and Instrumentation,* Sep 1994, vol. 2, pp. 984-987.

[7] Z. Zeljkovic, Z. Trpovski, and V. Senk, "Improved illumination independent moving object detection in real world video sequences," in *Proc. 4th EURASIP Conf. focused on Video/Image Processing and Multimedia Communications*, Jul 2003, vol. 2, pp. 473-476.