Using Outdoor Images for Flyer Classification

Payam Pourashraf and Noriko Tomuro
DePaul University, School of Computer Science, Chicago, IL
ppourash@depaul.edu
tomuro@cs.depaul.edu

Abstract

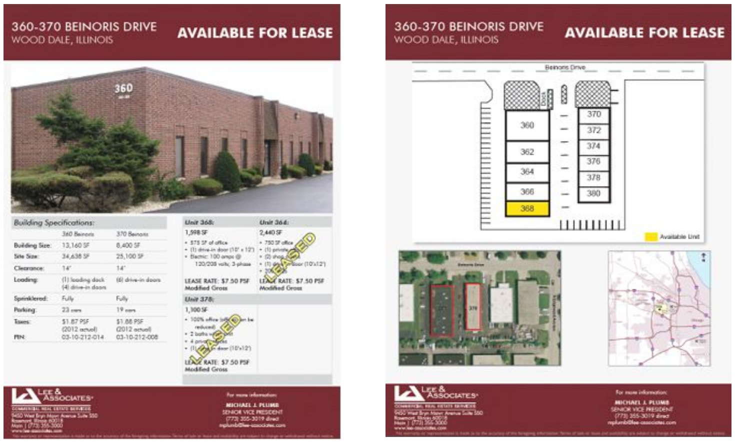
The goal of this paper was to create a new method for analyzing the online real estate flyers based on their property types. We created an algorithm which identifies the buildings and windows from the buildings in order to extract some useful features for classifying the flyers. Our novel approach for building recognition has two main steps: 1- Building Detector 2- Region Growing. Our novel window detection algorithm uses vanishing point to identify nearly the best angle for applying window detection. It transforms the 2D image into 3D and rotates the 3D image around the z-axis and pick the appropriate angle based on the vanishing points. Using these two novel techniques we were be able to extract a new feature vector which is used to build our final model. This final model is able to classify Retail spaces very well based on the Window and logo features.

Keywords: Flyer, Window Detection, Building Recognition, outdoor images, object detection.

1 Introduction

In a number of commercial industries such as real estate, marketing materials are the source of information. Brokers of commercial real estate have a collection of properties which they sell, and they make a flyer for each of them. These flyers have all relevant listing information to market the property. Nowadays brokers also gather information on other available properties from other brokers or public flyers. To attract customers, they build searchable databases. However, it is tiresome and error-prone to extract the relevant information out of a flyer and manually enter the data in a database. Automatic extraction and indexing the flyers is much easier and less prone to errors [2, 3].

A flyer is a mixed type of document containing textual descriptions and some images. The two modalities have a complementary role on conveying information. Texts explicitly provide relevant information by words, while images implicitly provide additional relevant information through visual representations. An example of real estate flyer is shown in figure 1.



360-370 BEINORIS DRIVE
WOOD DALE, ILLINOIS

AVAILABLE FOR LEASE

360

Building Specifications:

	360 Beinoris	370 Beinoris
Building Size:	13,160 SF	8,400 SF
Site Size:	34,638 SF	25,100 SF
Clearance:	14'	14'
Loading:	(1) loading dock (4) drive-in doors	(6) drive-in doors
Sprinklered:	Fully	Fully
Parking:	23 cars	19 cars
Taxes:	$1.87 PSF (2012 actual)	$1.88 PSF (2012 actual)
PIN:	03-10-212-014	03-10-212-008

Unit 368:
1,598 SF
- 575 SF of office
- (1) drive-in door (10' x 12')
- Electric: 100 amps @ 120/208 volts; 3-phase

LEASE RATE: $7.50 PSF
Modified Gross

Unit 364:
2,440 SF
- 750 SF office
- (1) private
- (2) shop
- (1) door (10'x12')

LEASED

LE RATE: $7.50 PSF
Modified Gross

Unit 378:
1,100 SF
- 100% office (off be reduced)
- 2 baths
- 4 priv
- (1) door (10'x12')

LEASED

LE RATE: $7.50 PSF
Modified Gross

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

For more information:
MICHAEL J. PLUMB
SENIOR VICE PRESIDENT
(773) 355-3019 direct
mplumb@lee-associates.com

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

A major problem in flyer classification is the problem of semantic gap between the low-level visual and textual features and the high-level labels of property type (e.g. retail, industrial, office; labels commonly used in the real estate industry). The problem is that property types are essentially based on the usage of the property, rather than the attributes or objects that are described in the text or captured in the images. Furthermore, since a flyer might include several images, the determination of the final property type by an automatic system is non-trivial. These factors together are making the relation between the two levels extremely complex and the semantic gap much more pronounced in the domain of real estate flyers.

The work we present in this paper is an attempt to reduce the semantic gap and improve the flyers classification, by focusing on images of Outdoor-building. Outdoor images by capturing the external features such as window properties could illustrate with the usage of the property. Thus focusing on the outdoor image genre has been chosen as a current step in the classification of the flyers, based on the image genres. The results of this work can improve the classification of flyers which has a practical import for the brokers. Another important benefit of this work is to reduce the semantic gap which has been an obstacle for modeling of this system.
The hypothesis of the current work is that focusing on the outdoor-building genre and extracting physical features from the images, such as window features, car and logo features, could help in the classification of flyers and help to reduce the semantic gap.

2 Related Works

Preliminary works have been done on text and image independently. In one of the previous works for the for the text side various techniques in IE and Text Categorization has been used. 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 accurately 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 [3]. The visual features which have been used on that work included: font size and Y coordinate. The SVM classifiers were used for the task of identifying 12 types of named entities, included: broker name, city and neighborhood. The results showed that overall visual features improved performance significantly [3].

In a 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) [2]. 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 the model has a significant out performance in comparison to the baseline the classifiers (Naïve Bayes, Decision Tree and KNN) [2].

Some approaches have been proposed for learning from multimodal data. In [4] they showed that utilizing tags with image features (such as Spatial Color Mode, MPEG-7 Edge Histogram, MPEG-7 Edge Histogram) could boost the classification performance of SVM model. In [5], an approach has presented that on different object recognition tasks, by adding the text modality the accuracy of SVM classifier can be boosted. In [6], in order to learn a model with multiple input modalities, a Deep Boltzmann Machine has been proposed.

As mentioned above, the independent previous works on the text side have obtained fairly good promising results but to problem of semantic gap the current work has been designed to improve the classification models. However, in going forward, the current system has been designed to improve the accuracy of classification of the flyers just based on the image side.

3 Methodology and Experimental Design

3.1 Image Dataset

In this work, we used the real estate flyer dataset used in [2]. From the original dataset we randomly selected 144 flyers that included outdoor images. We focused on outdoor images in this work, because we thought outdoor images were more suggestive of the property types specially the Retail than other kinds of images (such as maps and inside images). Then from each flyer, we extracted images by using software tools[1] and wrote our own code to filter 'noisy' non-content images (such as image fragments, color borders and company logos).

We put together an analysis that categorizes flyers by their property types. The flyers could have single label (Industrial, Office, ...) or multi-label (Industrial-Office, Industrial-Land). Below you can see the what results were:

Table 1: Distribution of different genres with their number per property type. I = industrial, O =Office, L = Land, M = Multi-family, R = Retail. I-O = Flyers with labels Industrial and Office.

	I	I-O	I-L	I-M	L	L-O	L-R	O	O-R	R	M
Aerial	107	25	3		115	2	2	60	23	110	3
Map	135	36	3		337	2	2	160	63	266	57
Inside	130	20			2			287	36	193	21
Outside	265	**50**		2	45		8	253	**85**	398	68

[1] https://www.gimp.org/

As can be seen there are considerable amounts of images which belongs to Office-Retail and Industrial-Office categories. This makes it difficult for the algorithms to differentiate between the Retail, Office and Industrial categories. To solve this, our hypothesis is that window size can be used by the algorithm to differentiate between retail, industrial and office. This is because retail buildings typically have large windows and office buildings have smaller frequent windows and industrial has smaller but less frequent windows. For example, in the figure 2 we can tell its retail because the windows size is large compare to the building size and also there are few windows. In figure 3 this is industrial building because the windows are small and infrequent. Finally, the figure 4 is Office since the windows are medium size and frequent.

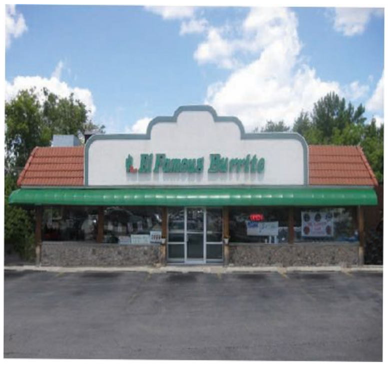


Figure 2: An example of a retail property

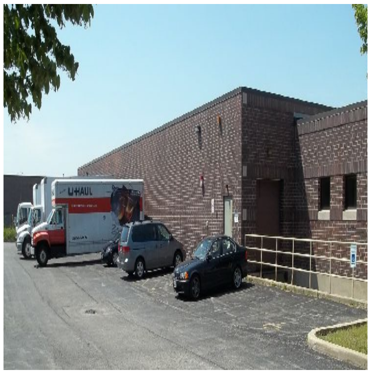


Figure 3: An example of industrial property

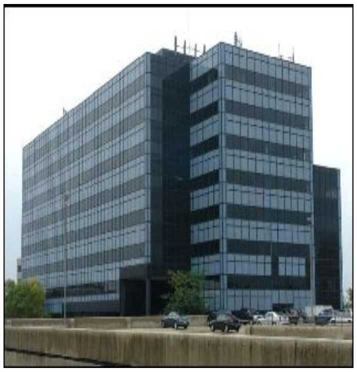


Figure 4: An example of office property

Our initial analysis included those multi-label flyers however for our final analysis we decided to analyze based on just single label flyers. The reason for this is because we felt that analyzing multi-label and single-label flyers together was too broad. Finally, we selected 844 outdoor images out of the original flyers. The distribution of the number of flyers and their proportion in the dataset are shown in Table 2.

Table 2. Distribution of different property types with their number of images and their proportion in the dataset

Property Type	Number of Images	Proportion of Images
Industrial	216	26%
Multi-family	65	8%
Office	182	21%
Retail	338	40%
Land	43	5%
Total	844	100%

Figure 5 show examples of the outdoor-building genre from our flyers.

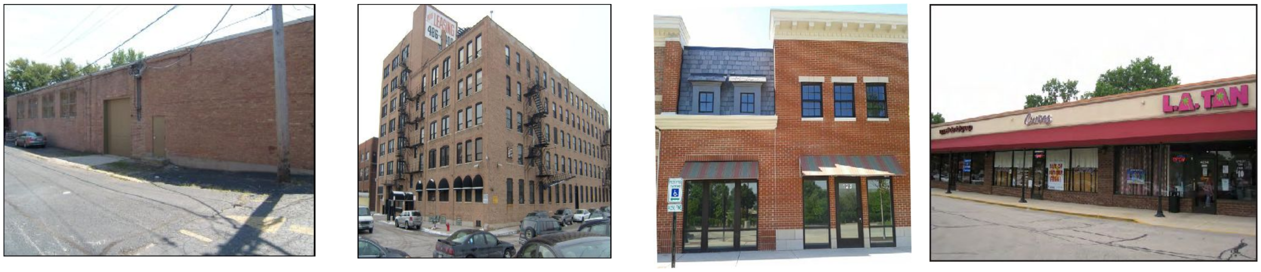


Figure 5: Examples of outdoor image genres

3.2 The Framework

The framework that we have developed will look for Objects like Cars, Logos, Features from Window and General features from building. We start with outside building image and we try to recognize the cars by our car detection algorithm (see section 3.3). Then from the outside image we recognize the building by our w recognition algorithm (see section 3.4) From that we extrapolate the Windows (see section 3.5) Logos (see section 3.6) and General features from Building. Finally, we extract 38 features from detected building, cars, logo, and windows. The framework of our approach can be seen in figure 6:

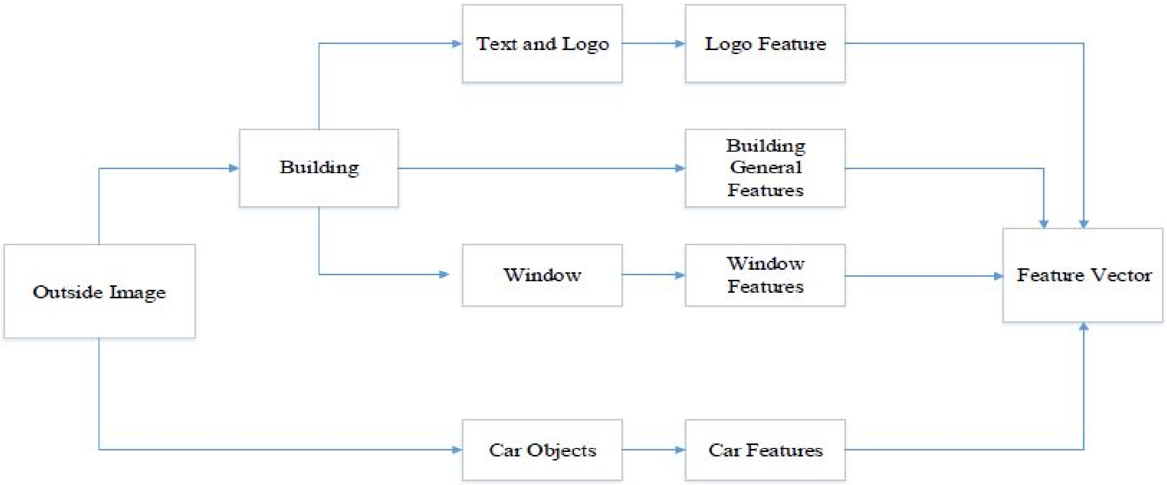


Figure 6: The framework of our work

3.3 Car Detection

Cars can tell us more about the use of the building. For example, an office might have lots of cars parked in front it, where land or multi-family may not have too much cars. We decided to use the work presented by Viola & Jones [7] for detection of cars. The referred object detection system uses HOG features and their respective rotated versions. For the training of the classifier we used 4000 car image objects from international SUN image databases [8]. After detection the cars the algorithm will categorize the number of cars; Low (0-3), Medium (3-7), High (7 and above).

3.4 Building Recognition

We thought that there are some useful features inside the facade building that can be helpful for the task of classifying flyers based on their property types. For instance, after identifying buildings the task of identifying windows would become easier and more accurate. Our approach for building recognition has two main steps: 1- Building Detector 2- Region Growing.

1- Building Detector: For building recognition we first apply the object detector by Viola & Jones which uses Haar-like features and trained by 6000 building images from SUN database [7,8]. The training set included both occluded and non-occluded samples. However, with this algorithm we are not able to find the area of the entire building. We made approximations of the regions that building could exist. In order to solve this issue, we applied the Region Growing algorithm.

2- Region Growing: This algorithm has been used multiple times for image segmentation [9,10]. In this algorithm, finding the initial seed is very vital. If the selected seed is inside a door or wall, then we will be unable to find the whole area. Because the doors and windows are negative spaces, the algorithm will only find the area of that space, unable to spread past the confines of the door or window. However, If the seed would be selected inside the wall then we are hopeful to capture the whole building area. We determined that the pixels with highest color frequency is more likely to be the wall because it is the largest part of the building. So out of the set of pixels with highest color frequency, we select the one which is the closests to the center of the building and apply the region growing algorithm to the seed.

3.5 Window Detection

After finding the building dimensions (sec. 3.4), the next goal would be to identify the window dimensions. In order to detect the windows**,** we proposed an alternative approach that functions based on the changing vanishing points. This alternative algorithm identifies the largest vanishing points of a rectangular shape. Specifically, we are analyzing windows and finding the best angle to apply window detection. The reason why we are performing this approach is because our window detector program performs optimally when looking at the object straight on. If the shape has an angle the program is unable to detect the object because our window detector is trained to recognize only rectangular shapes.

In order to begin, we first transform the 2D image into 3D by using Make3D[11]. Then we rotate the 3D objects around their z-axis and for each angle we compute the vanishing points. We use the vanishing points because as you rotate the object the angle changes from a fixed perspective. As the vanishing point decreases the angle of the window is sharper. But as the vanishing points becomes larger the object moves to a more head on view. So when we find the angle with the largest vanishing point we will transform the 3D object back into its original 2D dimension. After that we will apply the window detection on the newly rotated object. Similar to building and car detection we used Viola & Jones object detector which uses Haar-like features and trained by 15000 window images from SUN database [7,8]. The training set included both occluded and non-occluded samples. Other studies that have used window detection never rotate the original object because they simplified the problem by using the straight on images. However, our method is unique because we don't necessarily need the straight forward image for the task of object recognition. As a result, this method is unique and applicable to other object detection tasks. This process eliminates the need for huge training datasets and also simplifies the method used for object detection.

3.6 Logo Detection

For logos, the algorithm will simply detect the text from the building area (referenced in section 3.4) whether there is logo or not. We used the Tesseract Optical Character Recognizer(OCR) API for recognizing characters in the building image [12,13]. If there is a logo, this might tell us it is a retail, if not that flyer might be hinting at other property types. We are looking for texts that are bigger than a certain threshold and we want the logo dimensions to be proportional to the building size, because the small texts or numbers could be mistaken for building addresses. As it shown below in figure 7 this would be an example of a building address which we are not interested in. Figure 8 is showing the logo that we are interested in.

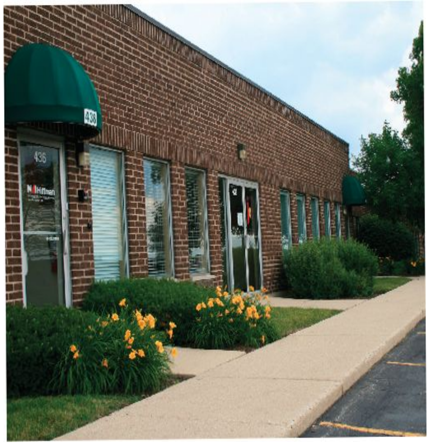


Figure 7: building with not interested embedded text

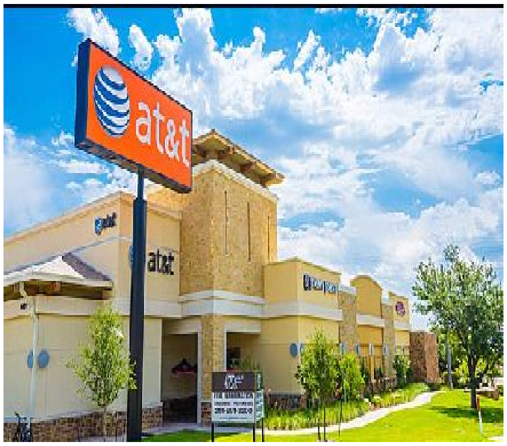


Figure 8: Building with interested embedded text

3.7 Feature Extraction

After building recognition, window detection, logo detection and car detection, we extracted some image features, which will be fed into the classification algorithm. In particular, for each image we calculated (1) Building general features, (2) Window features, (3) Logo feature and (4) car feature. For (1) Building features, we extracted the GIST features. Some prior studies [14, 15] 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 [15, 16] feature has been extracted from the images. We then quantized them into 32 features. For (2) window feature, we extracted 4 features which are: number of the windows in the building, the sum of the windows area proportional to the building size, the sum of the windows length proportional to the building length, and the sum of windows width proportional to the building width. For (3) logo feature, we have a binary feature which is 1 when a text is found and 0 when it does not exist. For (4) car feature, we counted the number of cars; and put them in three categories which are: Low (0-3), Medium (3-7), High (7 and above). By putting together these features, we obtained the final feature vector of length 38 (1x38) for each image.

4 Flyer Categorization

The flyer categorization task involves labeling all flyers with appropriate transaction property types, such as; retail, industrial, m, o, and land. This is a multi-label classification task as in all cases a flyer can have more than one label, however; some can have more than one but we decided to focus on single label flyers. We applied a supervised Machine Learning approach to the task utilizing Support Vector Machines (SVM) using the LibSVM library [17].

To see the effect of different features we ran 3 different test scenarios, because we were interested to reduce the number of features without sacrificing the classification accuracy. In the first scenario we used all 38 features. In the second scenario, we only included the Window, Logo and Car features and the third scenario we just included the Window and logo Features. Table 3 illustrates the results for these scenarios.

Table 3: Results from applying SVM on the task of identifying flyer property type (retail, office, industrial, land, multi-family). p = precision, R = recall, f= f1-score

		GIST-Window-Car-Logo	Window-Car-Logo	Window-Logo
Property type	P	0.461	0.478	0.486
	R	0.46	0.489	0.498
	F	0.46	0.481	0.485

Table 4: Confusion matrix for scenario 3

	Retail	Office	Multi-family	Industrial	Land
Retail	263	18	6	66	12
Office	52	53	19	51	7
Multi-family	23	16	12	12	2
Industrial	70	27	9	97	13
Land	10	1	0	10	22

As the table 3 shows there is an observable difference on between the accuracy levels depending the number of features. The accuracy level increases as features decreases. So we have fewer number of features. We also found that Retail category has been classified very well (table 4), but the others still need further experiments. Based on these results we can conclude that we can lessen the semantic gap using this approach.

5 Conclusions and Future Work

In conclusion, in this work we presented our work on classifying the outdoor images embedded in the real estate flyers by the property type. We were trying to find a new algorithm in order to categorize the online flyers based on their property types. We decided to use new feature vector which includes windows, logos, cars and general building features and run them into the algorithm. We also proposed new ideas for Building identification and Window recognition. The results of this study was windows were a feature that differentiated retail properties very well. Although other features did not provide the same distinct differentiation. Our results still can be useful for the flyer categorization.

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 outdoor-building (map, schematic drawing, aerial photo and indoor-building) would also be helpful for improvement of flyers classification.

References

[1] Pourashraf, Payam, Noriko Tomuro, and Emilia Apostolova. "Genre-based Image Classification Using Ensemble Learning for Online Flyers." International Conference on Image Processing (ICDIP), 2015.

[2] Apostolova, Emilia, and Noriko Tomuro. "Combining Visual and Textual Features for Information Extraction from Online Flyers." Empirical Methods in Natural Language Processing (EMNLP), 2014.

[3] [11] Huiskes, M. J., Thomee, B., & Lew, M. S. (2010, March). New trends and ideas in visual concept detection: the MIR flickr retrieval evaluation initiative. InProceedings of the international conference on Multimedia information retrieval(pp. 527-536). ACM.

[4] Guillaumin, M., Verbeek, J., & Schmid, C. (2010, June). Multimodal semi-supervised learning for image classification. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (pp. 902-909). IEEE.

[5] Srivastava, Nitish, and Ruslan R. Salakhutdinov. "Multimodal learning with deep boltzmann machines." Advances in neural information processing systems. 2012.

[6] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on (Vol. 1, pp. I-I). IEEE.

[7] Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010, June). Sun database: Large-scale scene recognition from abbey to zoo. In Computer vision and pattern recognition

(CVPR), 2010 IEEE conference on (pp. 3485-3492). IEEE.

[8] Mohammed, Mazin Abed, et al. "Automatic Segmentation and Automatic Seed Point Selection of Nasopharyngeal Carcinoma from Microscopy Images Using Region Growing Based Approach." Journal of Computational Science (2017).

[9] Duan, H. H., Gong, J., & Nie, S. D. (2016, August). Two-pass region growing combined morphology algorithm for segmenting airway tree from CT chest scans. In Control (CONTROL), 2016 UKACC 11th International Conference on (pp. 1-6). IEEE.

[10] Make3D: Learning 3-D Scene Structure from a Single Still Image, Ashutosh Saxena, Min Sun, Andrew Y. Ng, In IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2008.

[11] Google. "tesseract-ocr", https://code.google.com/tesseract-ocr/, Retrieved 2008-07-12.

[12] Kay, Anthony. "Tesseract: an open-source optical character recognition engine." Linux Journal 2007.159 (2007): 2.

[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] Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), 27.