ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, Issue 2, February 2013

Low and mid-level features for target detection in satellite images

Rajani.D.C

***Abstract*—Reliably detecting objects in broad-area overhead or satellite images has become an increasingly pressing need, as the capabilities for image acquisition are growing rapidly. The problem is particularly difficult in the presence of large intraclass variability. An automatic approach is used to detect and classify targets in high-resolution broad-area satellite images, which relies on detecting statistical signatures of targets, in terms of a set of biologically-inspired low-level visual features. Biologically-inspired feature extraction methods use the "gestalt" information (continuity, symmetry, closure, repetition) to conduct object detection. Gestalt psychology studies how the human visual system organizes the complex visual input into unitary elements. The goal of the visual system, computer or biological, is to transform a visual input into meaningful semantic information. A new methodology to learn relations inferred from Gestalt principles and an application to segment unknown objects, even if objects are stacked or jumbled and tackle also the problem of segmenting partially occluded objects. The relevance of the relations for object segmentation is learned with support vector machines (SVMs). Multispectral imaging is significant technology for the acquisition and display of accurate color information. This study shows that the proposed target search method can reliably and effectively detect highly variable target objects in large image datasets.**

***Index Terms*—Gist features, saliency features, satellite images, target detection, gestalt.**

IINTRODUCTION

OVERHEAD and satellite imagery have become ubiquitous, with applications ranging from intelligence gathering to consumer mapping and navigation assistance. This paper focuses on automatically detecting diverse types of targets with large intraclass variability in satellite images. Providing new means to automate this task is expected to facilitate and render more efficient the interpretation of satellite image by human analysts. An important technology for high-fidelity color reproduction, and multispectral cameras and scanners have been developed for digital archive, medical imaging, hardcopy reproduction, and manufacturing industry.

The problem of target detection is a difficult challenge in computer vision [1, 2]. A first, relatively straightforward approach is to use a provided (or trained) target template or model (hence, the feature is the image itself), to match against targets in the image of interest, at different locations, orientation and scales [3,4]. A second method for target detection is to use a model to extract a spatially sparse collection of invariant structural features (e.g., key point descriptors, bags of features) of the target even when viewpoint, pose, and lighting conditions vary [5-7]. In a third approach, using knowledge of target shape and characteristic geometry, several studies have proposed methods which learn and apply target geometric constraints on the keypoint feature locations [8]. In practice, the detection algorithms usually overlap these categories, and some approaches are intermediate between the geometry-based and "bag of features" approaches retaining

only some coarsely coded location information or recording the locations of features relative to the target's center [9]. In addition to these machine vision approaches, several biologically-inspired computational models have also started exploring target detection tasks in imagery, usually based on our knowledge of visual cortex, showing some promising experimental results [10-11]. Our approach extends these biologically-inspired frameworks. The Gestalt psychology studies how humans perceive the visual environment as unitary elements instead of individual visual measurements [12]. Through the years, Gestaltists have suggested different Gestalt principles for perceptual grouping and for figure-ground segregation [13]. The introduction of cheap and powerful RGBD-sensors which deliver a dense point cloud plus color for almost any indoor scene, recapturing of well-studied perceptual grouping techniques for segmentation of objects holds the promise to push the envelope slightly further. From Gestalt principles and by learning the relations with hand-annotated image data using support vector machines (SVMs). Finally a Graph-Cut algorithm is used to globally optimize segmentation of unknown objects. Based on the special properties of satellite image, several algorithms have been proposed to detect the targets in such kind of images. For example, for hyper-spectral satellite images, the features applied usually take advantage of the reflection characteristics of different materials [13-15] while for multispectral images, the features are usually extracted from fused spectra [16], [17]. Early on, the human visual processing system already makes decisions to focus attention and processing resources onto those small regions within the field of view which look more interesting or visually "salient" [18,19]. The mechanism of selecting a small set of candidate salient locations in a scene has recently been the subject of comprehensive research efforts and several computational models have been proposed [21-24]. One can make use of these models to predict possible target locations and target distributions. In this paper, saliency maps from several feature channels (intensity contrast, local edge orientation, etc.) are computed from a modified Itti-Koch saliency model [18], [22]. Given a static or dynamic visual scene, this model creates a number of multiscale topographic feature maps which analyze the visual inputs along visual feature channels known to be represented in the primate brain [22] and thought to guide visual attention and search [25] (luminance contrast, color-opponent contrast, oriented edges, etc.). Center-surround mechanisms and long-range competition for salience operate separately within each feature channel, coarsely reproducing neuronal interactions within and beyond the classical receptive field of early sensory neurons [26]. After these interactions, the feature maps from all feature channels are combined into a single scalar topographic saliency map. Locations of high activity in the saliency map are more likely to attract attention and gaze [20]. Thus far, saliency-based analysis of scenes has been predominantly applied to relatively small images, typically on the order of 1 megapixel (MP), with at least one study pushing to 24 MP [27]. Such smaller images are coarsely matching the amount of information which might arise from a primate retina (about 1 million distinct nerve fibers in each of the human optic nerves). With larger broad-area-search images, for example 400 MP–1000

MP satellite images, it becomes an interesting research question whether the mechanisms developed by the primate brain might scale up. Here, we address this question by developing a new algorithm, which analyzes large images in small chips, thus, mimicking the processing which human image analysts might operate when they deploy multiple eye fixations on an image, analysing each fixated location in turn. A second important research question is whether saliency maps might be useful at all for object classification, as opposed to being limited to just attention guidance as described previously. For example, target chips might have more numerous and sharper saliency peaks than nontarget chips. Our experiments and results test whether this approach is viable for complex target classification tasks where the intraclass heterogeneity is significant (e.g., find "boats," ranging from small pleasure craft to larger commercial or military ships). For each saliency map, mean, variance, number of local maxima, and average distance between the locations of local maxima are adopted to summarize saliency maps. In the full algorithm described in the following, all of these values from different feature channels' saliency maps are combined together to form the "saliency features" part of the proposed algorithm. For example, following presentation of a photograph for just a fraction of a second, a human observer may report that it is an indoor meeting room or an outdoor scene of a beach [28-30]. With very brief exposures (100 ms or below), reports are typically limited to a few general semantic attributes (e.g., indoors, outdoors, playground, mountain) and a coarse evaluation of the distributions of visual features (e.g., grayscale, colorful, large masses, many small objects) [31,32]. Gist may be computed in brain areas which have been shown to preferentially respond to "places," that is, visual scene types with a restricted spatial layout [33]. Like Siagian-Itti's gist formulation in computer vision [34], here we use the term "gist" to represent a low-dimensional (compared with the raw image pixel array) scene representation feature vector which is acquired over very short time. In our target detection scenario, this feature vector is computed for every image chip, and we explore how well it may represent the overall information of the chip so as to support classification (e.g., chips containing boats might have significantly different gist signatures than chips which do not). Saliency and gist features appear to be complementary opposites [34]: saliency features tend to capture and summarize the intensity and spatial distribution of those objects within a chip which stand out by being significantly different from their neighbors, while gist features capture and summarize the overall statistics and contextual information over the entire chip. To achieve this decision making task, a Support Vector Machine (SVM) [35, 36] is adopted as the classifier, while the biologically inspired saliency-gist features are explored to form the feature vector in the feature space. The system overview diagram can be seen in Fig.1

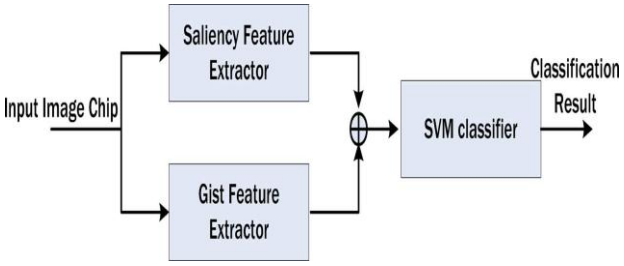

Fig. 1. Diagram of the image classification system applied to every image chip.

II. DESIGN AND IMPLEMENTATION

We compute saliency maps using several variants of the general Itti-Koch [22] architecture, and we then compute basic saliency map statistics for each variant. While in the original model only simple biological features (color, intensity, orientation) were employed, we here develop several new features which might be more effective in supporting the target/non-target classification task. Ten feature channels are adopted in this paper: intensity, orientation (0 , 45 , 90 and 135 , combined into one "orientation" channel), local variance, entropy, spatial correlation, T-junctions, L-junctions, X-junctions, endpoints and surprise. Some of these feature channels (variance, entropy, spatial correlation) are computed by analyzing 16*16 image patches, giving rise to a map that is 16 times smaller than the original image horizontally and vertically (one map pixel per 16*16 image patch). The remaining feature channels are computed using image pyramids and center-surround differences, as in the original Itti-Koch algorithm: for each of these feature channels, center-surround scales are obtained from Dyadic pyramids with nine scales, from scale 0 (the original image) to scale 8 (the image reduced by factor to 2^8=256 in both the horizontal and vertical dimensions).

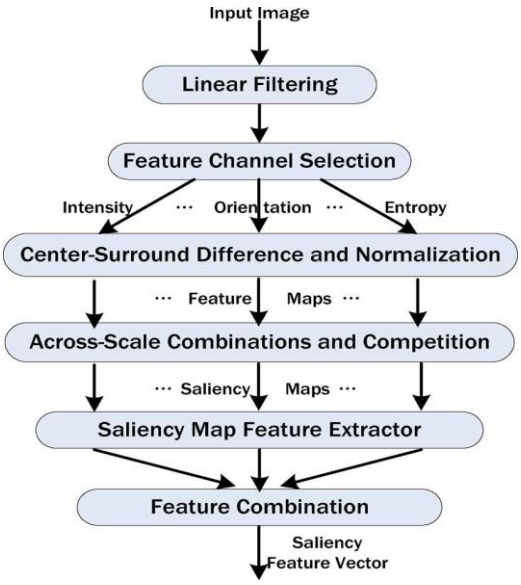

Fig.2. Block diagram of the saliency features computation model applied to every image chip

Six center surround difference maps are then computed as point-to-point difference across pyramid scales, for combination of three center scales (c={2,3,4}) and two center-surround scale differences (δ={3,4}).

Each feature map is additionally endowed with internal dynamics that provide a strong spatial within-scale competition for activity, followed by within-feature, across-scale competition. Finally, a saliency map feature extractor is applied to summarize each saliency map into a 4D vector with mean, variance, number of local maxima and average distance between locations of local maxima. All those feature vectors from the ten model variants are combined into a40D vector referred to as the "saliency features." More information about the model is described in details in the following

Intensity Channel: With the image chip as input, nine spatial scales are created using a dyadic Gaussian pyramid [18], which progressively low-pass filters and subsamples the input image, yielding

horizontal and vertical image-resolution factors ranging from 1:1 (scale zeros) to 1:256 (scale nine).

Intensity represents the amount of light reflected by the corresponding point on the object in the direction of the camera view and multiplied by some constant factor that depends on the parameters of the imaging system.

Orientation Channel: Orientation features are generally very effective feature in identifying objects, as demonstrated for example by humans' ability to understand line drawings. Here we adopt Gabor filters (θ_k=0^0,45^0,90^0,135^0)to extract the orientation feature. For each image Iintheimagepyramid the orientation feature maps can be obtained as follows [18]:

$M_{o,k}$=Gabor(I,θ_k).

Local Variance Channel: Local variance channel is used to capture local pixel intensity variance over 16 *16 image patches of the image chip of interest.

M_v(i,j)=sqrt($\sum_{sz} I^2(i,j) - S_{sz} * Mean(I_{sz}(i,j))/S_{sz} - 1$)

*EntropyChannel:*We follow the definition proposed by Privitera and Stark [38] who showed that such measure of entropy also correlates with human eye fixations. In image processing, entropy always indicates the probability distribution of the image intensity. The entropy value can be computed with the formula described in the following:

$$M_E(i,j) = -\sum_{I \in I_{sz}} p(I) * \log(p(I))$$

Where I_{sz} means the neighborhood of the pixel at (i,j) location, p(I) stands for the probability of possible intensity in its neighborhood

Spatial Correlation Channel: For two random variables X and Y, their correlation can be formulated as

$\rho X,Y = cov(X,Y)/\sigma X \sigma Y$
=E(XY)-E(X)E(Y)/$\sqrt{E(X^2)}$-E^2(X)$\sqrt{E(Y^2)} - E^2(Y)$

Junction Channels: four different kinds of junction channels are created, all included in the junction saliency map: L- junction, T-junction, X-junction and endpoint. For example, the T-junction detector will respond to 1) for an upright T, responses to the left (and from the orientation channel for horizontal orientation), right (horizontal orientation), and below (vertical orientation) the point of interest, plus 2) for a T rotated 90 clockwise, responses above, below, and to the left, plus 3) for an upside-down T, responses above, left and right, plus 4) for a T rotated 90 counter-clockwise, responsesabove, below and to the right.

Surprise Channel: We recently proposed an enhanced saliency model, which exploits a new Bayesian definition of surprise to predict human perceptual salience in space and time [39-41]. Very briefly, surprise quantifies the difference between prior and posterior beliefs of an observer as new data is observed. If observing new data causes the observer to significantly re-evaluate his/her/its beliefs about the world, that observation will cause high surprise.

Feature Maps Competition: To implement this, first normalize the feature map to a fixed range[0…M] , and then find the global maximum value M and the average value$\bar{m}$of other local maximums, finally

globally multiplying the map by(M-$\bar{m}$)^2, as was previously described in detail [18].

Saliency Map Feature Extractor: For each of the ten variants of the model, the obtained saliency map is relatively high-dimensional data (for example, a 512*512 image chip's saliency map size is 32*32=1024D), and this becomes especially true when all ten channels' saliency maps are combined.

$$m_k = \frac{1}{W * H}\sum_{i,j} SM_k(i,j)$$

$$v_k = \sqrt{\frac{1}{Sz-1}\sum_{i,j}(S\,M_k(i,j) - v_k)^2}$$

$$d_k = mean(\sqrt{(i_p - i_q)^2 + (j_p - j_q)^2})$$

p,q<n_k,p≠ q

Gist Feature Computation

The gist feature computation model [34] is related to the saliency computation model, except that it embodies concepts of feature cooperation across space rather than competition.

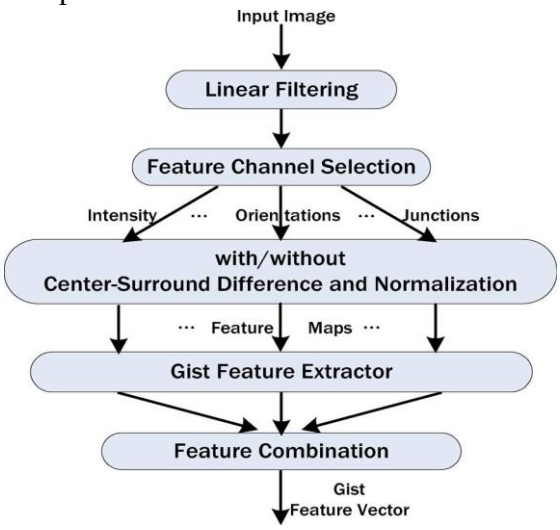

Fig. 3. Block diagram of gist features computation model applied to everyimage chip.

Unlike the saliency feature extraction model, both center-surround and raw (before center-surround) pyramid levels are exploited.

Since gist features describe an image chip's overall information, we only use mean value to represent each of the gist feature maps

$$G_{k,s,c} = \frac{1}{w*H}\sum_{i,j} GF_{k,s,c}(\text{i,j})$$

Where W and H are the gist feature map size, indices k, s, c denote feature map type, scale, center-surround type, respectively. Therefore, the gist feature vector dimensions are. Dim_{gist}= $N_{feature\ channels}$ X ($N_{Center\ Scales}$ X $N_{Surround\ Scales}$ + $N_{NoCS\ Scales}$) = 18 x (3x2+5) = 198.

In this paper, SVM provided by [36] were adopted for its easy to use. Furthermore, for the normalized input, the parameters of SVM can be optimized automatically and no tuning is needed.

III. DISCUSSION

Our results may show that the proposed algorithm performs better than the state-of-the-art (HMAX algorithm, SIFT algorithm, hidden scale salient structure algorithm and previously proposed gist algorithm alone) in difficult target search scenarios. This was achieved in situations where targets can vary greatly in their size, shape, and number of targets per chip. Taking all results together suggests that the proposed system may be further applicable to a wide range of images and target types.

The success of the proposed approach may be due to our use of two complementary sets of biologically-inspired features: gist features largely discard spatial information, while saliency features summarize it. Another mid-level features called as "gestalt" features (continuity, symmetry, closure, repetition, etc.) are added to conduct object detection.

It is likely that combining these feature types will get even better detection performance. Here we only show that the combination of gist feature and salient feature are complementary and can achieve good performance in target detection. It is possible that human saliency maps in posterior parietal cortex, the pulvinar nucleus, the frontal eye fields, or the superior colliculus [22] may also be analyzed in a holistic fashion and may contribute to the very rapid understanding of the rough layout of the scene. That is, the coarse structure of saliency maps may combine with the broad semantic information provided by the gist features to yield a coarse and rapid understanding of both a scene's gist and layout [42]. The task of creating Gestalt-based features that improve upon cutting-edge descriptors remains mostly unexplored.

The gestaltfeatures collect imageinformation that is not necessarily concentrated in space (HoG, SIFT, C1) or in scale (C1), but instead along other modes of image structure. Evidence pooled along lines or circular arrangements is combined to support hypotheses of image structures that are unlikely to be coincidental. Patchbased similarities are grouped according to spatial patterns of similarity in order to build an effective representations of repetition and symmetry.

The four feature image statistics improve substantially the performance of state-of-the-art detectors. The performance gain is consistent across several challenging real world datasets, indicating that the framework is not simply noise.

IV CONCLUSION

The proposed algorithm performs better than the state-of-the-art in difficult target search scenarios. A further improvement is expected in the device characterization, spatial uniformity, and some other factors of image quality, the accuracy of color reproduction seems to be more or less satisfactory for most color imaging applications. The visual comparison of different spectral reproduction of same chromaticity reveals that better matching of spectrum. Researchers world over are exploiting the behaviour of shock waves to develop novel experimental and modeling tools/technologies that transcend the traditional boundaries of basic sciences and engineering. Shock Waves are the only known wave phenomena in nature that have both micro and macroscopic scales in space and time.

REFERENCES

[1] Y. Amit, *2D Object Detection and Recognition, Models, Algorithmsand Networks.* Cambridge, MA: MIT Press, 2002.

[2] V. Vilaplana, F. Marques, and P. Salembier, "Binary partition treesfor object detection*," IEEE Trans. Image Process.*, vol. 17, no. 11, pp.2201–2216, Nov. 2008.

[3] R. Brunelli, *Template Matching Techniques in Computer Vision:Theory and Practice.* Hoboken, NJ: Wiley, 2009.

[4] A. Samal and P. A. Iyengar, "Automatic recognition and analysis ofhuman faces and facial expressions: A survey*," Pattern Recognit.,* vol.25, no. 1, pp. 65–77, 1992.

[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints,"*Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[6] K. Grauman and T. Darrell, "Pyramid match kernels: Discriminativeclassification with sets of image features*," in Proc. ICCV*, 2005, vol. 2,pp. 1458–1465.

[7] A. Opelt, A. Pinz, M. Fussenegger, and P. Auer, "Generic object recognitionwith boosting*," IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28,no. 3, pp. 416–431, Mar. 2006.

[8] G. Bouchard and B. Triggs, "Hierarchical part-based visual object categorization,"*in Proc. CVPR*, 2005, vol. 1, pp. 710–715.

[9] A. C. Berg, T. L. Berg, and J. Malik, "Shape matching and object recognitionusing low distortion correspondence," *in Proc. CVPR*, 2005, vol.1, pp. 26–33.
[10] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognitionin cortex," *Nature Neurosci.,* vol. 2, pp. 1019–1025, 1999.
[11] J. Mutch and D. Lowe, "Object class recognition and localization usingsparse features with limited receptive fields," *Int. J. Comput.Vis.*, vol.80, no. 1, pp. 45–57, 2008.

[12] M.Wertheimer, "Untersuchungenzurlehre von der gestalt ii," PsychologischeForschung, vol. 4, pp. 301–350, 1923, translation published in Ellis, W. (1938). A source book of Gestalt psychology (pp. 71-88). London: Routledge&Kegan Paul.

[13] "Modern theories of gestalt perception," in UnderstandingVision: An Interdisciplinary Perspective – Readings in Mind andLanguage, G. W. Humphreys, Ed. Oxford, England: Blackwell, 1992,pp. 39–70.
[14] C. Chang, H. Ren, and S. Chiang, "Real-time processing algorithmsfor target detection and classification in hyperstpectral imagery," *IEEETrans. Geosci. Remote Sens.*, vol. 39, no. 4, pp. 760–768, Apr. 2001.
[15] H. Li and J. H. Michels, "Parametric adaptive signal detection for hyperspectralimaging," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp.2704–2715, Jul. 2006.
[16] J. Lanir and M. Maltz, "Analyzing target detection performance withmultispectral fused images," *in Proc. SPIE*, 2006.
[17] S. Buganim and S. R. Rotman, "Matched filters for multispectral pointtarget detection," *in Proc. SPIE*, 2006.
[18] L. Itti, C.Koch, and E. Niebur, "A model of saliency-based visual attentionfor rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*,vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
[19] L. Itti, "Automatic foveation for video compression using a neurobiologicalmodel of visual attention," *IEEE Trans. Image Process.*, vol.13, no. 10, pp. 1304–1318, Oct. 2004.
[20] L. Itti, "Quantifying the contribution of low-level saliency to humaneyemovements in dynamic scenes," *Vis. Cogn.*, vol. 12, pp. 1093–1123,2005.
[21] J. Wolfe, "Guided search 2.0: A revised model of visual search," *Psychonom.Bull. Rev.*, vol. 1, no. 2, pp. 202–238, 1994.
[22] L. Itti and C. Koch, "Computational modeling of visual attention," *NatureRev. Neurosci.*, vol. 2, no. 3, pp. 194–203, Mar. 2001.
[23] N. Bruce and J. Tsotsos, "Saliency, attention and visual search: Aninformation theoretic approach," *J. Vis.*, vol. 9, no. 3, 2009.
[24] V. Mahadevan and N. Vasconcelos, "Spatiotemporal saliency in highlydynamic scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no.1, pp. 171–177, Jan. 2010.
[25] J. M.Wolfe and T. S. Horowitz, "What attributes guide the deploymentof visual attention and how do they do it?," *Nature Rev. Neurosci.*, vol.5, pp. 495–501, 2004.
[26] A. M. Sillito, K. L. Grieve, H. E. Jones, J. Cudeiro, and J. Davis, "Visualcortical mechanisms detecting focal orientation discontinuities,"*Nature*, vol. 378, pp. 492–496, 1995.
[27] L. Itti and C. Koch, "A saliency-based search mechanism for overtand covert shifts of visual attention," *Vis. Res.*, vol. 40, no. 10–12, pp.1489–1506, 2000.
[28] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holisticrepresentation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no.3, pp. 145–175, 2001.

[29] M. C. Potter, "Meaning in visual search," *Science*, vol. 187, no. 4180,pp. 965–966, 1975.
[30] A. Oliva and P. Schyns, "Coarse blobs or fine edges? Evidence thatinformation diagnosticity changes the perception of complex visualstimuli," *Cogn. Psychol.*, vol. 34, pp. 72–107, 1997.
[31] R. A. Rensink, "The dynamic representation of scenes," *Vis. Cogn.*, vol.7, pp. 17–42, 2000.
[32] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the humanvisual system," *Nature*, vol. 381, pp. 520–522, 1995.
[33] R. Epstein, D. Stanley, A. Harris, and N. Kanwisher, "The parahippocampalplace area: Perception, encoding, or memory retrieval?,"*Neuron*, vol. 23, pp. 115–125, 2000.
[34] C. Siagian and L. Itti, "Rapid biologically-inspired scene classificationusing features shared with visual attention," *IEEE Trans. Pattern Anal.Mach. Intell.*, vol. 29, no. 2, pp. 300–312, Feb. 2007.
[35] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines:An application to face detection," *in Proc. IEEE CVPR, 1997*, vol. 1,pp. 130–136.
[36] [Online]. Available: http://www.kernel-machines.org
[37] P. Reinagel and A. M. Zador, "Natural scene statistics at the centre ofgaze," *Network: Comput. Neural Syst.*, vol. 10, pp. 341–350, 1999.
[38] C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions-of-interest: Comparison witheye fixations," *IEEE Trans. PatternAnal. Mach. Intel*l., vol. 22, no. 9, pp. 970–982, Sep. 2000.
[39] L. Itti and P. F. Baldi, "A principled approach to detecting surprisingevents in video," *in Proc. IEEE CVPR*, 2005, vol. 1, pp. 631–637.
[40] L. Itti and P. F. Baldi, "Bayesian surprise attracts human attention,"*Adv. Neural Inf. Process.Syst.*, vol. 19, pp. 547–554, 2006.
[41] L. Itti and P. F. Baldi, "Bayesian surprise attracts human attention," *Vis.Res.*, vol. 49, no. 10, pp. 1295–1306, 2009.
[42] R. A. Rensink, "The dynamic representation of scenes," Vis. Cogn., vol.7, pp. 17–42, 2000.