2011 Canadian Conference on Computer and Robot Vision

Visual representation in the determination of saliency

Neil D. B. Bruce, Xun Shi, Evgueni Simine and John K. Tsotsos
Department of Computer Science and Engineering
York University
Toronto, Canada
Email: {neil,shixun,eugene,tsotsos}@cse.yorku.ca

***Abstract*—In this paper, we consider the role that visual representation plays in determining the behavior of a generic model of visual salience. There are a variety of different representations that have been employed to form an early visual representation of image structure. The experiments presented demonstrate that the choice of representation has an appreciable effect on the system behavior. The reasons for these differences are discussed, and generalized to implications for vision systems in general. In instances where a design choice is arbitrary, we look to the properties of visual representation in early visual processing in humans for answers. The results as a whole demonstrate the importance of filter choice and highlight some desirable properties of log-Gabor filters.**

***Keywords*-saliency; salience; linear filtering; Gabor filters; visual representation**

I. Introduction

Fast determination of visually salient targets may present an advantageous component of a combined attention and recognition system. The processing hardware involved in models that seek to characterize visual salience typically involve units that model local oriented structure in the scene. This may additionally include units sensitive to intensity, color, local contrast, edge content and a variety of other possibilities. A common element to most models that characterize saliency in some manner is a representation of oriented intensity gradients. This generally comes in the form of a bank of filters each selective for a particular combination of orientation and spatial frequency. There exist a variety of different means of constructing a representation of this form that are all similar in their nature but that differ in their precise details. It is apparent that spectral sampling may present a very important factor in determining salience. One might expect then, that the predictions of visually salient regions of a scene to depend on the precise representation used to model local oriented structure in the scene. To this end, we explore the extent to which this is the case, towards determining an optimal representation for a visual surveillance task. This is investigated in the context of the AIM model of visual saliency [1,2]. There exists prior efforts that define optimality according to a specific objective function [3,4,5]. In this case, the definition of *optimality* is in consideration to achieving even spectral coverage and insofar as it supports the estimation of salience in a visual surveillance task. While some specifics of parameterization are discussed, the emphasis is on the comparison of different classes of early Gabor-like visual filters.

The paper is formatted as follows: In section 2, a number of alternatives for representing local oriented structure are presented. This includes a look at the role that parameterization plays in dictating the nature of the filter banks in question. In section 3, experiments conducted based on visual surveillance scenarios are presented to demonstrate how the various representations differ in their ability to signal targets of potential interest to a recognition system. Finally, section 4 summarizes these results and discusses implications of the work that generalize to the modeling of salience at large.

II. Early Visual Representation

A variety of linear filters have been proposed as means of representing local frequency structure of an image. Gabor filters are ubiquitous in the modeling literature offering simultaneous localization of local structure in space and frequency. Starting with this definition, a variety of alternatives are described in this section with general practical issues that pertain to each choice being highlighted.

A. Gabor filters

Gabor filters are ubiquitous in the computer vision and image processing literature. They also figure prominently in modeling early visual representation in humans. Gabor filters produce a representation corresponding to a localized region of space and a specific spatial frequency band and orientation. The definition of a Gabor filter is as follows:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(\frac{x'^2 + \gamma^2 y'^2}{-2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right)$$

with $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$.

The aforementioned equation depicts only the real part of the filter with the cos becoming a sin in the imaginary counterpart. The parameters are as follows: λ is the wavelength of the sinusoidal component, θ dictates the orientation, ψ is the phase offset, σ controls the extent of the Gaussian envelope, and γ controls the aspect ratio.

978-0-7695-4362-8/11 $26.00 © 2011 IEEE
DOI 10.1109/CRV.2011.39

B. log-Gabor filters

The log-Gabor function is an alternative to the Gabor function proposed in [6]. Log-Gabor filters differ from Gabor filters in that they have a transfer function that is symmetric on a log frequency scale. For this reason, one cannot directly write the definition of log-Gabor filters in the spatial domain. Instead, log-Gabor filters are constructed in the frequency domain according to the definition:

$$G(w) = \exp\{-[log(w/w_o)]^2/2[log(k/w_o)]^2\}$$

with w_o the filters center frequency. k is chosen such that k/w_o is constant for varying choices of w_o. k/w_o effectively controls the filter bandwidth. The precise placement of bands then depends on the bandwidth and center frequencies chosen. This issue is discussed further at the end of this section.

The structure of natural images is such that the spectral power falls off as a function of $1/f$ with f being the radial frequency. This implies that a standard Gabor filter is biased in its sampling, with a higher likelihood of the filter being driven by lower spatial frequency content. In using a log-Gabor filter, one compensates for this bias resulting in a representation that is arguably a better encoding of visual content. It is interesting to note that the representation that appears in the early areas of human visual cortex is more consistent with a Gabor like representation that has a transfer function that is symmetric on a log scale [7]. This lends credence to the claim that log-Gabor filters are optimal in some sense as a representation of angular and radial frequency.

C. Difference of Gaussians

Inspired by the human visual system, a Difference of Gaussians based representation provides an alternative filter approach to extract oriented features. It has been shown [8,9] that the spatial organization of lateral geniculate nucleus (LGN) receptive fields accounts for orientation selectivity of V1 simple cells. Generally, a simple cell responds to a specific orientation from elongated patterns of converging symmetric LGN filters, which takes the form of an oriented Difference of Gaussians filter.

The center-surround receptive fields of LGN have response patterns that can be described as a Difference-of-Gaussian filter [10] given by:

$$\begin{aligned} f_{lgn}(x,y) = &\frac{1}{2\pi\sigma_c^2}\exp\{\frac{-(x^2+y^2)}{2\sigma_c^2}\} \\ &- \frac{1}{2\pi\sigma_s^2}\exp\{\frac{-(x^2+y^2)}{2\sigma_s^2}\} \end{aligned} \quad (1)$$

where σ_c and σ_s are the bandwidth (standard deviation) for the center and surround Gaussian function respectively. Through image convolution, signals containing spatial frequency confined to σ_c and σ_s are selected, which can be tuned for either magnocellular or parvocellular cells.

Orientation selectivity of a V1 simple cell can be represented by Oriented Difference of Gaussians (O-DOG) filter, which accumulates strengths over a group of LGN center-surround filters (DOGs) towards the desired orientation. The O-DOG filter is therefore defined as:

$$f_{v1}(x,y) = \sum_{\Delta_x}\sum_{\Delta_y} f_{lgn}(x+x_\theta, y+y_\theta)\cdot f_G(x,y) \quad (2)$$

where Δ_x and Δ_y denote the range of V1 receptive field with respect to number of LGN receptive fields in Cartesian coordinates. The orientation is governed by $x_\theta = \Delta_x\cos\theta + \Delta_y\sin\theta$ and $y_\theta = -\Delta_x\sin\theta + \Delta_y\cos\theta$.

Gaussian function $f_G(x,y)$ is used in (2) to attenuate responses that are far from the center of the filter, which is given by:

$$f_G(x,y) = \exp\{-(\frac{x_\theta^2}{2\sigma_x^2} + \frac{y_\theta^2}{2\sigma_y^2})\} \quad (3)$$

where σ_x and σ_y are the bandwidth (standard deviation) of the simple cell towards x and y axis respectively. The shape of the Gaussian profile is chosen to have the same orientation as the DOGs, which is governed by $x_\theta = x\cos\theta + y\sin\theta$ and $y_\theta = -x\sin\theta + y\cos\theta$.

D. Independent Component Analysis

Rather than a functional form, one can also learn a representation by optimizing an objective function

Independent Component Analysis (ICA) is one means of producing a filter set for which the response of the constituent units are as independent as possible. ICA is a form of projection pursuit that seeks basis filters that optimize an objective function. The objective is chosen to guide the algorithm towards the most independent basis set possible. This typically involves considering information theoretic quantities [11], or higher order statistics [12]. In the case of images, choosing random patches and seeking a set of basis filters to encode the patches in a manner in which the constituent filters exhibit responses that are mutually independent, results in a basis set of Gabor-like filters [13]. In the experiments considered in this paper, the basis sets are learned using a variety of algorithms using 21x21 random patches from RGB images. The details are described in the section on experimentation.

Existing instances of AIM have employed independent components towards a likelihood estimate formulated as a product of 1D marginal likelihood estimates

E. Filter banks and parameterization

One facet of the filter bank based representation produced by parametric models, that has not yet been discussed, is the role of parameters. The specific nature of the function that

characterizes content in a local neighbourhood has already been established in the preceding discussion. Also important is the set of parameters chosen to construct a filter bank. It is this set of parameters that dictates the spectral sampling including the precision of angular and radial frequency localization. The following discusses briefly steps towards achieving a sensible set of parameters for the filter banks to be used in the experimentation.

The possible space of parameters in characterizing Gabor and log-Gabor filters is large. A sensible guide for choosing these variables may correspond to values corresponding approximately to Gabor-like cells appearing in the early visual cortex of primates. This representation arguably has some optimality built into it that corresponds to the structure of the natural world [6].

The following considerations establish in short a set of constraints that define an idealized filter bank consistent with the visual representation appearing in early visual cortex of primates (referred to as V1 from hereon). Parameters tied to Gabor filtering that are discussed in the following are described in [14]. For further detail, readers may wish to refer to this work.

Beginning with Gabor filters, one might state the following:

- The half-magnitude frequency bandwidth is 1.4 cpd for all filters. In V1, the bandwidth of cells varies from 1.7 for cells that respond to low spatial frequencies to 1.2 for cells that correspond to high spatial frequencies. Choosing 1.4 then is a good compromise without the added complexity of representing a variable half-magnitude bandwidth.
- The half-magnitude orientation bandwidth is 40 degrees. Here the variability in visual cortex is large ranging from 10 degrees to 60 degrees depending on the cell under consideration. Again, 40 degrees is a good compromise to avoid the added complexity of variable orientation bandwidths.
- One may derive the fact that the filters have an aspect ratio of 1.24 from the above. Setting the half-magnitude contour of one frequency band to the lower contour of the following frequency band ensures even coverage.

The structure of log-Gabor filters lends itself naturally to designing a filter bank according to practical considerations. For a deep discussion of these considerations, see [15]. A summary of considerations that contribute to the design of a log-Gabor filter bank follow:

- For broad bandwidth log-Gabor filters, the tails may be extended to a significant degree in the spatial domain. One might then consider the bandwidth that requires the minimum spatial filter width to ensure good spatial localization. In practice this is in the range of 1 to 3 octaves and 2.1 octaves was used in our experiments. This determination derives from considering the width needed to represent 99% of the filter's absolute area.
- The Nyquist wavelength of 2 pixels defines the minimum filter wavelength. In practice, choosing 3 prevents strong aliasing effects. [15]
- The maximum wavelength is determined by an upper limit chosen by the number of frequency bands represented. This may be defined implicitly based on a chosen number of scales, and the smallest filter wavelength.
- Filter bandwidth is determined by the ratio of the standard deviation of the log-Gabor's transfer function to the filter centre frequency. A value between 1 and 2 octaves is sensible and setting the ratio to 0.65 is a good practical choice. Note that 0.55 corresponds approximately to 2 octaves and 0.75 to approximately 1 octave.
- There is a tradeoff between even spectral coverage and correlation between filters. The scaling factor that achieves even spectral coverage may be determined empirically and 2.1 does so for the ratio of 0.65 mentioned previously.
- The representation of angular frequency is determined by the spread of each filter and the number of orientations considered. Even coverage and correlation between filters again comes into play. By setting the ratio of the interval between orientations and the standard deviation of the angular Gaussian to 1.5 one achieves even spectral coverage.

Defining a rigorous set of criteria for the Difference of Gaussian based filters is a challenging issue. For this reason, the representation considered is crafted to maintain coverage as even as possible across the frequency plane.

This section has gone to significant lengths to outline the issues that come into play in designing a filter bank representation of the image. It is clear that there are many factors at play and one needs to be careful to ensure that the set of features produces a representation of image content that is adequate for the task at hand. In the following sections, experimentation and discussion pertaining to the impact of representation on task performance is presented.

F. Summary of early visual representations

In this section we have discussed a variety of different options for producing similar representations of local oriented intensity variation in an image. In the experimentation that follows, the set of operations that are considered includes: A bank of Gabor filters, A bank of log-Gabor filters, A Hierarchical Gabor based representation (fixed Gabor filters with varying image scale), Filters derived through ICA with RGB patches as input, Filters derived through ICA based on a YCbCr color space, and filters based on a spatially oriented sum over numerous circular symmetric DoG filters as described.

In the section that follows, the result of experiments are presented in which each of the aforementioned options forms the early visual representation used to derive a measure of visual salience based on the model described in [1,2].

III. Experimental Evaluation

Data employed in experimentation was collected from a number of elevated vantage points using a variety of cameras and varied imaging and environmental conditions. Additionally, data from aerial vehicles from a variety of public sources was used in evaluation. Qualitative evaluation was carried out on the entirety of the aformentioned data, while quantitative evaluation was performed on a representative subset of these videos for which ground truth was available. The ground truth for this data consisted of a set of bounding boxes for each frame of the video indicating the locations of pedestrians and vehicles in the video sequence. The intention of this labeling was to indicate targets that are not a fixed item or person (i.e. not part of the background). The evaluation then measures the extent to which the choice of visual representation impacts on the determination of salient targets (defined as people and vehicles) in this context. Labeling was carried out using the *SimpleLabel* labeling software developed as part of this assessment [16]. All of the results presented correspond to the output of the AIM algorithm [1,2], while varying the visual representation on which this definition is applied. The definition of saliency computed by AIM is as follows:

For a given local neighborhood of the image, one has a representation w corresponding to filters $w_1, w_2, ..., w_n$ that comprise a Gabor-like visual representation of local content. One may define the saliency according to the definition appearing in [1,2], as the $-log(p(w_1 = v_1, w_2 = v_2, ..., w_n = v_n))$ which quantifies the negative log likelihood of observing the local response vector with values $v_1, ..., v_n$ within a particular context. The presumed independence of the filter responses (originally inspired by the ICA representation) means that $p(w_1 = v_1, w_2 = v_2, ..., w_n = v_n) = \prod_{i=1}^{n} p(w_i = v_i)$. Thus, a sparse representation allows the estimation of the n-dimensional space described by w to be derived from n one dimensional probability density functions. An appropriate context may include a larger area encompassing the local neigbourhood described by w, or the entire scene in question. This latter definition of w is used in the experimentation appearing in this paper. The density estimate is computed based on a histogram estimate with 100 bins. Sample frames from the ensemble of videos used for qualitative evaluation appear in figure 1.

The quantitative assessment is based on two different standard metrics for assessing classifiers as follows: First a threshold is chosen to convert a saliency map to a binary classification. This is compared with the binary mask corresponding to the bounding boxes drawn for the same image. In the ideal case, the classification overlaps perfectly with the bounding boxes drawn. The nature of the classification produced by the saliency map depends on the threshold that is chosen. In choosing a large number of thresholds from 0 to 1, an entire performance curve may be drawn for each of the methods under consideration. The specific thresholds chosen are based on the $0th, 1st, 2nd, ..., 100th$ percentile values in the saliency map. The correspondence between the classification map and the bounding box map is carried out according to 2 separate metrics as follows.

Figure 1. A single frame from a number of the videos used in experimentation and in the qualitative assessment of algorithm performance. Videos consist of a variety of different viewpoints ranging from approximately 45 degrees from the ground, to an overhead birds-eye view.

A. ROC-curve

The ROC curve is constructed based on analysis that is done on a pixel by pixel basis. Given a particular threshold, pixels in the saliency map are set to a value of 0 or 1 (above or below threshold). The bounding box image also specifies a value of 0 or 1 for each pixel location based on its human labeled ground truth. Four different outcomes are defined as follows: TN: True negative, 0 in saliency map and 0 in bounding box map. FN: False negative: 0 in saliency map and 1 in bounding box map. TP: True positive: 1 in saliency map and 1 in bounding box map. FP: False positive: 1 in saliency map and 0 in bounding box map. Each threshold yields a set of 4 values given by these quantities. The ROC curve depicts the true positive rate versus the false positive rate. The two extremes correspond to a 0% TP and 0% FP rate and to a 100% TP and 100% FP rate. Choosing a variety of thresholds results in a smooth curve between these two extremes.

B. Precision-Recall curve

The precision-recall curve is in the same spirit as the ROC-curve. However, the precise quantity that is on display is qualitatively different. Precision refers to the fraction of pixels labeled as target by the saliency algorithm that are labeled as target in the ground truth. Recall refers to the fraction of pixels with a ground truth labeling as a target that were reported as a target by the algorithm. Therefore:

$$Precision = TP/(TP + FP)$$

and

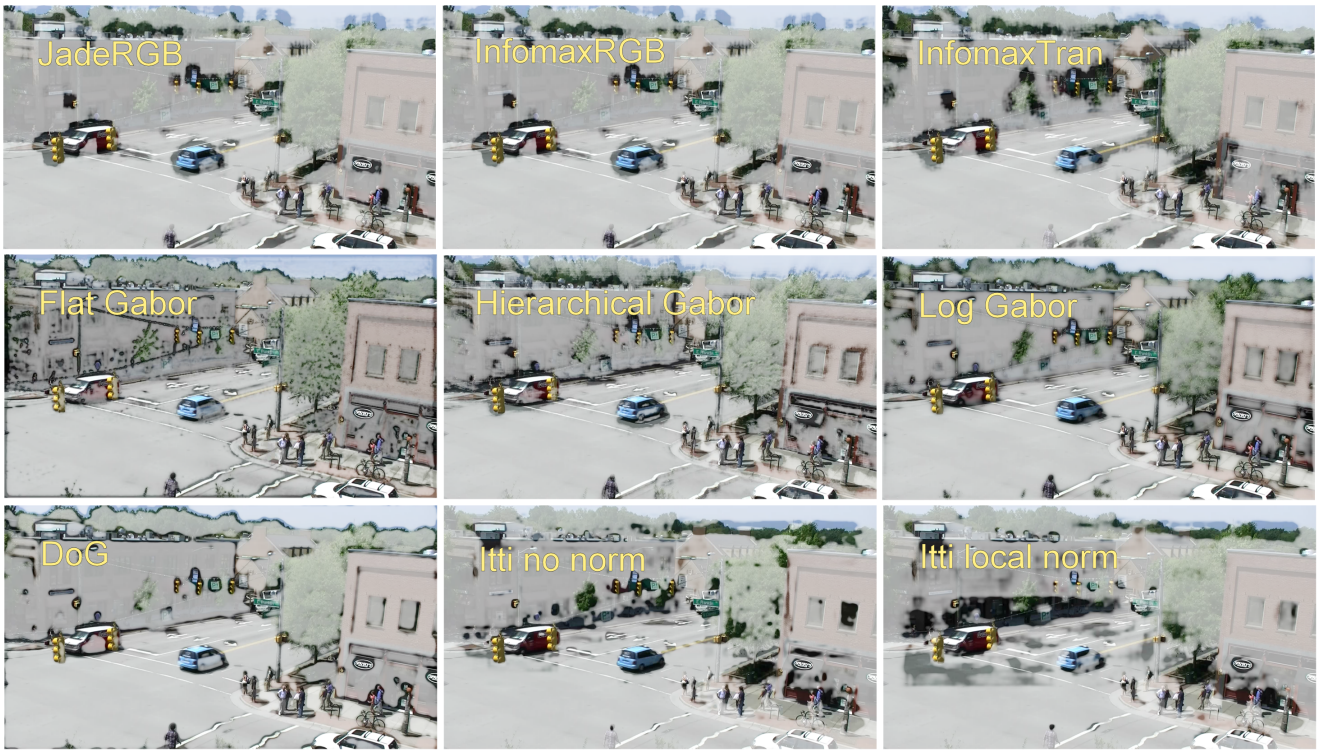

Figure 2. A depiction of the algorithmic determination of saliency attributed to the representation corresponding to the different choices of filters. The extent to which a region is *suppressed* indicates its determined salience by AIM. Also shown (bottom middle and bottom right) is the output of an alternative algorithm [17] for comparison. Regions that are non-salient are made to appear closer to white resulting in a cloudy or foggy appearance for non-salient regions. Labels are as follows: jadeRGB: ICA based filters learned using the Jade algorithm [12], infomaxRGB: ICA based filters learned using the extended infomax algorithm [11], Flat Gabor: A filter bank based on a standard Gabor based representation, Hierarchical Gabor: A filter bank based on a hierarchical Gabor wavelet decomposition, log-Gabor: A filter bank constructed using log-Gabor filters, DoG: A filter bank built as the sum of circular Difference of Gaussian filters arranged in an oriented pattern, infomaxTran: ICA based filters using the extended infomax algorithm on an alternative colorspace (YCbCr).

$$Recall = TP/(TP + FN)$$

This measure reflects the tradeoff between relevance and the proportion of positive cases retrieved.

Figure 2 depicts a representation of the saliency attributed to different regions of a frame from one of the sequences considered. The areas deemed non-salient by the algorithm are less transparent and in effect are closer to white. This then provides a depiction of the algorithms predictions localized on the image under consideration. One can see quickly from this image, that the qualitative differences between the output of the algorithms under consideration are significant. Each of the frames in figure 2 is labeled in yellow with the algorithm that depicted output corresponds to. Not shown is the output of the fasticaTran algorithm which is very similar to the infomaxTran output. At first inspection, it appears that the log-Gabor filters produce clearer localization of people and vehicles in the scene and also that spatial localization is much better than say the ICA based output. In other high contrast regions (e.g. road markings) there is also qualitative variability in the degree of attributed salience. The DoG based filters also produce output that is appealing from a qualitative perspective. A closer view of a region of this frame is depicted in figure 3. The qualitative observation of output corresponding to the complete set of aerial and surveillance videos is in agreement with these observations.

To verify that these qualitative conclusions concerning the filter outputs are correct, it is necessary to examine quantitatively the extent to which the regions deemed salient correspond to the people and cars labeled in the ground truth. An example of the ground truth for one frame of the evaluation appears in figure 4. To perform this quantitative evaluation, a representative video was selected for each of an approximately 45 degree view of the ground (as shown in figure 2), as well as a birds-eye view from above. Figure 5 demonstrates the ROC and Precision-Recall based comparisons of the various filter bank choices across several hundred frames of the video sequence. The algorithms are labeled as follows:

1) jadeRGB: ICA based filters learned using the Jade algorithm [15].
2) infomaxRGB: ICA based filters learned using the

Figure 3. A depiction of the algorithmic determination of saliency attributed to the representation corresponding to the different choices of filters. The extent to which a region is transparent indicates its determined salience by AIM. Regions heavily obscured by white are deemed non-salient by the algorithm. Labels are as in figure 2.

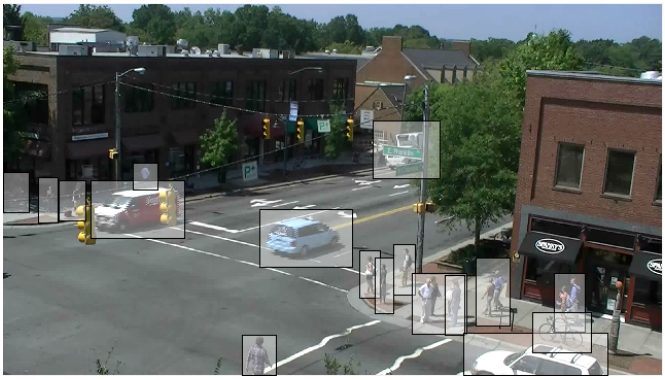

Figure 4. A depiction of the ground truth used in the quantitative assessment for one frame of the 45 degrees test video sequence. Boxed areas indicate areas where regions deemed as salient are present.

extended infomax algorithm [1].

3) Flat Gabor: A filter bank based on a standard Gabor based representation.
4) Hierarchical Gabor: A filter bank based on a hierarchical Gabor wavelet decomposition.
5) Log-Gabor: A filter bank constructed using log-Gabor filters.
6) DoG: A filter bank built as the sum of circular Difference of Gaussian filters arranged in an oriented pattern.
7) infomaxTran: ICA based filters using the extended infomax algorithm on an alternative colorspace (YCbCr).
8) fasticaTran: ICA based filters using the fastICA algo-

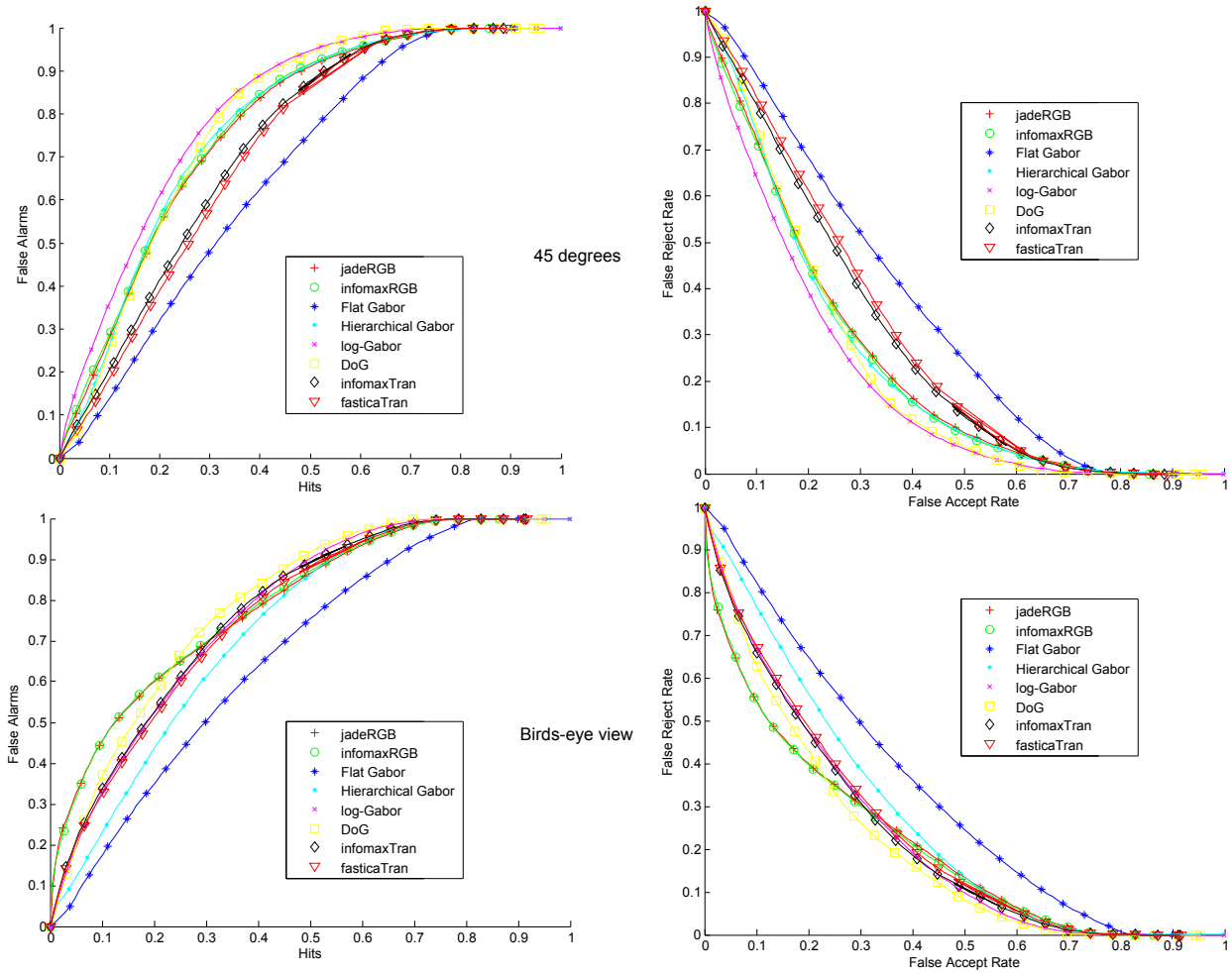

Figure 5. Top: Curves for the quantitative evaluation corresponding to the 45 degree angle intersection video sequence. Bottom: Curves corresponding to the birds-eye view video sequence. Left: ROC curves depicting the relative performance of the different algorithms. Right: Precision-Recall curves depicting the relative performance of the algorithms under consideration. Labels are the same as those listed in figure 2.

rithm [18] on an alternative colorspace (YCbCr).

The top row of figure 5 corresponds to the 45 degree angled video and the bottom row to the aerial view. On the left of each column is the ROC curve for the video under consideration and on the right is the Precision-Recall curve. As can be seen, the representation under consideration can appreciably affect algorithm performance. In general the quantitative assessment agrees with the intuition drawn from the qualitative assessment. As a whole, the log-Gabor filters are the best option for their simplistic parametric nature, and the quality of their output in the context of the algorithm under consideration. This is perhaps not surprising considering that cells in the visual cortex of humans appear to have a profile akin to log-Gabors and this conceivably yields a more even sampling of the spectral domain. It is interesting to note the apparently poor performance attributed to the standard flat Gabor filters here. This is apparently due to strong selectivity for certain frequency bands, notably on the higher spatial frequency end in this case. It is also conceivable that the performance difference might be smaller using masks that are a more precise fit to the boundaries of target items (i.e. not based on bounding boxes). In the case of the ICA based filters, the form, and also spatial localization can be poor. This is due in part to the fact that these filters have *steeper* boundaries when compared to those with a smoother Gaussian envelope that forms the spatially localized component of the filter definitions.

IV. Discussion

It is apparent from the results that the nature of early visual filtering has an appreciable impact on the deter-

mination of saliency. The intuition that derives from our extensive qualitative experimentation suggests that the nature of spectral sampling has a very important role in the determination of salient targets. From a system design perspective, representation based on independent component analysis (ICA) can be convenient since it allows the treatment of intensity variations as well as chromatic content to be dealt with using a single unified set of filters. On the other hand, in the absence of a parametric form describing the interpretive units involved in the system, it can be impossible to perform certain operations (e.g. suppressing a specific frequency band) without doing some further analysis first.

Owing to the $1/f$ structure of natural scenes, it is apparent that the sampling produced by standard Gabor filters is biased to certain radial frequencies determined by the filter parameters. When coupled with postprocessing that is highly non-linear, this bias can become exaggerated and lead to significant performance differences; this is apparently the case for the experimentation presented here. The results corresponding to log-Gabor filters appear to correspond more closely to targets of interest in a qualitative sense and this may be verified quantitatively based on the performance evaluation presented in this paper. Moreover, the fact that log-Gabor filters arguably provide a better fit for visual computation in the primate brain suggests that they may have some especially desirable properties.

The non-linear postprocessing that is performed by the model corresponds to a likelihood estimate based on a Generalized Gaussian distribution (GGD) followed by a logarithmic non-linearity. The probability density function corresponding to the response of any Gabor-like filter is fit well by a GGD. This implies that the strong sensitivity to any weak frequency bias (in the standard Gabor as opposed to the log-Gabor) is not isolated to the model at hand, but will likely arise in any model for which there is a likelihood estimate based on the response of a Gabor-like or bank of Gabor-like filters. The implication of this is that all of the issues discussed in this paper including the conclusions drawn are not system specific, but rather are issues that are deserving of close consideration in building any system that involves that analysis of outputs of Gabor or similar filters.

REFERENCES

[1] Bruce, N.D.B., Tsotsos, J.K., Saliency Based on Information Maximization. Advances in Neural Information Processing Systems, 18, pp. 155-162, June 2006.

[2] Bruce, N.D.B., Tsotsos, J.K., Saliency, Attention, and Visual Search: An Information Theoretic Approach, Journal of Vision 9:3, p1-24, 2009, http://journalofvision.org/9/3/5/, doi:10.1167/9.3.5.

[3] Dunn, D.F. and Higgins, W.E., Optimal Gabor-filter design for texture segmentation. ICASSP 0-7803-0946-4 (1993), pp. v37v40.

[4] D.A. Clausi and M.Ed Jernigan, Designing Gabor filters for optimal texture separability. Pattern Recognition 33 11 (2000), pp. 18351849.

[5] Lunscher, W.H.H.J and Beddoes, M.P., Optimal edge detector design I: parameter selection and noise effects, IEEE Trans. Pattern Analysis Mach. Intell. 8 (1986), pp. 164177.

[6] Field DJ. (1987). "Relations Between the Statistics of Natural Images and the Response Profiles of Cortical Cells". Journal of the Optical Society of America A, 4, 2379-2394.

[7] Parker, A. J. and Hawken, M. J. "Two-dimensional spatial structure of receptive fields in monkey striate cortex," J. Opt. Soc. Am. A 5, 598-605. 1988.

[8] Lampl, I., Anderson, J.S., Gillespie, D.C. and Ferster, D., Prediction of Orientation Selectivity from Receptive Field Architecture in Simple Cells of Cat Visual Cortex, Neuron, 30:1, 263–274, 2001.

[9] Somers, D.C., Nelson, S.B. and Sur, M., An emergent model of orientation selectivity in cat visual cortical simple cells, Journal of Neuroscience, 15:8, 5448-5465, 1995.

[10] Kaplan, E. and Marcus, S. and So, Y.T., Effects of dark adaptation on spatial and temporal properties of receptive fields in cat lateral geniculate nucleus, Journal of Physiology, 294:1, 561-580, 1979.

[11] Lee, T.-W., Girolami, M., Bell, A. J., Sejnowski, T. J., A Unifying Information-Theoretic Framework for Independent Component Analysis, Computers and Mathematics with Applications, 39, 1-21, 2000.

[12] Cardoso, J-F., High-order contrasts for independent component analysis, Neural Computation, 11:1, 157–192, 1999.

[13] Olshausen BA, and Field DJ. (1996). Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images. Nature, 381: 607-609.

[14] Movellan, J. R., Tutorial on Gabor Filters. Technical Report, 2002.

[15] Kovesi, P.D., MATLAB and Octave Functions for Computer Vision and Image Processing, Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia, http://www.csse.uwa.edu.au/ pk/research/matlabfns/

[16] Simine, E., SimpleLabel software package, Lab for Active and Attentive Vision, York University, 2010. http://www.cse.yorku.ca/LAAV/software/index.html

[17] Itti, L.,Koch, C., Niebur, E., A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 11, pp. 1254-1259, 1998.

[18] Hyvarinen, A. and Oja, E., A Fast Fixed-Point Algorithm for Independent Component Analysis, Neural Computation, 9(7):1483-1492, 1997.