Object Recognition on Satellite Images with Biologically-Inspired Computational Approaches

M.I. Sina, P. Payeur and A.-M. Cretu
University of Ottawa/Electrical Engineering and Computer Science, Ottawa, Canada
msina070@uottawa.ca, ppayeur@eecs.uottawa.ca, acretu@eecs.uottawa.ca

***Abstract*—The human vision system is often significantly superior in extracting and interpreting visual information when compared to classical computer vision systems. The exploitation of existing knowledge about human perception is expected to improve the performance of computational vision systems. Computational visual attention has been reported to improve scene understanding performance. This paper discusses some of the difficulties faced by the current generation of visual attention systems when applied on satellite images. Next, a novel technique for top-down attention is devised which is based on the energy of bottom-up feature maps and overcomes some of the limitations of previous approaches. The computed top-down map is then used as a method of object localization in the object recognition phase that makes use of texture and shape information using local binary patterns, Legendre moments and Hu moment invariants. The proposed algorithm is shown to perform better than other similar systems on satellite images in many aspects.**

I. INTRODUCTION

A. Context

The human visual system is capable of processing huge volumes of visual data efficiently and accurately. Among several processes that occur in the visual system, the one responsible for finding important and interesting regions and concentrating efforts for object recognition is known as visual attention. It helps to save valuable time by avoiding the demanding recognition process on irrelevant parts of the scene. On the other hand, classical image processing and computer vision techniques often struggle to fully understand even the simplest images. Therefore, several attempts to computationally reproduce such human-like behavior were reported in the literature. However, most of those systems have been tested on simplistic images. Experiments have also shown that they perform poorly on complex natural images, especially on satellite images. In this work a novel approach is described that outperforms other similar systems in the realm of computational visual attention and object detection.

The structure of the paper is the following: Section I.B provides the necessary background on the field of computational visual attention and a short literature review. Sections II.A and II.B focus on the description of one of the newest computational attention models, proposed by Frintrop, and its learning mechanism. In section II.C, the limitation of Frintrop's learning mechanism and its reasons are analyzed. Section III deals with the proposed learning technique, while the object recognition technique is presented in section IV. In section V, experimental results for both the proposed learning technique and the object recognition technique are provided. Finally, section VI concludes this paper.

B. Background

Visual attention is divided into two processes: bottom-up and top-down. Bottom-up attention is derived directly from the contents of a scene, whereas top-down attention is influenced by the knowledge, expectation, and current goal of the observer. Several computational attention systems have been proposed to implement bottom-up attention. These systems, given an input image, generate measures of local saliency in the form of a saliency map. The saliency map specifies the degree of conspicuity of any region within the image with respect to the region's surroundings. On the other hand, top-down attention is more goal-directed and attempts to highlight regions of the image containing objects of predefined interest.

The model for bottom-up visual attention proposed by Koch and Ullman [1], which is based on the Feature Integration Theory of Treisman and Gelade [2], is a seminal work in this field. Many other systems are derivatives of this model including the implementation of Milanese [3] which employs filter operations for the computation of feature maps. In this respect, Itti *et al.*'s attention system [4] is very popular. The latter formerly computed only bottom-up saliency, but was eventually extended to incorporate top-down saliency as well [5] [6] [7]. Later, Frintrop brought many improvements to Itti's attention system, and her model encompasses both bottom-up and top-down attention [8]. This paper is highly influenced by Frintrop's work on visual attention systems, which is described in the next section. The preceding work of this research can be found in [9].

II. FRINTROP'S ATTENTION SYSTEM

A. Frintrop's Bottom-Up Saliency

Frintrop's bottom-up saliency map [8], a derivative of Itti's attention system [4], is built using three feature channels – intensity, orientation and color. These features are the most commonly used in other visual attention systems as well. Frintrop's system (named VOCUS) employs both Gaussian and Laplacian image pyramids [10] for different feature channels to make the system scale-invariant. Specifically, the pyramids have five layers denoted by S_0, S_1, S_2, S_3, and S_4, where S_0 is the input image. The system discards the first two layers to reduce the impact of noise and consumption of computational resources. At first, for each feature channel, feature maps are computed (ideally in parallel) over layers S_2 to S_4, which are then summed up and normalized, resulting in conspicuity maps. Next, the final saliency map is computed by first applying a uniqueness function on each

978-1-4673-1014-7/12/$31.00 ©2012 IEEE

conspicuity map and then summing them up. The uniqueness function, which promotes maps with fewer peaks and demotes maps with many peaks, is defined as:

$$W(X) = X / \sqrt{m} \quad (1)$$

where X is a map and m is the number of peaks in the range $(M/2, M]$ within X with M being the global maximum of the map X. The computation of the feature conspicuity maps is described briefly below.

Intensity – The intensity feature deals with intensity variations over the image area. Computed intensity feature maps include the on-center map and the off-center map. The on-center map highlights regions that are surrounded by relatively darker regions; off-center map highlights regions that are surrounded by relatively brighter regions. The higher the difference in brightness in the image, the stronger is the response in the feature maps. At first, a Gaussian pyramid is built using the grayscale version of the input image and all computations are done on each layer, where a rectangular region is considered centered at each pixel. Mean intensity of this region is computed and compared to that of the center pixel. In case of on-center map, the center pixel is replaced with *max(0, pixel_intensity – mean_intensity),* and in case of off-center map the center pixel is replaced with *max(0, mean_intensity – pixel_intensity)*. Lastly, across-scale addition is performed for both on-center and off-center maps separately resulting in two feature maps, which then undergo the uniqueness function, *W(X)*, and are summed up. Conspicuity for intensity feature is thus achieved.

Orientation – Experimentations on human visual system showed that human vision does not respond equally to all orientations of edges. An orientation-selective edge detector, such as for example a Gabor-like filter [11], is essential to emulate such behavior. The procedure devised by Greenspan *et al*. [12] is then applied for the computation of orientation maps. First, a Filter-Subtract-Decimate (FSD) pyramid [13] which is a derivative of Burt and Adelson's pyramid [10] is computed on the grayscale version of the input image. Complex sinusoidal modulation with orientations 0, 45, 90 and 135 degrees is done on each layer of the pyramid resulting in four pyramids. After that, each layer of the pyramids is low-pass filtered with 5-sample Gaussian-like separable filter (1/16, 1/4, 3/8, 1/4, 1/16). Then, four feature maps, one for each orientation, are obtained by applying across-scale addition on each pyramid. Finally, these four maps undergo, *W(X)*, and are summed up, thus resulting in orientation conspicuity.

Color – Color plays an important role in the human visual system. Different experimentations advocated the existence of photosensitive cells for identification of greenness against redness (and vice versa) and blueness against yellowness (and vice versa). Emulation of this behavior requires conversion of input image from RGB color space to *CIE LAB* color space; where *L* denotes luminance or intensity of a pixel, *A* denotes greenness-versus-redness and *B* denotes blueness-versus-yellowness. The *L* component is dropped since intensity feature is already considered before and also to reduce the impact of changes in lighting conditions. A Gaussian pyramid is then constructed based on the *(A, B)* pair. After that, for each color and for each layer on the pyramid, color distance maps are calculated resulting in a four color pyramid. For example, for the red color, which in the *A-B* color space is (127, 0), all distances are measured from the (127, 0) coordinates. Next, on-center maps are calculated on top of color distance maps for each layer. Finally, each pyramid is across-scale summed and four conspicuity maps are obtained, one for each color. As before, *W(X)* is applied on all of them before they are summed up for the final color feature map.

Saliency Map – A final bottom-up saliency map is then computed based on the conspicuity maps. All three conspicuity maps undergo *W(X)* before being summed up in the saliency map. At this point the most salient region (MSR) is determined by first identifying the highest value in the map and then applying a flood-fill algorithm at that point within a threshold. An example of a final saliency map is given in Fig. 1 for a satellite image originating from [14], but without identification of any MSR at this stage.

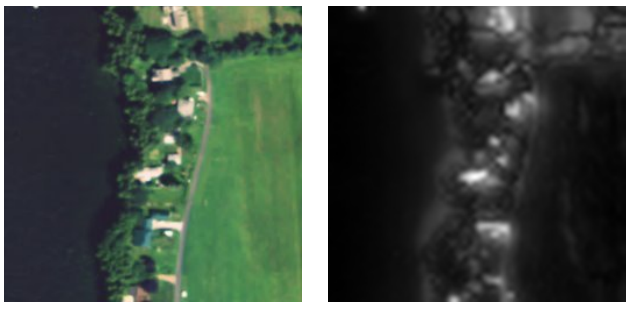
B. Frintrop's Learning Mechanism

In [8], Frintrop also proposed a learning mechanism based on bottom-up attention to compute top-down saliency. As mentioned earlier, top-down attention is goal-directed and helps in object search tasks. For example, if an observer is looking for a blue car in a parking lot, then objects that are blue and have the shape of a car are important and should be highlighted; even though they may not be among the most salient regions from the perspective of bottom-up attention. Top-down attention is computed by combining various unbiased bottom-up features with the right amount of weight. The challenge is to compute the weight vector. In Frintrop's system there are 13 such weights (two for intensity, four for orientation, four for color and three for conspicuity).

Learning – The goal of the learning phase is to identify the properties of the object of interest in terms of feature and conspicuity maps. In Frintrop's system, a training image is selected along with a rectangular region of interest that contains the target object. Next, VOCUS computes all 13 maps according to the description given in section II.A. After that, the weights are computed as follows:

$$w_i = m_{i_{MSR}} / m_{i(image - MSR)} \quad (2)$$

where $m_{i_{MSR}}$ denotes the mean intensity of the MSR within the given region of interest (ROI) and $m_{i_{(image-MSR)}}$ denotes the mean intensity of the rest of the image; all for the i^{th} map where $1 \leq i \leq 13$. It can be observed that the



(a) (b)

Figure 1. (a) A sample satellite image, and (b) its corresponding bottom-up saliency map. The green field and the lake are mostly non-salient, but small houses and other objects are highlighted in the saliency map.

denominator depends on the background contents of the object. This makes the learning phase highly dependent on the background and affects the searching phase when the background of the test image is different from the one in the training image. The learning mechanism is also extended such that the system can learn from multiple images.

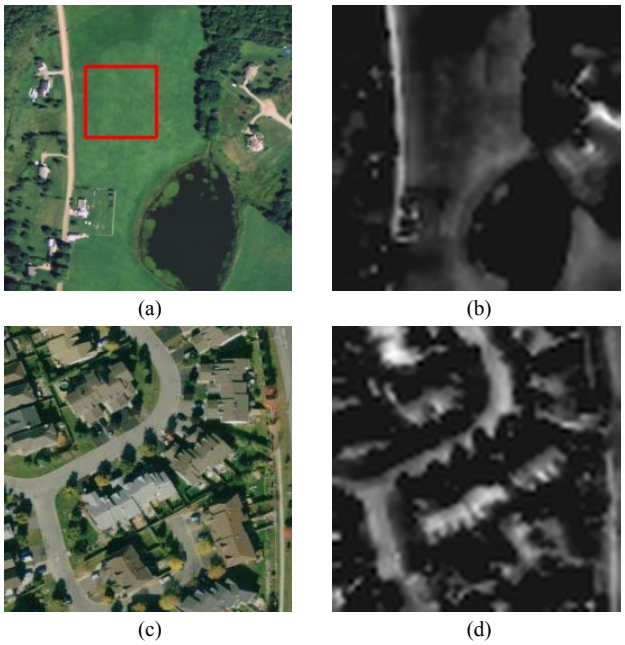
Searching – After the system learns the weight vector for a particular class of objects, it can be employed for object searching. Given a test image, the system computes all 13 maps as described in section II.A and applies the following formula:

$$I_{TD} = \sum_{w_i>1} w_i Y_i - \sum_{w_i<1} \frac{1}{w_i} Y_i \quad (3)$$

where Y_i is one of the 13 maps and w_i is its corresponding learned weight. The role of eq.(3) is to excite the maps that have large associated weights, and inhibit the maps that have small associated weights, therefore putting in evidence areas in the saliency map with similar properties to the one of the searched object and decreasing the intensity of those areas that have different properties from the object of interest. Negative values resulting from eq.(3) are replaced with zeros.

C. Limitations of Frintrop's Top-Down Attention

Fig. 2 depicts an example using Frintrop`s top-down attention approach where the selected search areas represent grass (red bounding box in Fig. 2(a)). For a self-test (Fig. 2(b)), most surfaces similar to grass are detected along with some outliers corresponding to roads. When tested with a different image (Fig. 2(c)), the resulting map (Fig. 2(d)) deteriorates and mainly brings in evidence irrelevant regions (e.g. roads and roofs). The limitation is inherent to the learning mechanism as the system also learns the background of the training image in the process. This can be observed from eq.(2) where the denominator is the mean intensity of the background, which affects the computed weight which is later used in eq.(3) to compute the final top-down map.



(a) (b) (c) (d)

Figure 2. (a) The training image with ROI shown in a red rectangle, (b) top-down map computed using (a) as the test image, (c) a different test image, and (d) the outcome of applying the search method. It can be seen that instead of highlighting grass areas exclusively, the system highlights other objects as well, such as roads and house roofs.

Another issue is the specification of the ROI. It is unclear how large the ROI should be. The numerator of eq.(2) is the mean intensity of the MSR within the ROI and hence the size of the ROI would affect its value. On the other hand, the MSR is restricted to a small contiguous region and hence, in certain applications, insufficient to represent an entire object, especially a complex one. In this paper a novel technique is proposed which not only is background independent but also free of ROI specification, and at the same time does not involve MSR.

III. Proposed Top-Down Attention System

A. Proposed Learning Mechanism

In the proposed learning mechanism, the importance of a feature map is defined in terms of energy of the signal. A high energy map is given more importance than a low energy one. Energy is defined as follows:

$$E(X) = \sum_x \sum_y X[x,y]^2 \quad (4)$$

High energy values in a given feature map, *X[x,y]*, signify that from the perspective of bottom-up attention, the map is important. For example, if the training image contains objects that have strong vertical edges, then the feature map for 90 degree orientation would have higher energy values than those of other orientations. Conversely, if the training image does not contain a red object then the feature map for the red color would be very weak. In both cases the energy of maps is capable to encode such information. At the same time, since the proposed approach does not involve any formally defined MSR, it is not restricted to a small region that defines an object.

The proposed learning mechanism is well-suited for learning from a large number of images. The algorithm only uses feature maps. Therefore the computation of conspicuity maps is not necessary. The proposed solution can be described as follows: for each image in the set, ten feature maps (two intensity maps, four orientation maps and four color maps) are computed according to the process described in section II.A. Equation (1) is then applied on each of them. Next, the energy of each map is computed using eq.(4), where X represents each of the feature maps respectively. A 10-dimensional energy vector, $\boldsymbol{B}$, is formed for each training image. All energy vectors are combined to compute the average energy vector, $\boldsymbol{R}$, for the training set, using the following equation where n represents the number of training images:

$$R = \frac{1}{n} \sum_k B_k \quad (5)$$

It is important to note that all these vectors would not be too dissimilar as will be further discussed in section III.B. Finally, the average 10-dimensional energy vector, $\boldsymbol{R}$, is mean-shifted as follows:

$$r_i = r_i - \bar{r}; \quad \bar{r} = \frac{1}{10} \sum r_i \quad (6)$$

where r_i are components of $\boldsymbol{R}$, and $\bar{r}$ is the average vector computed over the components of $\boldsymbol{R}$ corresponding to the 10 different feature maps. The resulting r_i components from eq.(6) constitute the final learned weight vector. Note that the sum of the r_i components is zero (since they

are mean-shifted), which also means that at least one of them is negative and at least one of them is positive. This is helpful to inhibit the features from the final top-down map that are weak and excite the feature maps from the final top-down map that are strong. After the weight vector is learned for a training set that represents a class of images, the top-down map is computed using the following formula:

$$TD(I) = \sum_{1 \le i \le 10} r_i X_i \quad (7)$$

The inhibition and excitation are applied automatically in the summation. The test image, *I*, is expected to belong to the same class as the training set, in which case the objects that were interesting in the training set would be highlighted. In case the test image belongs to some other classes, the top-down map would highlight at least similar looking objects. The proposed technique is interesting in that it does not need to know explicitly what features of the objects to learn. Based on the strengths of bottom-up features, the proposed energy-based system identifies and enhances the "important" features that characterize objects in images from a given dataset automatically.

B. Selection of Training Set

For the significant regions to be identified properly, the training images are chosen such that their contents are similar and relevant for the task that the system is trained for. For example, if one wants to process satellite images of cities, as it is the case in the current application, the training set should contain ideally only images of cities and not contain other types of images such as desserts or farm lands. The use of other images than those relevant for the task results in changes in the 10-dimensional vector, ***B***, that characterizes each image and that, in turn, leads to the bias of the average energy vector, ***R***, to faithfully represent any of the classes of interest. When representative for the task, "objects of interest" are automatically learned by the system. The latter objects are defined by features that are strong with respect to their neighborhoods and therefore highlighted in the bottom-up saliency map and further enhanced by the strategy proposed in section III.A. For example, in city images, among the many objects present, objects of interest can for example be houses, roads, or parking lots.

IV. Proposed Object Recognition

A. Procedure Overview

Recognizing objects in a scene is one of the fundamental problems in computer vision. Usually the first step towards successful object recognition is object localization. In the present context, localization can be accomplished with the help of the proposed top-down learning mechanism. First, the system is trained with a suitable set of training images according to the procedure described in III.A and III.B. Next, a test image is fed in the system and its top-down saliency map is obtained. The bright regions in this map are expected to highlight the objects of significance. The map is converted to a binary image by replacing small values of significance with 0s and all other values with 1s. The threshold was determined by trial and error method and set to 0.25. Finally, contiguous regions filled with 1s are extracted one after another; thus completing the object localization step.

Next, corresponding source image regions are examined for further recognition. The recognition phase also involves a learning phase, where an object signature is learned. Both texture and shape are important for a signature to be effective; since two different objects may have similar texture but different shapes, or conversely similar shape but different textures. In the proposed technique two measures are used: Local Binary Pattern (LBP) [15] as texture signature and image moment-invariants [16] as shape signature. The learning process is described in the following section.

B. Learning the Object Signature

Learning of the object signature is a separate and independent process whose goal is to describe objects based on their physical appearances rather than bottom-up or top-down features. Training images for object signature learning are in the form of mask images (see section V for such mask examples). Mask images specify where the objects in the training images are located. For example, if one wants to learn roads from satellite images, then the mask images contain only roads and black pixels everywhere else. Masks are prepared by manual segmentation and preferably each class of object is represented in multiple training masks. The system goes through each mask and computes a signature of the objects associated with the mask in the form of a vector, and finally computes the resultant vector as a representative signature of that object class. Computed object signatures consist of texture and shape descriptors and are discussed in following sections.

C. Texture Learning

Local binary patterns (LBP) were shown to be effective in texture classification in [15]. They are therefore used in the context of this work as a basis for texture learning. In general, computation of LBP is as follows: first, within each 3x3 window, 8 neighbors of the center pixel are numbered from 0 to 7 in a given fixed order. Then for each neighbor of the center pixel, if its value is larger than that of the center pixel, the LBP value of the center pixel is added with 2^i, where i is the number associated with the neighbor pixel being compared. In this way, an LBP value is computed for each pixel of an image resulting in an LBP map. In the context of this work, the LBP coding is computed over the source image for each of its red, green and blue color channels, resulting in three maps containing values in the range [0, 255]. Next, a histogram of the LBP values is calculated over the regions of interest in an image (where the corresponding mask value is non-zero). Finally, the histogram is converted to a probability density function (PDF) by dividing the frequency values by the sum of all frequency values and a 256-dimensional vector is formed for each color channel. A 3-color image requires a 768-dimensional vector which would slow down the system severely. To work around this problem, 1-dimensional Legendre moments [17] up to a certain order (e.g. 10) are calculated on top of the histogram PDF. Legendre moments are based upon Legendre polynomials which are orthogonal to each other in the range [-1, 1] and thus are able to classify optimally with a low number of moments. The Legendre moment of order n is calculated using the following equation:

$$L(n) = \int_{-1}^{1} P_n(x) * f(x) dx \quad (8)$$

where $f(x)$ is the histogram PDF and $P_n(x)$ is the n^{th} order Legendre polynomial. Ultimately, texture is defined by an effective but low-dimensional vector.

D. Shape Learning

Incorporating shape information is advantageous as it can be observed from our everyday life experience. Human beings in many cases can recognize objects even from their outlines only. Moment invariants have been successfully used as shape descriptors in applications such as character recognition or image registration [17]. Hu moment invariants [16], which are based on 2-dimensional geometric moments, are selected in the context of this work. Hu derived seven moment invariants that are invariant to translation, rotation and scaling. From the training mask images defined in section IV.B, each contiguous region is extracted and their contours are identified. After that, seven Hu moment invariants are computed from those contours and a 7-dimensional vector is formed for each object present in the mask. Finally, the resultant of all vectors is computed as a vector average, to represent the shape of objects.

Figure 3. A subset of training images that depict top-view of a city area.

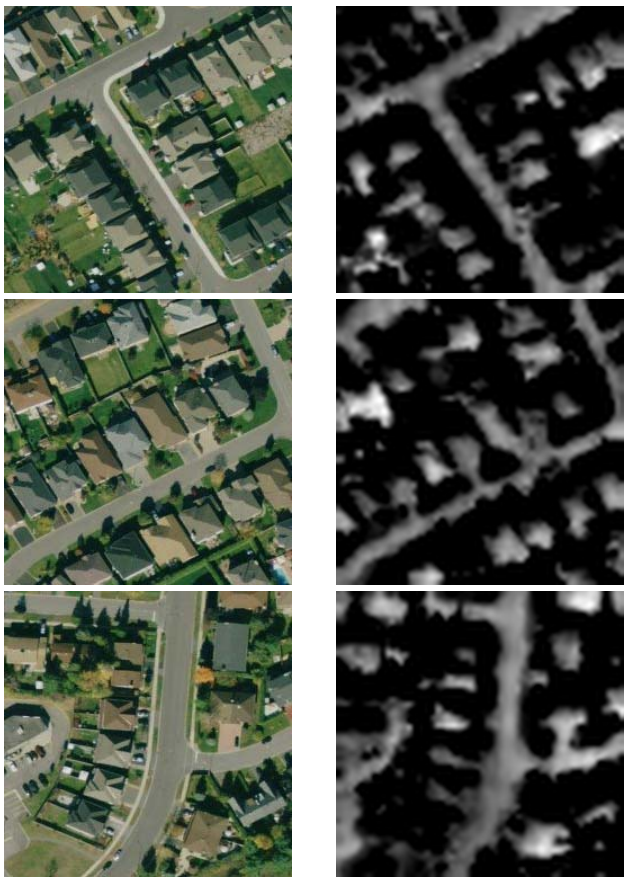


Figure 4. Test images (which were not used in the training set) are shown on the left column and corresponding top-down maps on the right. The proposed technique was able to identify most of the objects of interest. For the training set used, important objects are houses and streets.

E. Final Recognition

The vectors representing texture and shape are concatenated in a larger vector to obtain the final object signature. Learned object signatures are stored in a database. Next, given the test image, its bottom-up feature maps are computed and then using eq.(7) the top-down map is obtained as detailed in section IV.A. The extracted regions then undergo the same procedure as for object signature learning and several vectors of unknown classes are obtained from that test image. The database of pre-learned vectors is then iterated once for each of the objects found in the test image and the closest match is found by means of Euclidean distance. The corresponding object is then labeled with the color associated to its recognized category. Related results are presented in section V.B.

V. Results

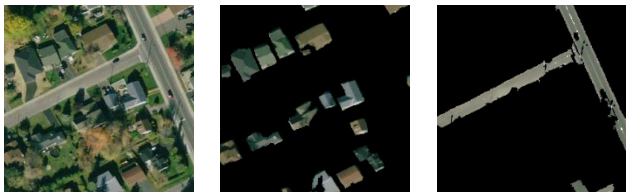
A. Top-Down Learning

In the experiments, a set of training images (consisting of 17 images) [14] is used, of which a few are illustrated in Fig. 3. Other test images and their corresponding top-down maps are shown in Fig. 4. Brighter areas in the top-down maps are associated with regions where the underlying feature map values are well aligned with the learned weight vector (high value for positive weights and low value for negative weights). It can be observed that computed top-down maps can be used for effective object localization which would later be helpful for object recognition. The system identified houses and streets as "important" objects, which means that visual properties of these objects were stronger in terms of bottom-up feature maps.

B. Object Recognition

The recognition technique was tested on a binary classification that separates houses from streets. First, object signatures for houses and streets are learned using the 17 images from the training set. As mentioned earlier, this is done through the use of mask images. Examples of mask images for both houses and streets are presented in Fig. 5. Finally, Fig. 6 shows the object recognition results. The recognition rate is good as most of the streets and houses are labeled correctly. In order to quantify the performance of the object recognition technique, the number of houses and the pixel area of streets are initially computed in each input test image. After the recognition phase is complete, the number of correctly classified houses and the area of streets are obtained and the average recognition rates are reported. Based on 10 test images, from the total number of houses, on average 68% of the houses are correctly classified; from the total number of pixels belonging to streets, 81% are correctly classified as such. In a small number of cases, the system confuses between streets and houses.

Texture similarity between roofing and asphalt is the primary reason behind this behavior. Due to the



(a) (b) (c)

Figure 5. Examples of mask images: (a) original image, (b) segmented houses, and (c) segmented streets.

incorporation of shape information, the system learned that blob-like shapes are houses and longer shapes are streets.

In some cases, the binary converted top-down maps contain regions that are thin and long and with a texture similar to that of a street causing the system to classify them wrongly as street, instead of houses. On the other hand, the shapes of the recognized objects do not match perfectly to those of the source image. To overcome this problem, some additional post-processing steps are required and will be added as a future extension to this work.

VI. CONCLUSION

This paper presents a novel biologically-inspired technique for object recognition from satellite images. In the course of this recognition, the key step is the computation of a top-down map which helps in localizing objects of interests. The technique devised for top-down map computation overcomes some of the limitations of its predecessors. Early experimental results using the technique are plausible and demonstrate the potential of this computational representation of knowledge.

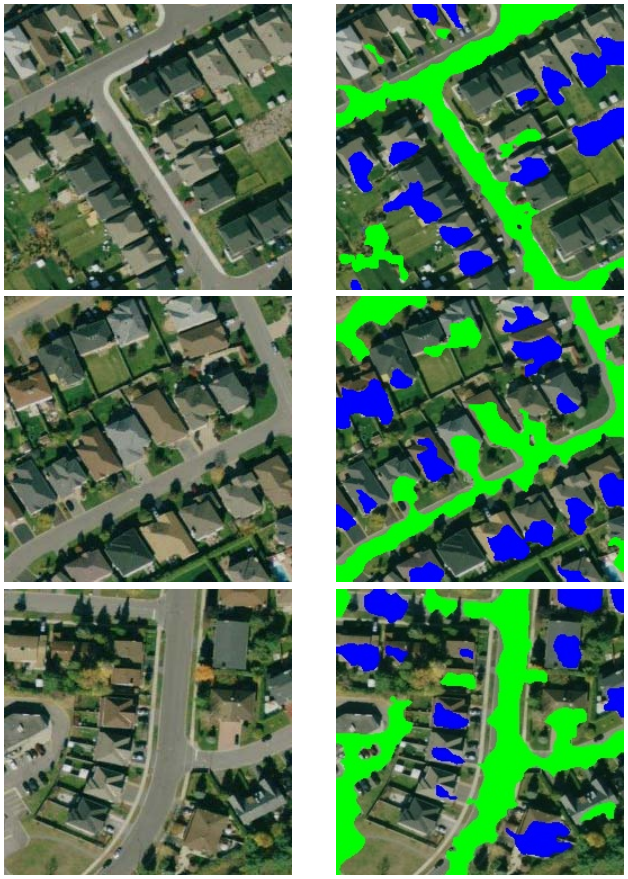


Figure 6. Left column shows some test images reproduced from Fig. 4. Right column shows the resulting labeled images. Green and blue colors represent streets and houses respectively.

The incorporation in the object signature of both texture and shape information make the recognition process more effective. The use of Legendre moments on top of LBP histogram PDF helps represent texture information compactly yet effectively. Shape information is encoded using tested and well-understood Hu moment invariants.

REFERENCES

[1] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human Neurobiology,* vol. 4, no. 4, pp. 219-227, 1985.

[2] A. M. Treisman and G. Gelade, "A feature integration theory of attention," *Cognitive Psychology,* vol. 12, pp. 97-136, 1980.

[3] R. Milanese, "Detecting salient regions in an image: from biological evidence to computer implementation," PhD Thesis, University of Geneva, Switzerland, 1993.

[4] L. Itti, C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. PAMI,* vol. 20, no. 11, pp. 1254-1259, 1998.

[5] V. Navalpakkam and L. Itti, "Modeling the influence of task of attention," *Vision Research,* vol. 45, pp. 205-231, 2005.

[6] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," *IEEE Trans. PAMI,* vol. 29, no. 2, pp. 300-312, 2007.

[7] C. Siagian and L. Itti, "Biologically inspired mobile robot vision localization," *IEEE Trans. Robotics,* vol. 25, no. 4, pp. 861-873, 2009.

[8] S. Frintrop, "A visual attention system for object detection and goal-directed search," PhD Thesis, University of Bonn, Germany, 2006.

[9] M. I. Sina, A. -M. Cretu and P. Payeur, "Biological visual attention guided automatic image segmentation with application in satellite imaging," in *IS&T / SPIE Electronic Imaging*, Burlingame, CA, January 2012.

[10] P. J. Burt and E. A. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. Communications,* vol. 31, pp. 532-540, 1983.

[11] J. G. Daugman, "Uncertainty relations for resolution in space, spatial frequency and orientation optimized by two-dimensional visual cortical filters," *Journal of Optical Society of America,* vol. 2, pp. 1160-1169, 1985.

[12] H. Greenspan, S. Belongie, R. Goodman, P. Perona and C. H. Anderson, "Overcomplete steerable pyramid filters and rotation invariance," in *IEEE Computer Vision & Pattern Recognition*, pp. 222-228, 1994.

[13] C. H. Anderson, "A filter-subtract-decimate hierarchical pyramid signal analyzing and synthesizing technique". USA Patent 4,718,104, 1987.

[14] "MapQuest," [Online]. Available: http://www.mapquest.ca. [Accessed February 2012].

[15] T. Ojala and M. Pietikainen, "Unsupervised texture segmentation using feature distributions," *Pattern Recognition,* vol. 32, no. 1999, pp. 477-486, 1998.

[16] M. -K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Information Theory,* vol. 8, no. 2, pp. 179-187, 1962.

[17] J. Flusser, T. Suk and B. Zitova, *Moments and moment invariants in pattern recognition*, West Sussex, U.K., PO19 8SQ: John Wiley & Sons Ltd., 2009.