In press, Proc. IEEE Int’l Conf. on Computer Vision and Pattern Recognition (CVPR), June 2005

A Principled Approach to Detecting Surprising Events in Video

Laurent Itti
Computer Science and Neuroscience
University of Southern California
Los Angeles, CA 90089

Pierre Baldi
Computer Science and Inst. for Genomics & Bioinformatics
University of California, Irvine
Irvine, CA 92697

Abstract

Primates demonstrate unparalleled ability at rapidly orienting towards important events in complex dynamic environments. During rapid guidance of attention and gaze towards potential objects of interest or threats, often there is no time for detailed visual analysis. Thus, heuristic computations are necessary to locate the most interesting events in quasi real-time. We present a new theory of sensory surprise, which provides a principled and computable shortcut to important information. We develop a model that computes instantaneous low-level surprise at every location in video streams. The algorithm significantly correlates with eye movements of two humans watching complex video clips, including television programs (17,936 frames, 2,152 saccadic gaze shifts). The system allows more sophisticated and time-consuming image analysis to be efficiently focused onto the most surprising subsets of the incoming data.

1. Introduction

Attention in biological and artificial systems serves to rapidly identify subsets within sensory inputs that contain important information [14, 12], in order to allocate slower processing resources [35]. Although computationally challenging given vast amounts of data carried by up to millions of sensory receptors, efficient and rapid attentional allocation is key to predation, escape, and mating – in short, to survival. Thus central to perception is providing a computationally tractable definition of “important information,” at the neuronal as well as behavioral levels.

The present study proposes such a definition under the label of “surprise” and develops a computational model to detect surprising locations in video streams, validated against eye movements of human observers.

2 Background and Rationale

Several approaches have been proposed to computationally characterize the potential behavioral importance or surprise of visual stimuli. Saliency is one such metric [15, 36, 17, 13], which defines important stimuli as statistical outliers, over the extent of an image and along one or more visual feature dimensions [12, 29]. Thus, for a stimulus to become salient, at least one of its visual attributes needs to be unique or at least rare over the entire visual scene: a red coat is perceptually salient among black suits but not among many others red coats. Computationally, detecting salient outliers may be achieved by analyzing an image along a number of feature channels, each locally sensitive to a distinct image attribute. Example features include local luminance contrast, color contrast, orientation, or direction of motion [37]. Salient locations, then, are those which elicit isolated peaks of activity in some feature map. These special locations are detected in biology *via* inhibitory spatial competition for representation within each map [31, 17, 12]. Hence, isolated active regions receive no competition and remain strong, whereas active regions within a similarly active neighborhood are inhibited. Saliency models explain a wealth of psychophysical results on human visual search behavior [34, 37, 38], and have been shown to correlate with human eye movements over static images [25]. They are also appealing because of their close link to the neurophysiology of early visual processing in the primate brain [12]. Indeed, monkey recordings have identified correlates of saliency in brain areas that include the superior colliculus [5, 22], frontal eye fields [1], V4 [21], and posterior parietal cortex [9].

Complementing this spatial definition of importance through saliency, a number of computer vision models have emphasized the temporal dimension, defining important or “novel” [2, 19] objects as those which stand out given an adaptive model of an image’s background scenery. Novelty of a stimulus is then defined by the degree to which its visual appearance does not fit the statistics of previously received image samples at a given location [6, 7]. Novelty detection starts by assuming a model for the data; for example, the distribution of a pixel’s intensity values over time may be modeled by a mixture of Gaussians [10, 20]. New data samples are then evaluated against the current model: if the probability of the observed data is low given the model, the pixel is labeled as containing a novel stimulus. The same data samples are then used to adapt the model's parameters; e.g., the means and variances of the Gaussian mixture are updated using the Expectation-Maximization (EM) algorithm [2, 33]. This approach is effective at learning scenery backgrounds, even when their temporal dynamics are not trivial. For example, trees or grass waving in the wind can be well captured; yet another gust of wind elicits little novelty whereas a pedestrian suddenly occluding a patch of tree or grass is reliably detected as novel.

Studied thus far largely independently, the notions of saliency and novelty provide two complementary answers to the question of how behaviorally important stimuli may be characterized computationally. Novelty resembles saliency in time, while saliency is somewhat like novelty over space. Below we develop a theory and model which resolve this duality, combining into a principled Bayesian notion of *surprise* the strengths of both approaches. To test our model, we analyze complex video stimuli, including television broadcast. We find that humans gaze towards surprising locations in the video streams, in a highly statistically significant manner. We finally discuss how surprise theory may find broader applicability beyond video analysis.

3. Methods: Theoretical Foundations

The proposed theory of surprise relies on a first-principles analysis, to attempt to solidify some of the more *ad-hoc* or empirically derived aspects of saliency and novelty. In particular, the within-feature spatial competition implemented by saliency models arises from an analysis of horizontal neural connections in primary visual cortices of monkeys and cats, but lacks a theoretical foundation. Conversely, novelty computation at present typically relies on an *ad-hoc* choice and parameterization of a good model for the distribution of data samples received at one location. Indeed, a model-free approach seems out of the question in the context of video processing, as it would take unreasonably too many data samples and hence unreasonably too long to accumulate sufficient data and allow accurate model-free estimation of the underlying probability density function (PDF) of the data. Yet, the rather arbitrary choice of a model for the data, often guided by intuition, prior observations, or computational complexity reasons, is problematic: what if, rather rapidly over time, the process generating the data changed in its statistical properties, requiring a new model altogether every few data samples? Or, what if multiple, contradictory models could co-exist at a given moment?

To address these issues, it is useful to go back to the foundations of our current understanding of the notion of information. Shannon's theory of communication focuses on "reproducing at one point either exactly or approximately a message selected at another point [30]." Accordingly the amount of information contained in a single dataset D is measured by the quantity $-\log P(D)$ and the average information over all datasets $\mathcal{D}$ is the entropy:

$$H(\mathcal{D}) = -\int_{\mathcal{D}} P(D) \log P(D) dD \qquad (1)$$

In this definition, the probability originates from a single observer (e.g., the Bell Labs engineer) with a single probability model. There are situations, however, characterized by the presence of multiple models or observers and where the subjective and semantic dimensions of the data are more important than its transmission. In these situations, a suitable measure of information ought to remain probabilistic in nature but to depend on the observer (be it a single cell, complex organism, or artificial entity) and its prior beliefs or expectations.

Regardless of whether one subscribes to the frequentist or subjectivist [4, 8] approach to probabilities, the fundamental impact data has on an observer is captured by Bayes' theorem for computing the posterior probability $P(M|D) = P(M)P(D|M)/P(D)$ of a hypothesis or model M given the data. In this view, the information contained in a dataset is not its entropy but rather what changes the observer's belief in M from prior $P(M)$ to posterior $P(M|D)$. Thus, a complementary way of measuring information carried by data is to measure the difference between prior and posterior distributions over the set $\mathcal{M}$ of all models, which is best done using the relative entropy or Kullback-Liebler (KL) divergence [16]. Thus surprise is defined by the average of the log-odd ratio:

$$\begin{aligned} S(D, \mathcal{M}) &= KL(P(M), P(M|D)) \\ &= \int_{\mathcal{M}} P(M) \log \frac{P(M)}{P(M|D)} dM \qquad (2) \end{aligned}$$

taken with respect to the prior distribution over the model class $\mathcal{M}$. Note that KL is not symmetric but has well-known theoretical advantages over other possible measures, including invariance with respect to reparameterizations. $KL(P(M|D), P(M))$ could also be used or the symmetric version $[KL(P(M), P(M|D)) + KL(P(M|D), P(M))]/2$, or any other measure of similarity between $P(M)$ and $P(M|D)$. While the term "surprise" has sometimes been informally used in relation to Shannon entropy, key to our new definition is how surprise requires averaging over the space of models, whereas entropy averages over data without any explicit notion of model.

Surprise can always be computed numerically, but also analytically in many practical cases, in particular those involving probability distributions in the exponential family [3] with conjugate or other priors.

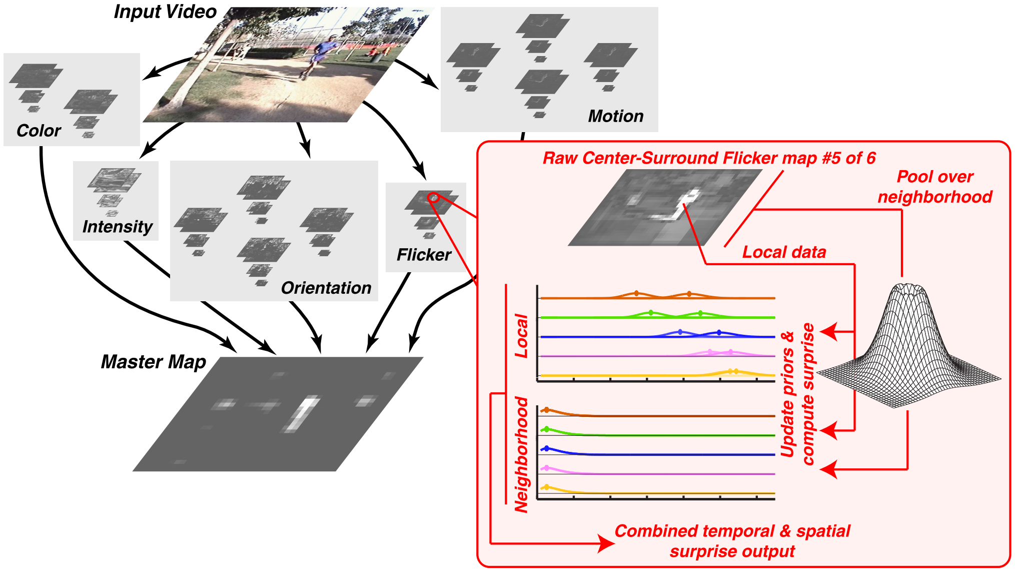
Input Video
Color
Motion
Intensity
Orientation
Flicker
Master Map
Raw Center-Surround Flicker map #5 of 6
Pool over neighborhood
Local data
Local
Neighborhood
Update priors & compute surprise
Combined temporal & spatial surprise output

Figure 1. Overview of the model. Incoming 640×480-pixel video frames are processed by five feature channels for color, flicker, etc. In each channel, six, 12, or 24 feature maps are computed using center-surround linear filters. After rescaling all maps to 40×30 pixels, in the surprise model a cascade of five surprise detectors (inset) is attached to every pixel in each of the 72 feature maps, and the resulting surprise values sum across feature channels, spatial, and temporal scales into the master map.

4. Computational Model

Armed with this theoretical framework, we can revisit our previously proposed model of saliency-based visual attention, where activity in a topographic master saliency map guides attention bottom-up [13]. The master map (40×30 lattice of temporally low-pass leaky integrator artificial neurons, given 640×480 stimuli) receives inputs from five center-surround feature channels, operating in parallel over the visual field at six spatial scales and thought to guide human attention [37, 12]: intensity contrast (six feature maps), red/green and blue/yellow color opponencies (12 maps), four orientation contrasts (24 maps), temporal onset/offset (six maps), and motion energy in four directions (24 maps), totalling 72 feature maps.

Here we retain the raw center-surround features of that model, but attach local surprise detectors to every location in each of the model's 72 neural feature maps. In addition, to replace the long-range inhibitory neural connections that give rise to phenomena of attention capture and spatial popout in the saliency model, we introduce below a second, spatial, type of surprise detectors. As individual biological neurons are unlikely to learn multimodal distributions of inputs [18], we consider unimodal model families, but multiple ones and operating at different time scales.

We model data received from feature map f at location (x, y) and time t as Poisson distributions $M(\lambda)$ (which well describe cortical pyramidal cell firing statistics [32]), parameterized by firing rate $\lambda \geq 0$. λ is estimated over the duration of each video frame (33.185 ms) as $\overline{\lambda} = f(x, y, t)$. It is important to note a central difference between surprise and previous outlier-based novelty detection approaches. Here, fitting a PDF to the data samples is a trivial step, and $\overline{\lambda}$ which parameterizes the data PDF is estimated independently at every video frame, from a sample of Poisson spikes received over the duration of the frame. Thus, when the distribution of data samples changes rapidly (e.g., from a quiescent state with low $\overline{\lambda}$ to a highly active state with very high $\overline{\lambda}$ at the next video frame), there is no inertia in our estimation of $\overline{\lambda}$. This contrasts with what would happen in outlier-based novelty models which incrementally update a longer-term estimate of the data PDF cumulated over many successive frames. Hence, a rapid change in the statistics of the data is not necessarily surprising. For this change to become surprising, it must also yield a change in the distribution of beliefs about which models are likely, as described by the prior distribution $P(M)$ over models.

Proceeding as prescribed by our surprise theory outlined above, we here consider conjugate priors, whereby the posterior belongs to the same functional family as the prior. In such case, the posterior at one frame can directly serve as prior for the next frame, as is customary in Bayesian learning. Thus, here we consider for $P(M)$ a functional form such that $P(M|D)$ has the same functional form when D is Poisson-distributed. It is easy to show that $P(M)$ satisfying this property is the Gamma probability density:

$$P(M(\lambda)) = \gamma(\lambda; \alpha, \beta) = \frac{\beta^{\alpha} \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)} \tag{3}$$

with shape $\alpha > 0$, inverse scale $\beta > 0$, and $\Gamma(.)$ the Euler Gamma function. To allow the model to detect surprise at several time scales, we implement cascades of five surprise detectors at every pixel in every feature map: the first (fastest) is updated with feature map data, and detector $i + 1$ samples from i, so that time constants increase exponentially with i. In total, the attention model comprises

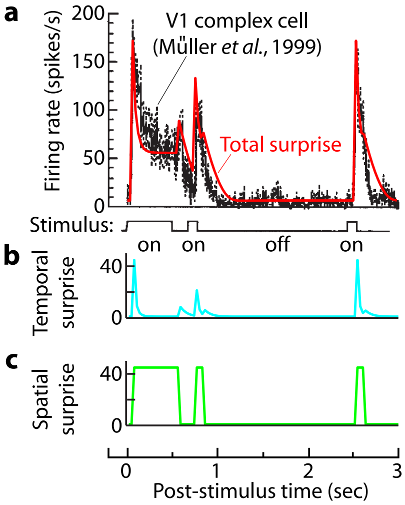

Figure 2. Combination of temporal and spatial surprises. We calibrate our model against single-neuron monkey data, then will test it against behavioral human data. **(a)** Extracellular mean firing rate (solid-black curve, ±S.D. dashed-black) of a macaque V1 complex cell during three successive brief presentations of an isolated grating stimulus demonstrates rapid adaptation (from first to second presentations) and recovery (from second to third) [23]. This suggests that cortical neurons do not passively signal Shannon information, which here would directly follow the stimulus (some information when stimulus on, none otherwise), but instead are actively affected by prior exposures to a stimulus. **(b)** Temporal surprise signals in real-time how rapidly the neuron's adapting "belief" in the presence or absence of a stimulus is changing as the stimulus is flashed on and off. Hence temporal surprise here follows stimulus transients within the receptive field, more weakly when temporally closer. **(c)** Spatial surprise signals how much a local model which hypothesizes in real-time the presence or absence of a grating stimulus disagrees with a broader model by which the display overall is blank. Hence spatial surprise here follows the stimulus. A reasonable fit of the neuron's mean firing rate (**a**; red curve) is obtained with the additive temporal/spatial surprise combination rule of Eq. 8: spatial surprise provides a sustained component of the firing rate while the stimulus is on, and temporal surprise provides firing transients at stimulus onsets and offsets.

$72 \times (40 \times 30) \times 5 = 432,000$ surprise detectors. Given an observation $D = \overline{\lambda}$ at one of these detectors and prior density $\gamma(\lambda; \alpha, \beta)$, the posterior $\gamma(\lambda; \alpha', \beta')$ obtained by Bayes theorem is also a Gamma density, with:

$$\alpha' = \alpha + \overline{\lambda} \quad \text{and} \quad \beta' = \beta + 1 \tag{4}$$

To prevent these from increasing unboundedly over time, we introduce a forgetting factor $0 < \zeta < 1$, yielding:

$$\alpha' = \zeta\alpha + \overline{\lambda} \quad \text{and} \quad \beta' = \zeta\beta + 1 \tag{5}$$

ζ preserves the prior's mean α/β but increases its variance α/β^2, embodying relaxation of belief in the prior's precision (we use $\zeta = 0.7$; see below). Local temporal surprise S_T resulting from the update can be computed exactly when using the KL divergence to quantify the differences between prior and posterior distributions over models:

$$S_T(D, \mathcal{M}) = KL(\gamma(\lambda; \alpha, \beta), \gamma(\lambda; \alpha', \beta')) \tag{6}$$

$$= \alpha' \log \frac{\beta}{\beta'} + \log \frac{\Gamma(\alpha')}{\Gamma(\alpha)} + \beta' \frac{\alpha}{\beta} + (\alpha - \alpha')\Psi(\alpha) \tag{7}$$

with $\Psi(.)$ the digamma function. For example, local temporal surprise arises when new observations are received such that an image patch previously well modeled as stationary-black becomes better modeled as flickering-red.

Spatial surprise S_S is computed similarly. For every t, (x, y), f, and i, a Gamma neighborhood distribution of models is computed as the weighted combination of distributions from the next-faster local models, over a large neighborhood with two-dimensional Difference-of-Gaussians profile ($\sigma_+ = 20$ and $\sigma_- = 3$ feature map pixels). As new data arrives, spatial surprise is the KL between prior neighborhood distribution and the posterior after update by local samples from the neighborhood's center. For example, spatial surprise arises when a model of the entire image as stationary-black must be reconsidered at some locations better modeled as flickering-red.

Our theory does not constrain how temporal and spatial surprises combine. We hence turn to empirical single-unit recordings of complex cells in striate cortex of anesthetized monkey [23]. This also allows us to set the time scale at which the model operates (defined by ζ in Eq. 5). From the fit shown in **Figure 2**, total surprise S is:

$$S = \left[S_T + \frac{S_S}{20} \right]^{\frac{1}{3}} \tag{8}$$

which results from a least-squares fit of a function of the form $S = [a_1 S_T + a_2 S_S + a_3 S_T S_S]^{a_4}$ to the neural data. Interestingly, parameter a_3 was near zero, suggesting that spatial and temporal surprise combine additively rather than multiplicatively in single neurons. We further posit that surprise combines multiplicatively across time scales, such that an event is surprising only if at all relevant time scales, allowing the model to learn periodic stimuli of various frequencies. We finally assume that surprise sums across features, such that a location may be surprising by its color, motion, or other. The sum is passed through a saturating sigmoidal nonlinearity to enforce plausible neuronal firing dynamics, then provides input to the master map. **Figure 3** illustrates the model's internals with simple stimuli.

5. Experimental Validation

We compare spatiotemporal distributions of surprise in the model's master map to eye-movement recordings from two naïve human observers watching 20 video clips, including eight outdoors scenes, four video games, and eight television programs including newscast, sports, and commer-

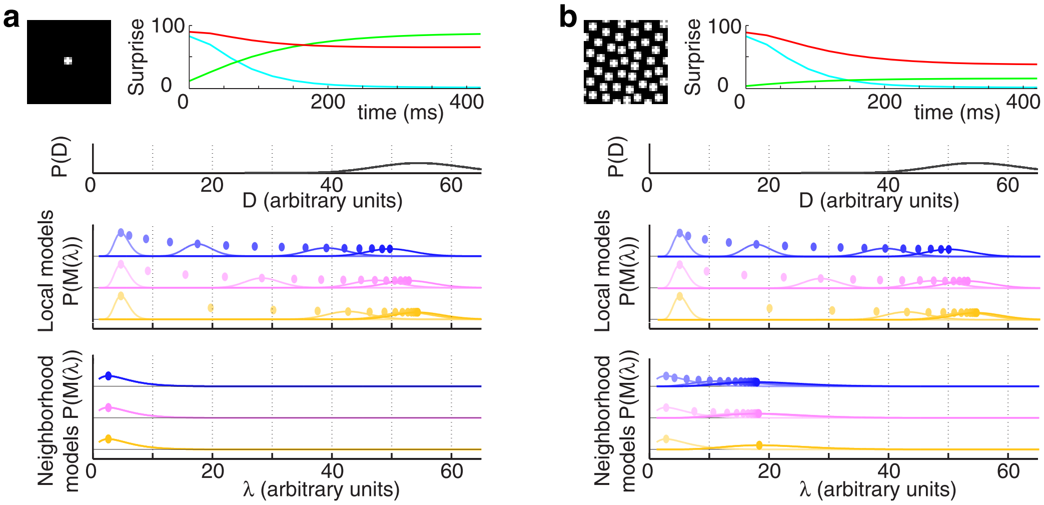

Figure 3. Model behavior with simple stimuli presented for 15 frames (450 ms). A feature map is considered (image insets) corresponding to: **(a)** a small stationary isolated stimulus on black background, and **(b)** a stationary array of small stimuli. Behaviors of local and neighborhood models at the center location are shown. In each panel, the top graph shows normalized local temporal surprise (cyan), spatial surprise (green) and total surprise (red) over time. $P(D)$ represents the Poisson data distributions received at the center location (one black curve per frame; all identical for these stationary stimuli). The Gamma prior distributions $P(M(\lambda))$ over local and neighborhood models are shown, considering $n = 3$ cascaded surprise detectors for clarity although our full model uses $n = 5$ (fastest in orange, second in purple, third in blue; more saturated colors correspond to later frames; only four Gamma distributions for frames 1, 5, 10, and 15 are plotted for clarity, with means of the others plotted as small ellipses; initial condition is a black-image prior). **(a)** Local priors quickly adapt towards the mean of the locally received data samples, generating decaying local temporal surprise. Neighborhood priors remain unchanged given the black background, generating increasing spatial surprise, as local and neighborhood priors increasingly differ. **(b)** The local situation is identical but neighborhood priors now quickly adapt towards an average data value. Hence, spatial and total surprises are lower than in **(a)**, allowing the model to predict pop-out.

cials (between 164 and 2,814 frames per clip, 17,936 frames or about 10 minutes in total). Obviously, bottom-up sensory processing may only contribute a fraction among all competing influences on attentional allocation [24, 11, 28]. Nevertheless, models computing local image information in Shannon's sense predict human gaze fixations significantly more reliably than chance [27, 26, 25], providing a challenging baseline for our model. Gaze was recorded with a 240 Hz infrared-video-based eye-tracker (ISCAN, Inc. model RK-464) **(Figure 4)**, yielding 2,152 saccadic gaze shifts. To determine whether humans preferentially oriented towards surprising stimulus elements, at onset of every human saccade we sampled (circular aperture, diameter $5.6°$) model-predicted master map activity around the saccade's future endpoint, and around a random endpoint (uniform probability). Saccade initiation latency was assumed accounted for by the master map's leaky integrators.

We quantify differences between distributions of master map samples from human and random saccades using again the KL distance: models which better predict human scanpaths exhibit higher distances from random, as observers non-uniformly gaze towards a minority of regions with highest model responses. Three models are compared, illustrated in **Figure 4**: Shannon entropy as computed by Privitera and Stark [26], saliency as computed by Itti and Koch [12], and surprise as presented here. Surprise performed superiorly to entropy and saliency, exhibiting a stronger human bias towards surprising locations than towards entropic or salient regions. These results have been confirmed in a larger study with more subjects and video clips (manuscript submitted).

6. Discussion and Conclusion

We have proposed a method for computing low-level visual surprise in unconstrained video stimuli. The model combines a notion of local temporal surprise, inspired from previous work on novelty detection, to that of spatial surprise, inspired from previous work on saliency computation. We find that a combination of both factors yields a model that is superior to both a rather straightforward measure of local entropy over image patches, and a biologically-inspired measure of bottom-up saliency.

Applications of the surprise model include the rapid, knowledge-free pruning or compression of video streams so as to focus computational resources or transmission bandwidth onto only a few surprising locations and events. One difference between our approach and many other machine

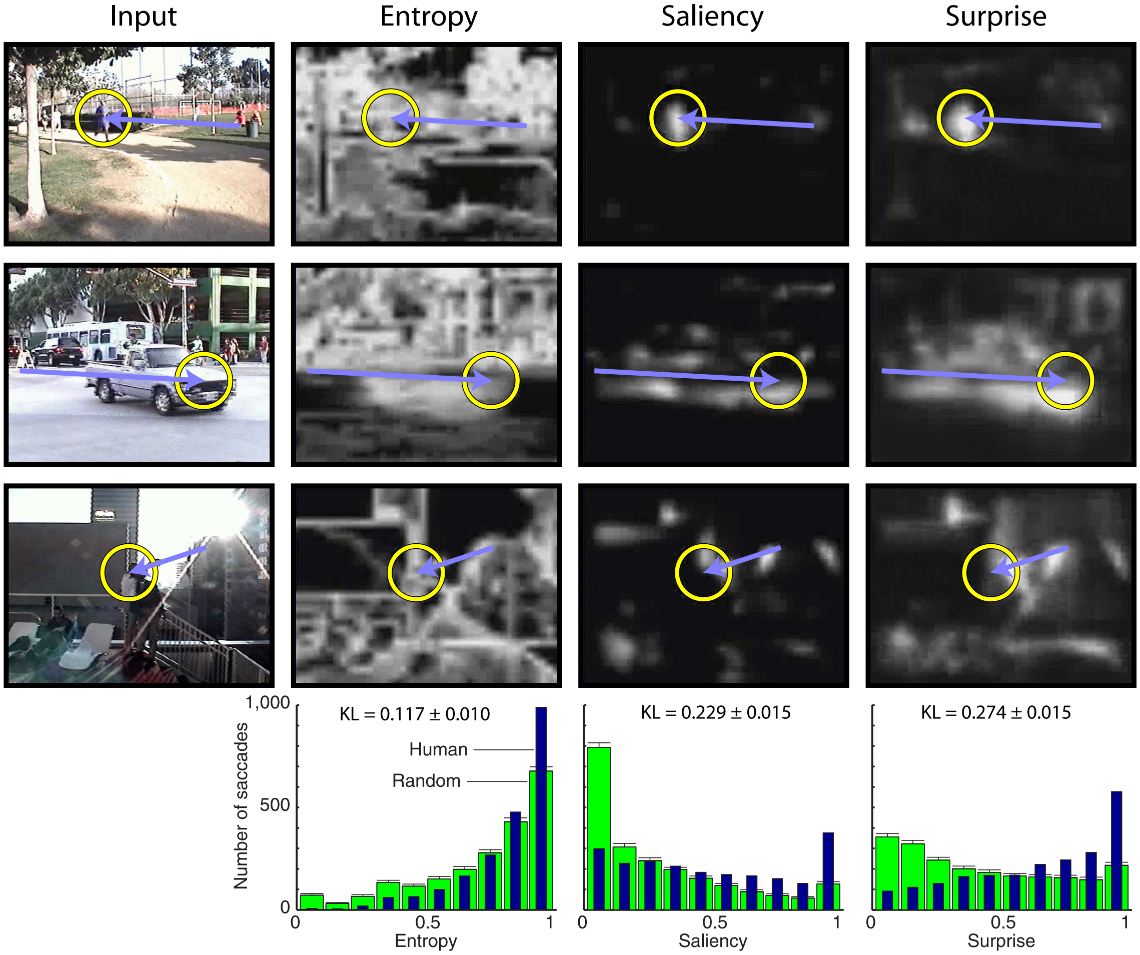

Figure 4. Sample frames from our clips (first column), with corresponding human saccades (arrows) and predictions from the entropy, saliency, and surprise models (second to fourth columns). Entropy maps exhibited many active locations, hence was poorly discriminative (lower KL score between human and random). In contrast, saliency and surprise maps were sparser and more specific. KL distances between histograms of sampled model activity at the endpoints of 2,152 human (thin blue histograms) and random (fat green histograms which control for sparseness) saccades were all significantly higher than zero, which would indicate a model not predicting human saccades better than chance (t-test, $p < 10^{-10}$ or better). KL distances all significantly differed from one another (t-tests, $p < 10^{-10}$ or better), indicating a strict ranking of entropy $<$ saliency $<$ surprise.

vision efforts is that nowhere in this work did we tune the algorithm for specific classes of objects or backgrounds. That is, using a simple yet fairly general definition of surprise we were able to reliably predict where human observers would look in complex video stimuli as diverse as outdoors scenes in parks, crowded streets, an open rooftop bar, video games including first-person and racing, television commercials, sports, news, and others. Yet, no training or software tuning was necessary to process the different stimuli.

We believe this is the first time a novelty detection algorithm is tested against human scanpaths (but see, e.g., [25] for testing of saliency models using static scenes). Obviously, our results indicate that humans did not exclusively orient towards surprising stimuli. Indeed, often they looked at locations which had been present and stationary for a while (e.g., a label indicating the name of a television speaker), which were similar to others (e.g., one of the players in a football game), or which contained no image information but were predicted to contain some in the future (e.g., the expected endpoint of the ball in a football game). Nevertheless, our results indicate that surprise accounts for human fixations highly significantly better than entropy and saliency. Inspection of our video clips and associated model predictions suggests a number of qualitative distinctions, in addition to the quantitative data reported here. Entropy was high in most locations except highly uniform ones; hence the significantly better than chance performance of the entropy model is to a large extent due merely to the fact that humans most of the time avoided empty regions in the displays (e.g., the empty sky, an empty wall). Saliency often performed similarly to surprise, suggesting that long-range inhibitory interactions may have evolved in biological cortex to compute something equivalent to our spatial surprise. Its lack of temporal dynamics, however, often made saliency inferior to surprise, in that salient stimuli would remain so for extended time periods while they would quickly become uninteresting to humans. One example includes life/health indicators in the video games, highly colorful and salient throughout the game but only fixated by humans at onset of a clip or when significant changes in their appearance over time occurred.

At the foundation of our model is a simple theory which describes a principled approach to computing surprise in data streams. While surprise certainly is not a new concept, it had lacked a formal definition, broad enough to capture the intuitive meaning of the term, yet quantitative and computable in a principled manner. The advantage of our definition is its generality which results in widespread applicability not limited to early vision. For example, a vehicle's sudden trajectory deviation and lane change on a freeway may elicit surprise that could be computed from distributions of trajectory models. Beyond vision, computable surprise could guide the development of future data mining systems, as it can in principle be applied to any type of data, including visual, auditory, or text.

Supported by NGA, NSF, and NIH.

References

[1] N. P. Bichot and J. D. Schall. Effects of similarity and history on neural mechanisms of visual selection. *Nat Neurosci*, 2(6):549–554, Jun 1999.

[2] C. Bishop. Novelty detection and neural network validation. In *Proc. IEE Conference on Vision and Image Signal Processing*, volume 141, pages 217–222, 1994.

[3] L. D. Brown. *Fundamentals of Statistical Exponential Families*. Institute of Mathematical Statistics, Hayward, CA, 1986.

[4] R. T. Cox. Probability, frequency and reasonable expectation. *Am J Phys*, 14:1–13, 1964.

[5] J. H. Fecteau, A. H. Bell, and D. P. Munoz. Neural correlates of the automatic and goal-driven biases in orienting spatial attention. *J Neurophysiol*, 92(3):1728–1737, Sep 2004.

[6] N. Friedman and S. Russell. Image segmentation in video sequences: A probabilistic approach. In *Annual Conference on Uncertainty in Artificial Intelligence*, pages 175–181, 1997.

[7] R. Gaborski, V. Vaingankar, and A. Tentler. Detection of inconsistent regions in video streams. In *Proc. SPIE Human Vision and Electronic Imaging*, 2004.

[8] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall, London, 1995.

[9] J. P. Gottlieb, M. Kusunoki, and M. E. Goldberg. The representation of visual salience in monkey parietal cortex. *Nature*, 391(6666):481–484, Jan 1998.

[10] W. E. L. Grimson, C. Stauffer, R. Romano, and L. Lee. Using adaptive tracking to classify and monitor activities in a site. In *Proc. CVPR, Santa Barbara, CA*, 1998.

[11] J. M. Henderson and A. Hollingworth. High-level scene perception. *Annu Rev Psychol*, 50:243–271, 1999.

[12] L. Itti and C. Koch. Computational modeling of visual attention. *Nat Rev Neurosci*, 2(3):194–203, 2001.

[13] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans Patt Anal Mach Intell*, 20(11):1254–1259, 1998.

[14] W. James. *The Principles of Psychology*. Harvard University Press, Cambridge, MA, 1890/1981.

[15] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.*, 4(4):219–27, 916 1985.

[16] S. Kullback. *Information Theory and Statistics*. Wiley, New York:New York, 1959.

[17] Z. Li. Contextual influences in v1 as a basis for pop out and asymmetry in visual search. *Proc Natl Acad Sci U S A*, 96(18):10530–10535, Aug 1999.

[18] L. Maffei, A. Fiorentini, and S. Bisti. Neural correlate of perceptual adaptation to gratings. *Science*, 182(116):1036–1038, 1973.

[19] M. Markou and S. Singh. Novelty detection: a review - part 1: statistical approaches. *Signal Processing*, 83(12):2481–2497, 2003.

[20] S. Marsland, U. Nehmzow, and J. Shapiro. Detecting novel features of an environment using habituation. In *Proc. Simulation of Adaptive Behavior*. MIT Press, 2000.

[21] J. A. Mazer and J. L. Gallant. Goal-related activity in v4 during free viewing visual search. evidence for a ventral stream visual salience map. *Neuron*, 40(6):1241–1250, Dec 2003.

[22] R. M. McPeek and E. L. Keller. Saccade target selection in the superior colliculus during a visual search task. *J Neurophysiol*, 88(4):2019–2034, Oct 2002.

[23] J. R. Muller, A. B. Metha, J. Krauskopf, and P. Lennie. Rapid adaptation in visual cortex to the structure of images. *Science*, 285(5432):1405–1408, 1999.

[24] D. Noton and L. Stark. Scanpaths in eye movements during pattern perception. *Science*, 171(968):308–11, 1971.

[25] D. Parkhurst, K. Law, and E. Niebur. Modeling the role of salience in the allocation of overt visual attention. *Vision Res*, 42(1):107–123, 2002.

[26] C. M. Privitera and L. W. Stark. Algorithms for defining visual regions-of-interest: comparison with eye fixations. *IEEE Trans Patt Anal Mach Intell*, 22(9):970–982, 2000.

[27] P. Reinagel and A. M. Zador. Natural scene statistics at the centre of gaze. *Network*, 10:341–350, 1999.

[28] R. A. Rensink. The dynamic representation of scenes. *Visual Cogn*, 7:17–42, 2000.

[29] P. Sajda and F. Han. Perceptual salience as novelty detection in cortical pinwheel space. In *Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering*, pages 43–46, 2003.

[30] C. E. Shannon. A mathematical theory of communication. *Bell Syst Tech J*, 27:379–423, 623–656, 1948.

[31] A. M. Sillito, K. L. Grieve, H. E. Jones, J. Cudeiro, and J. Davis. Visual cortical mechanisms detecting focal orientation discontinuities. *Nature*, 378(6556):492–6, 1995.

[32] W. R. Softky and C. Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *J Neurosci*, 13(1):334–50, 1993.

[33] L. Tarassenko. Novelty detection for the identification of masses in mammograms. In *Proc. 4th ICANN Conference*, volume 4, pages 442–447, 1995.

[34] A. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognit. Psychol.*, 12(1):97–136, 1980.

[35] J. K. Tsotsos. Computational resources do constrain behavior. *Behav Brain Sci*, 14(3):506, 1991.

[36] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. H. Lai, N. Davis, and F. Nufb. Modeling visual-attention via selective tuning. *Artificial. Intelligence.*, 78(1-2):507–45, 1995.

[37] J. Wolfe. Visual search. In H. Pashler, editor, *Attention*. University College London Press, London, UK, 1998.

[38] J. M. Wolfe and T. S. Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nat Rev Neurosci*, 5(6):495–501, Jun 2004.