Use of a Large Image Repository to Enhance Domain Dataset for Flyer Classification

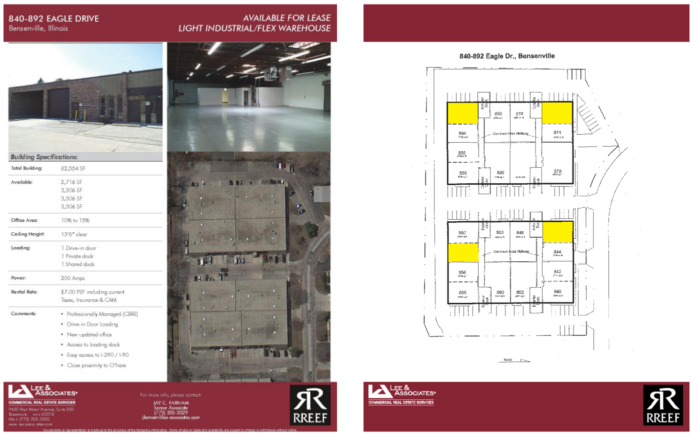
Payam Pourashraf and Noriko Tomuro

DePaul University, 243 S. Wabash Ave, Chicago, IL 60604 USA
ppourash@cdm.depaul.edu, tomuro@cs.depaul.edu

Abstract. This paper describes our exploratory work on supplementing our dataset of images extracted from real estate flyers with images from a large general image repository to enhance the breadth of the samples and create a classification model which would perform well for totally unseen, new instances. We selected some images from the Scene UNderstanding (SUN) database which are annotated with the scene categories that seem to match with our flyer images, and added them to our flyer dataset. We ran a series of experiments with various configurations of flyer vs. SUN data mix. The results showed that the classification models trained with a mixture of SUN and flyer images produced comparable accuracies as the models trained solely with flyer images. This suggests that we were able to create a model which is scalable to unseen, new data without sacrificing the accuracy of the data at hand.

1 Introduction

Flyers are a popular form of advertising material, intended to be distributed to a wide audience. Flyers are often multimodal – description of the subject matter in text is accompanied by related images. For example, a flyer of a commercial real estate typically indicates the listing information such as address, square footage, price, property type (industrial, office, etc.) and broker's name in text, and includes some pictures of the property. Figure 1 shows an example flyer of an industrial property. In recent years, with the wide spread of desktop publishing tools and the prevalence of the internet, flyers have become more popular, especially in the electronic file format. For example, brokers of commercial real estate create a flyer for each property they sell in the (commonly) pdf format and post it on their website; they also collect flyers of other brokers or in the public domain (again commonly in the pdf format) to create a large, online searchable database of available properties to attract customers. However, manually searching and indexing flyers in building a database is a tedious, time-consuming task -- a better solution would be to automatically extract the relevant information.



840-892 EAGLE DRIVE
Bensenville, Illinois

AVAILABLE FOR LEASE
LIGHT INDUSTRIAL/FLEX WAREHOUSE

Building Specifications:

Total Building:	82,554 SF
Available:	2,716 SF 3,306 SF 3,306 SF 3,306 SF
Office Area:	10% to 15%
Ceiling Height:	13'8" clear
Loading:	1 Drive-in door 1 Private dock 1 Shared dock
Power:	200 Amps
Rental Rate:	$7.00 PSF including current Taxes, Insurance & CAM
Comments:	• Professionally Managed (CBRE) • Drive-in Door Loading • New updated office • Access to loading dock • Easy access to I-290 / I-90 • Close proximity to O'hare

LEE & ASSOCIATES
COMMERCIAL REAL ESTATE SERVICES
9450 Bryn Mawr Avenue, Suite 550
Rosemont, Illinois 60018
Main (773) 355-3000
www.lee-associates.com

For more info, please contact:
JAY C. FARNAM
Senior Associate
(773) 355-3029
jfarnam@lee-associates.com

RREEF

No warranty or representation is made as to the accuracy of the foregoing information. Terms of sale or lease and availability are subject to change or withdrawal without notice.

840-892 Eagle Dr., Bensenville

Fig. 1. A commercial real estate flyer of an industrial property (© Lee & Associates)

In this paper, we describe our work on classifying images embedded in real estate flyers by the property types (retail, office, industrial and multi-family). It is a part of a larger project which aims to develop a high-performance multimodal system which extracts information from real estate flyers by using both texts and images. In this paper, we focus on the image classification component. In the full system, the results of the image classifier will be combined with the textual information to extract various information from a flyer regarding the property and listing.

Recently, there has been a lot of effort, in both research and industry, on how to mine from multimodal or multimedia data and utilize the extracted content for various purposes [1-3]. However, there are several major difficulties in the pursuit. First, automatic media/image extraction/cropping tools (from various file formats) are generally not very accurate. Not only are they unable to differentiate irrelevant graphic elements such as banners and border strips from real content images, most tools also have difficulty with images which are overlapping or laid out in an unusual or creative way. In fact, the latter situation occurs commonly with flyers because they are essentially free-form. Consequently the 'ratio of successful automatic harvesting' of good, clean content images from embedded files is usually only moderate. For example, the dataset we used in our previous work showed about 30% of the automatically cropped images were fragments, and around 15-20% were failures (where typically an original file with multiple embedded images is returned without cropping). This will become a problem when one wants to collect sample images of many categories. The second major difficulty is the problem of semantic gap. It refers to the difficulty in image analysis in relating the low-level pixel data with a high-level concept which the image represents (e.g. 'sunset') [4, 5]. In our case, the gap is particularly deep because of the subtle similarity or overlaps between the property types. For example, retail and industrial spaces often have an office (room) inside, while an office may have a storage room which looks like a small warehouse.

The work we present in this paper is an attempt to address the first problem (of obtaining clean content images), by supplementing the images extracted from the flyers with images from a large publicly available image repository, in particular the Scene UNderstanding (SUN) database[1] [6]. The SUN database contains over 130,000 images, which are annotated with over 900 scene categories. Also the scene categories are organized in a hierarchy. We selected several scene categories from SUN which seem to match with our flyer images, and added the images of those categories in our dataset. Actually, enhancing the dataset would also help narrow the semantic gap as well in our case, because generally a real estate property consists of spaces of several types (for different usage) and our flyer dataset did not have enough samples of various space types (partly due to the low cropping accuracy).

However, there is a critical assumption in adding SUN images in our dataset, that SUN images are similar to our flyer images. Since SUN is essentially an external source, mixing in with our primary dataset, which is from a particular domain of real estate flyers, might not bring improvement in classification. To test the assumption and see the usefulness of the SUN dataset in our research, we added a selected subset of the SUN images to our flyer image dataset and ran a series of experiments. The results showed that the classification models trained with a mixture of SUN and flyer images produced comparable accuracies as the models trained solely on flyer images. This is a promising result, which indicates that we would be able to utilize the SUN database to boost our data and create a larger training dataset without sacrificing classification accuracy.

2 Related Works

Preliminary works have been done on text and image independently. In one of the previous works for the text side various techniques in Information Extraction and Text Categorization has been used to do the task. The combination of textual (e.g. token and token kind) and visual features (e.g. font color, size, position in the flyer) were used to extract various information about the property, such as property type (e.g. retail, industrial, office, land), address, space size (square footage, acres), and the name and contact information of the broker [7]. The visual features which have been used in that work included: font size and Y coordinate [7].

In a recent work on the image side, the images embedded in real estate flyers were classified into five genres (map, schematic drawing, aerial photo, indoor-building and outdoor-building) [8]. At the start of this experiment, the features were extracted from over 3000 images from publicly available online real estate flyers, including Autocorrelogram, Tamura, Local Binary Patterns, Histogram of Oriented Gradients, number of lines (by using Hough Transform) and the number of points with high cornerness (by using Harris corner detection). A two-level ensemble classifier model was built in which the first (Tier-1) consisted of several binary classifiers, each of which was trained to classify data for a given genre, and the second (Tier-2) classifier combined the output of the Tier-1 classifiers to produce the final output. The result showed that

[1] http://groups.csail.mit.edu/vision/SUN/

the model has a significant out performance in comparison to the baseline the classifiers (Naïve Bayes, Decision Tree and KNN) [8].

The SUN database from which we have selected the images has been widely used as a benchmark dataset for different tasks such as scene understanding [9] and object recognition [10]. SUN-397 database is a subset of this dataset which has also been used as a benchmark dataset [6, 11]. In [12] the authors selected over 500 images from the SUN categories of "bedroom" and "living room" for the task of 3D model scene understanding.

3 Methodology and Experimental Design

3.1 Image Dataset

In this work, we used the real estate flyer dataset used in [7]. From the original dataset we selected 144 flyers that included indoor images. We focused on indoor images in this work, because we thought indoor images were more suggestive of the property types than other kinds of images (such as maps and outside images). Then from each flyer, we extracted images by using software tools[2] and wrote our own code to filter 'noisy' non-content images (such as image fragments, color borders and company logos). Finally, we got 686 indoor images out of the original flyers. The distribution of the number of flyers and their proportion in the dataset are shown in Table 1.

Table 1. Distribution of the number of flyers and their proportion in the dataset

Property Type	Number of Flyers	Proportion of Flyers	Number of Images	Proportion of Images
Industrial	37	21%	143	26%
Multi-family	7	8%	60	5%
Office	58	41%	280	40%
Retail	42	30%	203	29%
Total	144	100%	686	100%

We supplemented the flyer images with images from the SUN database. SUN has 3 categories in its' third level of hierarchy, which are indoor, outdoor man-made, and outdoor natural. We selected some categories from the SUN indoor categories which seem to match with our flyer images. We tagged each of the selected categories by one property type (industrial, multi-family, office and retail). In order to have a balanced

[2] We used pdftohtml (http://sourceforge.net/projects/pdftohtml/) to convert pdf to html, and Gimp (http://www.gimp.org/) to crop individual images.

dataset, we randomly selected the images inside each property type and made the same number of images for each of them. The total number of images was 5140 (1285 for each of the four property type). Selected SUN categories with their assigned property types are shown in Table 2.

Table 2. Selected SUN categories

Property Type	Number of Images	SUN Categories
Industrial	1285	Basement, Assembly line, Furnace room, Parking garage, Storage room, Warehouse
Multi-family	1285	Kitchen, Bedroom
Retail	1285	Cafeteria, Coffee shop, Lobby, Restaurant, Shopping mall
Office	1285	Computer room, Conference room, Cubicle, Office, Office cubicles

Figures (2, 3) show examples of the indoor-building genre from our flyers and example images of the SUN indoor categories, respectively.

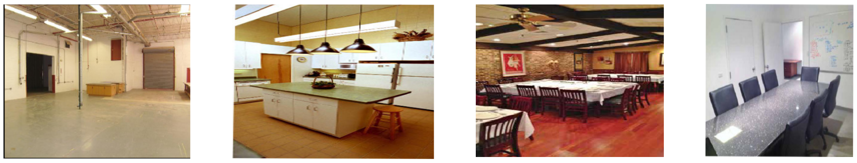


Fig. 2. Example images of the indoor-building genre from our flyers

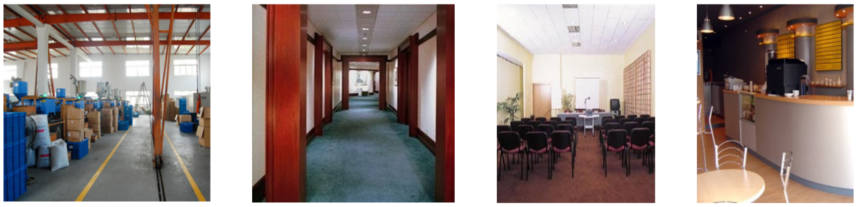


Fig. 3. Example images of the SUN indoor categories

3.2 Feature Extraction

The indoor-building genre images of our work and the selected list of categories from SUN database have then been incorporated to extract the GIST and color features. The details of the extracted features are described as follows:

GIST: Some prior studies [13, 14] have proposed that the scene recognition is initiated from the getting of the global configuration of the scene. The task Scene recognition can be done by looking at their GIST. Thus, the general 512 dimensions' GIST [14, 15] feature has been extracted from the images.

Color features: For color features, we have extracted 6 color moments (the 2 first color moments from each R,G,B) and 32 color Autocorrelogram [16] of 1 and 3 distances. Finally we got 38 color features.

3.3 Experimental Setup

To see the effect of SUN images mixed in with the flyer images to build classification models, we ran several experiments with varying degrees/proportions of SUN vs. flyer mixes in the training set. In particular, we added SUN images in the training set gradually with a 20% increment, that is, all of the flyer images + 0% of SUN images, + 20% of SUN, + 40% of SUN, and so on until (all of flyer images) + 100% of SUN – thereby a total of 6 mixture configurations. We also wanted to see the effect of color features (in addition to the GIST features) in the classification for our problem – thus two feature configurations: GIST only or GIST + Color. Putting all these together, we had a total of 12 (= 6 mixture * 2 feature) configurations. Table 3 below summarizes the configurations and provides the breakdown of the training set for each configuration. Note the selection of the SUN images in each configuration was random but stratified (evenly across the property types, because the same number of images were selected from the SUN database for each property type).

Table 3. Various configurations of the training dataset

SUN Mix Image Features	0%	20%	40%	60%	80%	100%
GIST only (512 features)	686 flyer images + 0 SUN images	686 flyer images + 1028 SUN images	686 flyer images + 2056 SUN images	686 flyer images + 3084 SUN images	686 flyer images + 4112 SUN images	686 flyer images + 5140 SUN images
GIST + Color (550 features)						

To build classification models, we used Support Vector Machine (SVM) with a linear kernel. We chose SVM because it has been shown to produce good results in many works (including in several of our previous work).

4 Results and Discussion

To evaluate the classification models, we used (5-fold) cross validation (CV), with two data partition schemes: one by images (but stratified with respect to property types, which is the target category) and another by flyers (from which the images were extracted). We thought of partitioning at the flyer level because it would better simulate the realistic situations - a trained model receives images from a new (totally unseen) flyer.

The accuracy results for various experimental configurations are demonstrated in Table 4 and Figure 4. Note that cross validation was done for flyer images only, not for SUN images. For each scheme (1 through 4), for each fold the training was done using all of the SUN images (always), plus a 4/5 of the flyer images partitioned based on the images/property types or the original flyers from which the images came from, and the testing was done using the remaining 1/5 of the flyer images. Also the p-values were computed between the 0% SUN vs. each of the SUN mix for each scheme.

Table 4. Accuracy results for various experimental configurations

Scheme	CV Partition	SUN Mix / Image Features	0%	20%	40%	60%	80%	100%
1	Image level	GIST only (p-value)	51.45 (--)	37.75 (.009)	32.78 (.002)	31.77 (.001)	30.01 (.002)	30.89 (.001)
2		GIST + Color (p-value)	52.62 (--)	51.73 (.663)	48.97 (.177)	47.65 (.111)	47.08 (.032)	47.07 (.010)
3	Flyer level	GIST only (p-value)	43.86 (--)	41.94 (.568)	43.91 (.989)	42.31 (.663)	43.15 (.863)	42.45 (.758)
4		GIST + Color (p-value)	45.68 (--)	45.66 (.995)	42.50 (.290)	41.15 (.315)	42.24 (.366)	42.94 (.514)

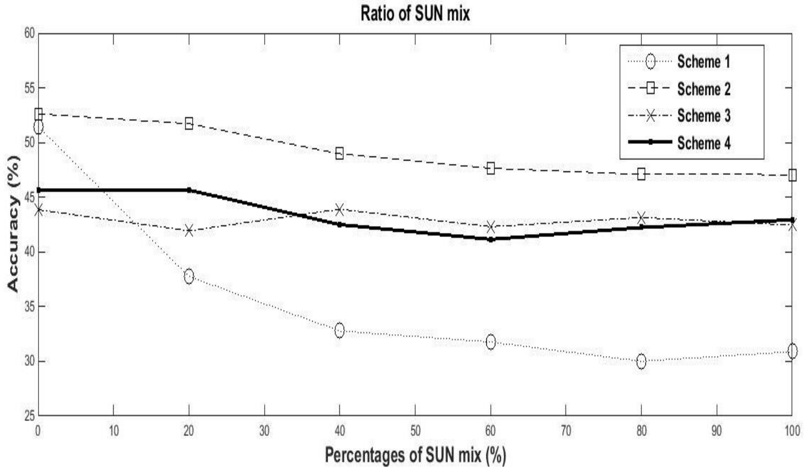


Fig. 4. Accuracies of various experimental schemes

For all schemes, accuracy went down as more SUN images were added in the training set (except for Scheme 3: accuracy went up slightly for 40% SUN mix). It was an expected result because SUN images were essentially 'out-of-domain' data from the perspective of flyer images. Also the decrease was more drastic when the test data was partitioned on the image level (Scheme 1 and 2) – the decrease was statistically significant (p-value < .05 for all non-zero SUN mixes for Scheme 1, and 80 and 100% SUN mixes for Scheme 2). A notable result is that, for Scheme 3 and 4, accuracy didn't go down significantly, or yet stayed about the same, as more SUN images were added in the training set (as evidenced by the p-values > .05 for all non-zero SUN mixes). This is an encouraging result -- Since Scheme 3 and 4 used the partition based on flyers (and simulate realistic and predictive situations), maintaining approximately the same level of accuracy as more SUN data was injected means that these models are scalable to unseen, new data.

As for the effect of color features, there was a large effect in the classification accuracy when the test data was partitioned at the image level (Scheme 1 vs. 2), but not when it was partitioned at the flyer level (Scheme 3 vs. 4).

5 Conclusions and Future Work

In conclusion, in this work we presented our work on classifying the indoor images embedded in the real estate flyers by the property type. The proposed model could be scaled to new data without compromising the accuracy of data classification.

For future work, we plan to bridge the problem of semantic gap by building a multimodal system using both texts and images. Other future studies with focus on the other image genres besides indoor-building (map, schematic drawing, aerial photo and outdoor-building) would also be helpful for improvement of flyers classification.

References

1. Manjunath, T. N., Hegadi R. S., Ravikumar G. K.: A Survey on Multimedia Data Mining and Its Relevance Today. IJCSNS 10, no. 11, 165-170 (2010)
2. Bhatt, Chidansh Amitkumar, and Mohan S. Kankanhalli.: Multimedia data mining: state of the art and challenges. Multimedia Tools and Applications 51, no. 1, 35-76 (2011)
3. Guillaumin, Matthieu, Jakob Verbeek, and Cordelia Schmid. Multimodal semi-supervised learning for image classification. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pp. 902-909 (2010)
4. Dorai, C., Venkatesh, S.: Bridging the semantic gap with computational media aesthetics. IEEE multimedia 10, no. 2, 15-17 (2003)
5. Zhao, R., Grosky W. I.: Bridging the semantic gap in image retrieval. Distributed multimedia databases: Techniques and applications, 14-36 (2002)
6. Xiao, J., Ehinger, K. A., Hays, J., Torralba, A., Oliva, A.: SUN database: Exploring a large collection of scene categories. International Journal of Computer Vision, 1-20 (2014)
7. Apostolova, E., Tomuro, N.: Combining Visual and Textual Features for Information Extraction from Online Flyers. Empirical Methods in Natural Language Processing (EMNLP), (2014)
8. Pourashraf, P., Tomuro, N., Apostolova, E.: Genre-based image classification using ensemble learning for online flyers. In Seventh International Conference on Digital Image Processing (ICDIP15), International Society for Optics and Photonics, (2015)

9. Li, C., Parikh, D., Chen, T.: Automatic discovery of groups of objects for scene understanding. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2735-2742 (2012)
10. Santiago, M., Guillaumin, M., Van Gool, L.: Prime Object Proposals with Randomized Prim's Algorithm. In Computer Vision (ICCV), 2013 IEEE International Conference on, 2536-2543 (2013)
11. Su, Y., Jurie, F.: Improving image classification using semantic attributes. International journal of computer vision 100, no. 1, 59-77 (2012)
12. Scott, S., Lin, J., Hebert, M.: Data-driven scene understanding from 3D models. In BMVC, (2012)
13. Irving, B.: Aspects and extensions of a theory of human image understanding. Computational processes in human vision: An interdisciplinary perspective, 370-428 (1998)
14. Khosla, A., K., Das Sarma, A., Hamid, R.: What makes an image popular?. In Proceedings of the 23rd international conference on World wide web, 867-876 (2014)
15. Oliva, A., Torralba, A..: Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of computer vision42, no. 3, 145-175 (2001)
16. Huang, J., Kumar, S. R., Mitra, M., Zhu, W., Zabih, R.: Image indexing using color correlograms. In Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, 762-768 (1997)