Evaluation of the Windowpane IPM

Jan Čech and Radim Šára
cechj@cmp.felk.cvut.cz
tel.: +420 2 2435 5726

Center for Machine Perception
Prague, Czech Republic

TN-eTRIMS-CMP-02-2007
version 1.0

May 2, 2007

Abstract

This document describes quantitative evaluation of the Windowpane Image Processing Module (IPM). There are ground-truth experiments, i.e. experiments where manually annotated images were available. We defined several error statistics, which evaluate the error rate, sensitivity and accuracy. We tested the method under conditions where the assumptions of orthographic rectification is violated. We also evaluate the quality of method's internal statistics with respect to ground-truth data.

1 Introduction

The method background is described in [2], where we show the benefits of using a structure model for segmentation. The Windowpane IPM which uses this method is described in [1], where there additional notes on incorporating the toolbox into the SCENIC system: places of possible feedback and measures of detection quality as a statistic for self-evaluation [3].

In this report, we use the structure model and the image model of all labels except for the facade label learnt the same as in [2], i.e. offline from an exemplar image. Since there is a large variability of the facade appearence, the image model of the facade labels, were estimated differently. The Gaussian distribution parametres were estimated (online) from the entire input image, regardless it contains also windowpanes. This is the simplest way to make the method adaptable to various facade colors, although the model is too wide and biased due to contamination by windowpanes. This is the first step in an intended EM-iterative scheme which would focus the model by re-estimating model parametres with respect to current detection results.

2 Error statistics

Manually annotated windowpanes as axis parallel rectangles is assumed as ground-truth detection. The output of the windowpane IPM is a list of axis parallel rectangles representing detected windowpanes. We define error statistics reflecting the (dis)agreement of the IPM detection and ground-truth data.

We define *False Negative* rate as a ratio of missing windowpanes, i.e. the windowpanes which were not detected by the IPM, but are in the ground-truth

$$FN = \frac{|T \setminus D|}{|T|}, \tag{1}$$

where T is the ground-truth detection, D is the detection by the IPM. Relation $T \setminus D$ is derived from the element equality relation as follows. Two elements $t \in T$ and $d \in D$ are assumed equal here if their overlap is higher than $\tau = 50$ percent.

Similarly, we define *False Positive* rate as a ratio of wrong detections, i.e. detections found by IPM which are not in the ground-truth

$$FP = \frac{|D \setminus T|}{|D|}. \tag{2}$$

The last statistic is *pixel accuracy*. This is a difference of the labeling map obtained by the IPM and the ground-truth labeling map normalized by the number of pixels in the image

$$a = \frac{|L - L_T|}{|L|}. \tag{3}$$

The labeling map is an image of the same size as the input image and each pixel have the value of the label identifier [2, 1].

3 Experimental Evaluation

We performed three experiments. First, we calculate the above error statistics on several facade images, Sec. 3.1. Then, we show how sensitive the method is to rectification condition violation, Sec. 3.2. Finally, we demonstrate qualities of internal evaluation statistic, Sec. 3.3.

3.1 Results of the evaluation

The above error statistics computed for IPM detection results are summarized in Fig. 1. There are both the error statistics 1(a) and values of their denominators 1(b). Corresponding images with detection results are shown in Fig. 2, 3.

We can see, the error varies from zero to about 60 for false negative rate, and from zero to about 40 percent for false positive rate. The accuracy varies from 1.5 to 7.7 percent. Note, that the worst case is the image number 7, 3. The

image i.d.	FN	FP	a	$\|T\|$	$\|D\|$	$\|L\|$
1	33.9	12.7	5.2	245	79	188324
2	10.7	24.8	3.3	140	101	220158
3	0.0	24.5	1.5	42	49	110565
4	23.4	9.1	5.9	47	44	102795
5	32.2	0.0	6.6	90	24	71142
6	4.4	6.7	3.9	68	60	88920
7	58.3	37.5	7.7	24	16	166635
8	33.3	0.0	6.8	60	36	162027
9	18.4	2.4	7.4	98	84	127608
10	11.4	13.4	4.8	70	67	114708

(a) Error statistics (in percent). (b) Denominator values.

Figure 1: Evaluation results.

reason is that its facade color is very close to the mean value of the windowpane color model. On the other hand, the detections are quite good in case of images 3, 6, 10. The reason is that the images are in accordance with the image and structure models [2]. The pixel error is below 10 percent for all images which is mostly caused by the correct structure model reflecting the reality by preferring facade labels to windowpane labels.

Computational time strongly depends on the difficulty of the scene, i.e. the agreement of the data with the model. In case of tested image, the time is below 5 sec per image, which is form 0.1 to 0.2 megapixels.

3.2 Sensitivity to improper rectification

We measured the error rate for image number 6 which was rotated by angle from zero to 45 degrees. Results are plotted in Fig. 4(a). Image with detection results under rotation angle 15 degrees is in Fig. 4(b).

We can see, the method is not much sensitive to precise rectification. The errors do not increase dramatically for small rotations, up to 15 degrees approximately, Fig. 4(a). However, we can see the windowpanes tend to split, Fig. 4(b), since with increasing rotation angle they are not axis parallel rectangles anymore, which is required by the model. Note that the error statistics are defined with $\tau = 50$ percent overlap of the ground-truth and detection results.

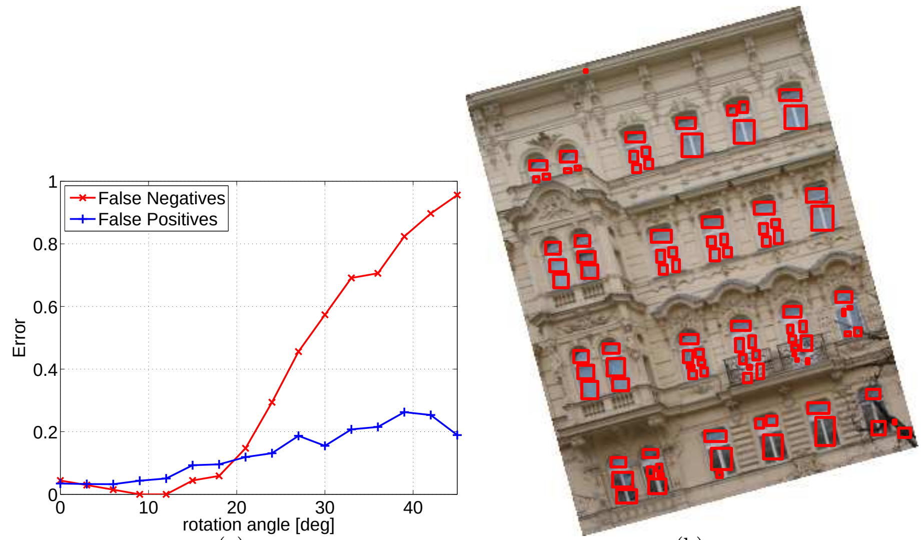
3.3 Quality of internal measures

Each detected windowpane has associated a confidence measure C_R which serves as a self-evaluation (internal) statistic [1]. We define the confidence

$$C_R = \frac{1}{|R|} \sum_{t \in R \subset I} g_t(x_t^*), \tag{4}$$

Figure 2: Windowpane detection results. Image i.d. per rows: 1, 2, 3, 4, 5, 6.

Figure 3: Windowpane detection results. Image i.d. per rows: 7, 8, 9, 10.



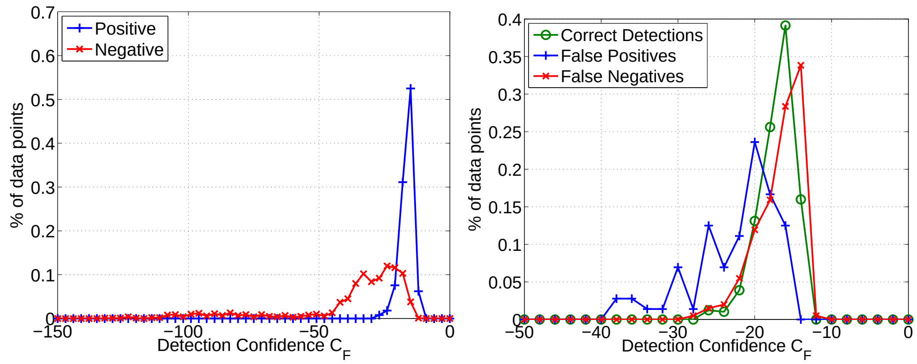
(a) (b)

Figure 4: Sensitivity to improper rectification. The error rate plots as a function of rotation angle (a), The image with detection under rotation angle 15 degrees (b).

where R is a rectangular region of the windowpane plus 1-pixel neighbourhood which belong to image I. Function $g_t(x_t^*)$ represents an agreement of the image pixels with the image model for the resulting labeling x_t^*. Based on the experiments, we decided to omit the agreement with the structure model $g_{tt'}(x_t^*, x_t'^*)$. The reason is that it had a low discriminability, i.e. it made a small margin between confidence of a truth windowpane region and wrong region.

Histogram of confidence for the ground-truth windowpane and randomly generated regions from input images is in 5(a). We used 884 ground-truth windowpane regions and 1000 randomly generated regions. We can see the distance between the two distribution is rather low, which means that the distance between the distributions of the facade image model and windowpane image model is small. The distributions are relatively flat, so further focusing and reestimating its parameters based on the detection results might improve the detection results and the confidence C_R.

Histograms in Fig. 5(b) show the confidence C_R for the detection results which were correct, detection results which were false positives and ground-truth windowpane regions which were not detected, false negatives. We can see, the distance of correct detections and false positives is again quite small. This is also caused that some of the false positive detection fulfill very well the image model and cannot be identified at that low-level of the system, see Fig. 2, 3. The histogram of correct detection are almost identical. The reason why the false negative regions were not detected is in most likely by the influence of the



(a) (b)

Figure 5: Histogram of confidence C_R for positive ground-truth windowpanes and negative examples (a). Histogram of confidence C_R for correct detections, false positive detections and false negative ground-truth detections (b).

surrounding regions in the image. The prior model overweighted the data in the global optimality criterion.

4 Conclusion

The conclusion of this performance evaluation is that we believe that the algorithm is hopefully usable in the SCENIC system. The error rates are not perfect, but further improvement can be made. The bottleneck here is the statistical image model. The idea is to develop an adaptable image model, which would adapt from own detection results in the EM-like steps as briefly discussed in Sec. 1. The convergence of such an iterative process is unclear and it might be slow, but it should improve the quality of the detection results.

The internal statistic C_R is perhaps usable in spite of having a relative low discriminability. We could additionally increase it by adding some extra features and construct a new confidence statistic. Straightforward way is to incorporate a size or aspect ratio of the detected windowpane, or some quality of detected windowpanes alignment. But, this not the low-level knowledge. The question is whether we should use it in the low-level IPM.

References

[1] J. Čech and R. Šára. Specification of windowpane IPM for use in SCENIC. Technical Report TN-eTRIMS-CMP-02-2006, 2006.

[2] J. Čech and R. Šára. Windowpane detection based on maximum aposteriori labeling. Technical Report TR-CMP-2007-10, Center for Machine Per-

ception, K13133 FEE Czech Technical University, Prague, Czech Republic, 2007.

[3] Wolfgang Förstner. Performance characteristics for classification and learning. Hamburg meeting talk, March 2007.