SYMBOL RETRIEVAL BY IDENTIFYING REPEATING PATTERNS

Nayef N., Breuel T. -- Tech. University of Kaiserslautern
{nnayef , tmb}@iupr.com

IUPR FACHBEREICH INFORMATIK

Abstract

Content analysis of images is essential for search engines and retrieval applications.

This work presents a method for content analysis in technical line drawings.

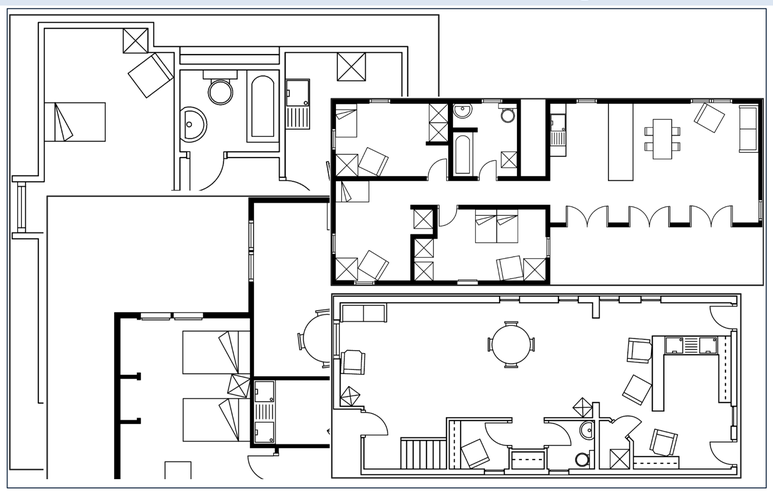
This is achieved by:
1) Identifying patterns by statistical grouping
2) Clustering the patterns using geometric matching.

The clusters form a **symbol library** to be used in symbol retrieval.

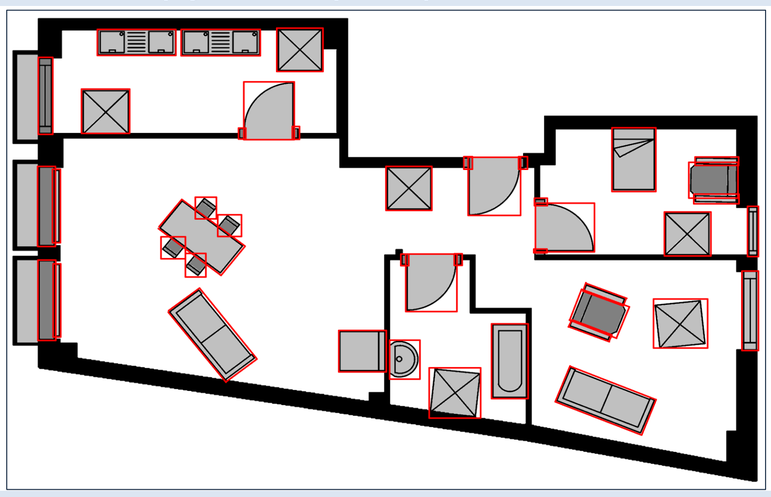
The results on a standard dataset of architectural drawings are significantly better than all the results reported so far.

What does it do ?

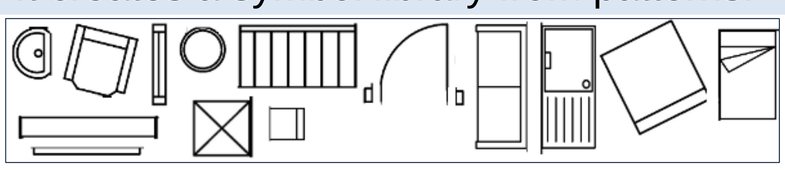
From a collection of line drawings:



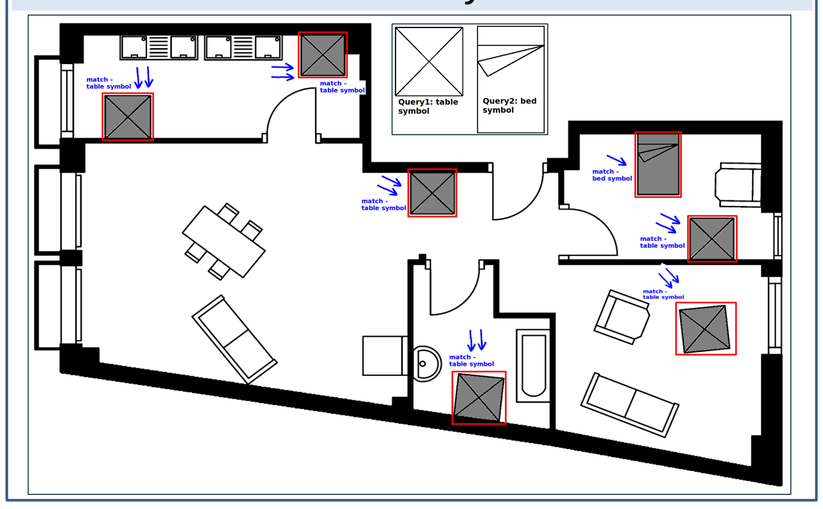
It finds –in each drawing- meaningful patterns (symbols parts):



It creates a symbol library from patterns:



And uses it to retrieve symbols:



Grouping: Identify Patterns

1) Preprocessing**:** edge detection, then sampling **segments** along the edges.
2) Find convex groups from the edges using Jacobs' algorithm [1].
3) Clean up found groups (trivial, subsets, cyclic permutations).
4) Keep the outer groups ⟹ **patterns**.

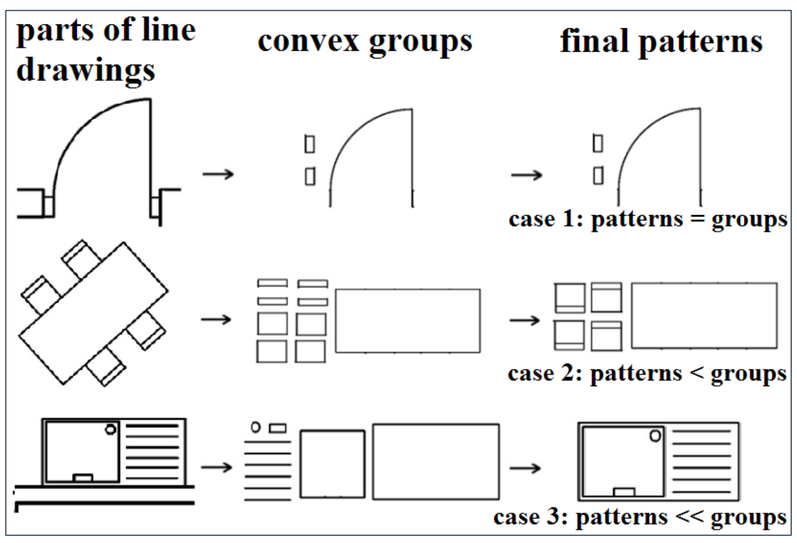


Fig: how grouping works on 3 different cases

A "Pattern" is:

"The set of all line segments that are inside a group including the segments that constitute the group itself".

Clustering: Repeating Patterns

Similar patterns can be rotated / scaled.
Match each pattern with all other patterns:
- Use geometric matching framework [2].

Build a cluster out of the current pattern and all its matching patterns.
- Do not consider the matched patterns in building other clusters.

The clusters form a symbol library.
- Compact representation of the dataset.

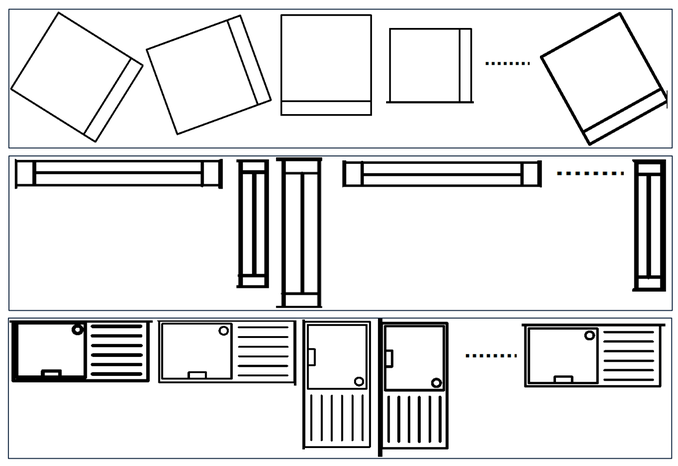


Fig: sample clusters

Retrieval

Match [2] a query symbol against the symbol library.
Retrieve all symbols from the matching cluster.

Results

Dataset:
- 300 images of complete floor plans.
- Public [3] – used in GREC-2011 symbol spotting contest, and by researchers [4].

Metrics:
- Clustering / Retrieval: Recall & Precision.
- Grouping: Recall for non-missing symbols, Precision for the relevant found patterns.

Our detailed work is in [5,6].

	Ground Truth	Results
No. of Patterns	12513	13780
No. of Clusters	25	30
Grouping Evaluation (offline)		
Covered Symbols	Recall (M=6987)	Precision (N=13780)
	98.8%	97.31%
Avg. time per image		22.75 sec.
Clustering Evaluation (offline)		
All Clusters	Avg. Recall	Avg. Precision
	95%	96.5%
Avg. time per forming 1 cluster		45 min.
Retrieval Evaluation (online)		
No. of Queries	Avg. Recall	Avg. Precision
7 (a subset)	99%	98%
Avg. time per query from 300 images. depends on # clusters not # images)		3 min.

Take Home Messages

1) Symbol retrieval in large databases is made practical.
- With very high accuracy.

2) Finding symbols patterns from the background has converted recognition-in-context to isolated recognition.
- Using statistically justified methods [1].

3) Geometric matching is optimal for clustering shapes under similarity transformations.
- Based on well founded theories [2].

References

[1] D. W. Jacobs, "Robust and efficient detection of salient convex groups," PAMI, vol. 18, no. 1, pp. 23–37, 1996.

[2] T. M. Breuel, "Implementation techniques for geometric branch-and-bound matching methods," Computer Vision and Image Understanding (CVIU), vol. 90, no. 3, pp. 258–294, 2003.

[3] http://mathieu.delalandre.free.fr/projects/sesyd/index.html

[4] M. M. Luqman, T. Brouard, J. Ramel, and J. Llodos, "A content spotting system for line drawing graphic document images," in ICPR, 2010, pp. 3420–3423.

[5] N. Nayef and T. M. Breuel, "Statistical grouping for segmenting symbols parts from line drawings, with application to symbol spotting," accepted in ICDAR, 2011.

[6] N. Nayef and T. M. Breuel, " Creating Symbol Libraries From Technical Drawings By Identifying Repeating Patterns", accepted in GREC, 2011.

Contact: Nibal Nayef (nnayef@iupr.com) **https://sites.google.com/a/iupr.com/nayef**