Perceptual Scale Space and its Applications

Yizhou Wang[1] Siavosh Bahrami[1] Song-Chun Zhu[1,2]
Department of Computer Science[1] and Statistics[2]
University of California, Los Angeles
{*wangyz*, *siavosh*, *sczhu*}@cs.ucla.edu

Abstract

In this paper, we study a perceptual scale space by constructing a so-called sketch pyramid which augments the Gaussian and Laplacian pyramid representations in traditional image scale space theory. Each level of this sketch pyramid is a generic attributed graph – called the primal sketch which is inferred from the corresponding image at the same level of the Gaussian pyramid. When images are viewed at increasing resolutions, more details are revealed. This corresponds to perceptual transitions which are represented by topological changes in the sketch graph in terms of a graph grammar. We compute the sketch or perceptual pyramid by Bayesian inference upwards-downwards the pyramid using Markov Chain Monte Carlo reversible jumps. We show two example applications of this perceptual scale space: (1) motion tracking of objects over scales, and (2) adaptive image displays which can efficiently show a large high-resolution image in a small screen (of a PDA for example) through a selective tour of its image pyramid. Other potential applications include super-resolution and multi-resolution object recognition.

1. Introduction

In this paper, we study a perceptual scale space by augmenting the traditional image scale space representations[5, 6, 9], i.e. the Gaussian and Laplacian pyramids with a multi-level sketch pyramid illustrated in Fig. 1. Let $\mathbf{I}_0, \mathbf{I}_1, ..., \mathbf{I}_n$ be discrete levels of the Gaussian pyramid with increasing resolutions. $\mathbf{I}_k$ is a smoothed version of $\mathbf{I}_{k+1}$ by an isotropic Gaussian kernel or equivalently by running a heat diffusion process. For clarity of notation, in this paper we omit the down-sampling step and thus assume that all images are defined on the same lattice. The difference (band-pass) images $\mathbf{I}_k^+ = \mathbf{I}_{k+1} - \mathbf{I}_k$ for $k = 0, 1, ..., n-1$ form the Laplacian pyramid. When images are viewed at increasing resolutions in the Gaussian pyramid, more semantic content will be revealed. This evokes quantum jumps in visual perception amid continuous intensity changes (diffusion).

In the literature of multi-scale image representation and wavelet coding[8, 10], an image is modeled by a linear addition of independent image bases. According to this model, a new set of image bases (such as Gabor or Laplacian of Gaussian) are added when the resolution is refined. Sometimes, a Markov tree structure is assumed for the wavelet coefficients at different levels.

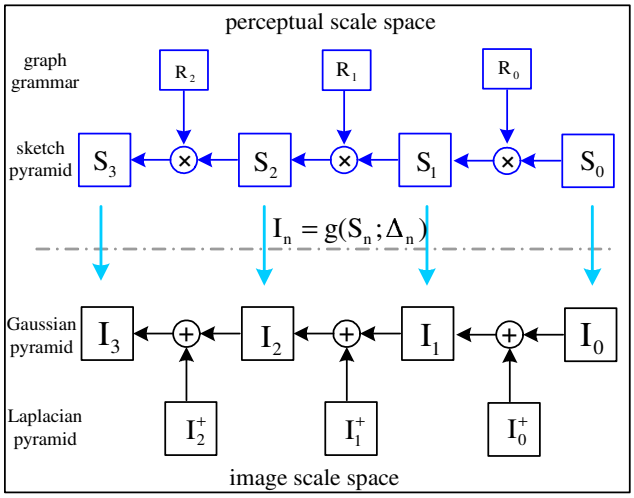

Figure 1: Augmenting the image scale space to perceptual scale space which includes a sketch pyramid and a series of graph grammars for perceptual transitions.

In this paper, we adopt a primal sketch representation studied in [2]. It divides an image into (i) structural parts (sketchable) for object boundaries and (2) textural parts (non-sketch) for stochastic textures. The structural parts are represented by a dictionary Δ of occluding image primitives (textons). Each primitive has a number of landmarks (anchor points) and includes attributes for the photometric changes and geometric warpings. It is similar to the AAM model for faces but is low dimensional and more generic. These primitives are aligned through the anchor points to form a graph representation. The texture area is summarized by some histograms of filter responses. This will yield a hidden sketch representation $\mathbf{S}_k$ which constructs the image $\mathbf{I}_k$ with dictionary Δ_k at each level.

$$\mathbf{I}_k = \mathrm{g}(\mathbf{S}_k; \Delta_k), k = 0, 1, ..., n. \quad (1)$$

where g is the generative process from sketch to image. The

perceptual transitions (jumps) over scales are represented by a set of context sensitive grammar rules which expand the graph with increasing resolution.

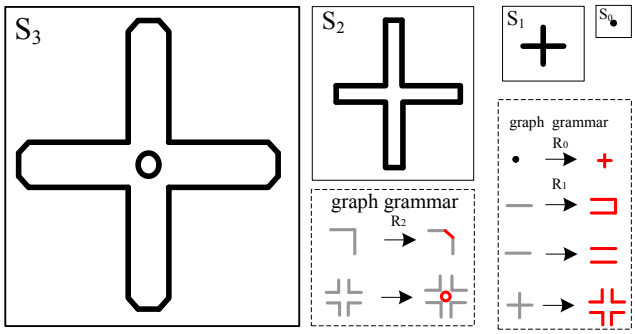

Figure 2: An example of a 4-level sketch pyramid and corresponding graph grammar for perceptual transitions.

Fig. 2 illustrates a four-level sketch pyramid $\mathbf{S}_0, \mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3$ and a series of graph grammar $\mathbf{R}_0, \mathbf{R}_1, \mathbf{R}_2$ for the graph expansion. Each $\mathbf{R}_k$ includes production rules $\gamma_{k,i}, i = 1, 2, ..., m(k)$ and each rule extends a subgraph g (could be $g = \emptyset$) conditional on its neighborhood ∂g.

$$\mathbf{R}_k = \{\ \gamma_{k,i} : g_{k,i} | \partial g_{k,i} \rightarrow g'_{k,i} | \partial g_{k,i}\ \}. \qquad (2)$$

The expansion of a graph is realized through a series of rules,

$$\mathbf{S}_k \overset{\gamma_{k,1} \cdots \gamma_{k,m(k)}}{\longrightarrow} \mathbf{S}_{k+1}. \qquad (3)$$

In this paper we study the following three issues for the perceptual scale space.

(1) Inferring of the sketch pyramid so that the graphs over scales are optimally matched and have consistent correspondence. We adopt a Bayesian framework and use Markov Chain Monte Carlo reversible jumps to compute the optimal representation upwards-downwards the pyramid to ensure consistency.

(2) Studying three categories of perceptual transitions in $\mathbf{R}_k$. (i) Sharpening of image primitives without structural changes, i.e. $\Delta_k \rightarrow \Delta_{k+1}$. See Fig. 3 for examples. (ii) Graph grammars for the graph topological changes. See Fig. 2 for examples. (iii) Catastrophic changes from texture to structures with explosive births of image primitives. See Fig. 4 for examples.

(3) Studying the criterion and mechanisms for the transitions in the context of model selection in minimum description length or maximum posterior probability.

We demonstrate two applications of the perceptual scale space.

(A) Motion tracking of objects over scales, e.g. a car driving towards the camera. Traditional motion tracking relies on fixed structures (such as contours), but tracking over scales requires the ability to account for the increasing

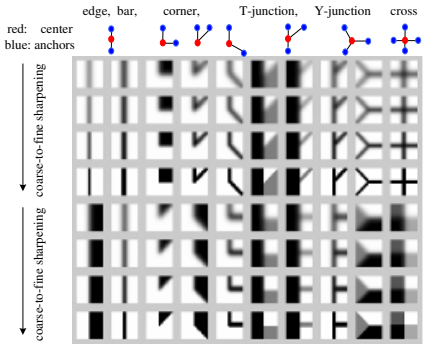

Figure 3: Image primitives in the dictionary are sharpened with 4 increasing resolutions from top to bottom.

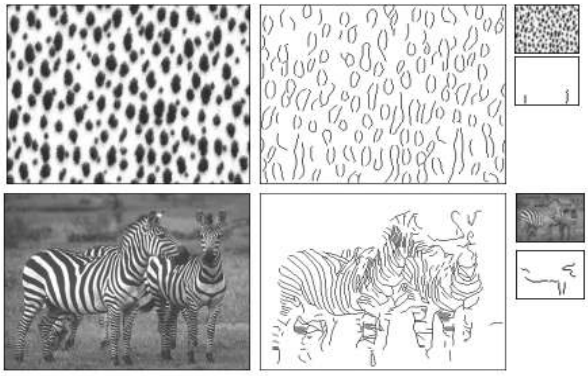

Figure 4: Catastrophic changes with explosive births of image primitives from one scale to another.

details of the object's representation. The MCMC grammar inference algorithm we describe explains the perceptual changes occurring, and computes a consistent sketch pyramid without flickering effects.

(B) Adaptive image display. The goal is to show a large high-resolution image within a small screen where different areas are shown at different resolutions to maximize the information in limited space/time. The display resolution for an area is chosen so that its sketch sub-graph no longer expands (or expands little) when the image is refined. The graph transition provides a measure (description length) of perceptual information gained from top to bottom in the pyramid. The display is stopped at a level when no new semantic content is revealed even though the image may continually be refined.

Other potential applications include super-resolution and multi-resolution object recognition[7, 4].

The perceptual sketch space should also help clarify many concepts and phenomena in vision. For example, it is known in the literature that certain image features/structures only exist within a small range of scales[13]. This is clearly demonstrated in Fig. 2 by the limited lifespan of the "dot", "cross", and "L-junctions". The ability to explicitly represent this phenomena is important for multi-scale object recognition.

The paper is organized as follows. We first introduce

the perceptual pyramid representation in Section 2, and the algorithm for inferring the sketch pyramid in Section 3. We show a number of results and applications in Section 4, and conclude the paper with a discussion in Section 5.

2. Perceptual Pyramid Representation

The perceptual scale space representation includes a pyramid of multi-level sketches $(\mathbf{S}_0, \mathbf{S}_1, ..., \mathbf{S}_n)$, and a series of graph grammar rules for perceptual transitions $(\mathbf{R}_0, \mathbf{R}_1, ..., \mathbf{R}_{n-1})$ between adjacent levels. As there are multiple paths from $\mathbf{S}_k$ to $\mathbf{S}_{k+1}$, both the optimal sketches and the transitions have to be computed together by maximum a posterior probability upwards-downwards the pyramid to form a consistent perceptual pyramid.

2.1 The primal sketch representation

We adopt a primal sketch model studied in [2] as a generic and parsimonious representation in early vision as Marr conjectured. When we talk about the perceptual transitions, we mean the changes of this generic representation without involving the concept of objects, although object templates can be represented as deformable graphs as part of the sketch.

Given an input image $\mathbf{I}$ on a lattice Λ, the primal sketch model divides it into two parts: the "sketchable" part $\mathbf{I}_{\Lambda_{\rm sk}}$ for structures and the "non-sketchable" part $\mathbf{I}_{\Lambda_{\rm nsk}}$ for textures.

$$\mathbf{I} = (\mathbf{I}_{\Lambda_{\rm sk}}, \mathbf{I}_{\Lambda_{\rm nsk}}), \quad \Lambda = \Lambda_{\rm sk} \cup \Lambda_{\rm nsk}. \tag{4}$$

The structural part assumes an occlusion model where $\Lambda_{\rm sk}$ is divided into a number of disjoint domains,

$$\Lambda_{\rm sk} = \cup_{i=1}^{N_{\rm sk}} \Lambda_{{\rm sk},i}, \quad \Lambda_{{\rm sk},i} \cap \Lambda_{{\rm sk},j} = \emptyset, i \neq j.$$

Each domain is covered by a patch from the dictionary of image primitives Δ.

$$\mathbf{I}(u,v) = \mathbf{B}_k(u,v), \quad (u,v) \in \Lambda_{{\rm sk},i}, i = 1, ..., N_{\rm sk}. \tag{5}$$

This is an occlusion model in contrast to linear additive model, and k indexes the primitives in dictionary Δ for translation x, y, rotation θ, scaling σ, photometric contrast α and geometric wrapping $\vec{\beta}$,

$$k = (x_i, y_i, \theta_i, \sigma_i, \alpha_i, \vec{\beta}_i).$$

Each primitive has a center plus 0-4 anchor points (see Fig. 3 for examples) for connections with other primitives and these points are aligned to form a sketch graph structure with attributes specifying the photometric and geometric properties. The remaining texture area $\Lambda_{\rm nsk}$ is clustered into $N_{\rm nsk} = 1 \sim 5$ homogeneous stochastic textures areas,

$$\Lambda_{\rm nsk} = \cup_{j=1}^{N_{\rm nsk}} \Lambda_{{\rm nsk},j}.$$

Each follows a MRF model (FRAME) with parameters λ_j using the structural part as boundary condition. For details of the primal sketch model, please refer to [2].

$$\mathbf{I}_{{\rm nsk},j} \sim p(\mathbf{I}_{\Lambda_{{\rm nsk},j}} | \mathbf{I}_{\Lambda_{\rm sk}}; \lambda_j), j = 1, ..., N_{\rm nsk}. \tag{6}$$

In summary, the sketch $\mathbf{S}_k$ at each layer includes all the above representation including the attributed graph and textures which we summarize as a generative model,

$$\mathbf{I}_k = {\rm g}(\mathbf{S}_k; \Delta_k), \quad k = 1, 2, ..., n. \tag{7}$$

Given $\mathbf{I}_k$, $\mathbf{S}_k$ is inferred by maximizing a posterior probability,

$$\mathbf{S}_k^* = \arg\max p(\mathbf{I}_k | \mathbf{S}_k; \Delta_k) p(\mathbf{S}_k), \ k = 0, 1, ..., n. \tag{8}$$

The key component in $\mathbf{S}_k$ is the sketch graph $G_k =< V_k, E_k >$ where V_k is the selected image primitives and E_k is the connections between adjacent primitives whose anchor points are aligned. This graph follows an inhomogeneous Gibbs model enforcing some Gestalt properties such as smoothness, continuity, and canonical junctions. Compared with the image pyramid representation with linear additive models[8], the sketch representation has two evident advantages: (1) The number of sketches used to reconstruct an image is much fewer due to hyper-sparsity of the dictionary learned from images. (2) The sketch graph topology captures properties of human perception in contrast to the independent additive image bases (wavelets). Consequently, it is more meaningful to use the sketch graph to study the perceptual transitions due to scale change.

2.2 Sketch pyramid and graph grammar

Because of the intrinsic uncertainty in the posterior probability, the sketch pyramid $\mathbf{S}_k, k = 0, 1, ...n$ will be inconsistent if each level is computed independently from Eqn. (8). For example, the graphs $G_k, k = 0, 1, ..., n$ may not have good correspondence, and this may cause a "flickering" effect when we view the sketches from coarse-to-fine (see Figs 6.b and 7.b). Therefore we must enforce steady and monotonic graph transitions over the sketch pyramid. This is realized by computing the graph grammars.

We denote the discrete Gaussian pyramid by $\mathbf{I}[0, n] = (\mathbf{I}_0, ..., \mathbf{I}_n)$ and the sketch pyramid by $\mathbf{S}[0, n] = (\mathbf{S}_0, ..., \mathbf{S}_n)$. The transition from $\mathbf{S}_k$ to $\mathbf{S}_{k+1}$ is a generative model with a sequence of $m(k)$ production rules $\mathbf{R}_k$,

$$\mathbf{R}_k = (\gamma_{k,1}, \gamma_{k,2}, ..., \gamma_{k,m(k)}). \tag{9}$$

The order of the rules matters and they constitute a path in the space of sketch graphs from $\mathbf{S}_k$ to $\mathbf{S}_{k+1}$. Each rule is applied to a subgraph $g_{k,i}$ with neighborhood $\partial g_{k,i}$ and replaces it by a new subgraph $g'_{k,i}$. So it is context sensitive. $g_{k,i}$ may be empty for some "birth" grammar, for example,

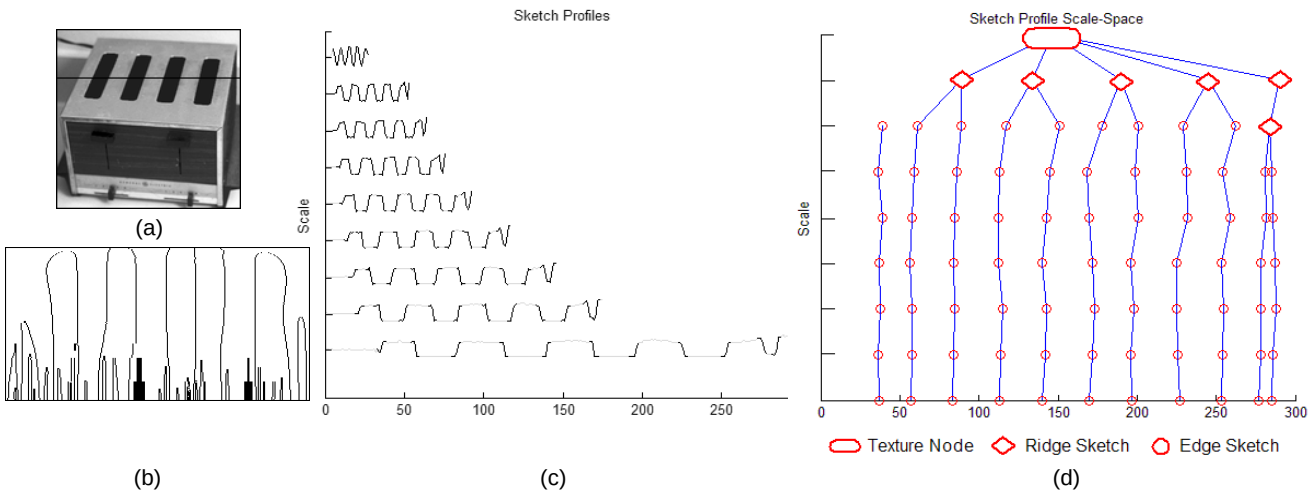

Figure 5: Scale-space of a 1D signal. (a) A 1D signal (marked as a black line) from the toaster image. (b) Trajectories of the 2nd derivative zero-crossing of the 1D signal [13]. The finest scale is at the bottom. (c) The 1D signal at different scales. The black segments on the curves correspond to the primal sketch primitives. (d) Symbolic representation of the sketch in scale-space with three types of transitions.

in which case it will add a new element to the graph. As the graph is expanding, we request $g'_{k,i}$ is no less than $g_{k,i}$.

$$\gamma_{k,i} : g_{k,i}|\partial g_{k,i} \rightarrow g'_{k,i}|\partial g_{k,i}, \quad |g'_{k,i}| \geq |g_{k,i}|. \tag{10}$$

Some generic and common grammar rules are

$$\begin{aligned} \Sigma_{gram} \quad = \quad & \{\mathcal{T}_{\emptyset}, \mathcal{T}_{bn}, \mathcal{T}_{bj}, \mathcal{T}_{ext}, \\ & \mathcal{T}_{rec}, \mathcal{T}_{spt}, \mathcal{T}_{cspl}, \mathcal{T}_{bcata}, ...\} \end{aligned}$$

They stand respectively for null operation (no topology change), birth of a node, birth of a junction, extending a node, splitting a ridge terminator into a pair of step-edges with a set of corners, splitting a ridge into a pair of step-edges, split of cross to L-junctions, catastrophic birth event, and catastrophic death event, etc.

Each rule is associated with a probability depending on its attributes,

$$\gamma_{k,i} \sim p(\gamma_{k,i}) = p(g_{k,i} \rightarrow g'_{k,i}|\partial g_{k,i}). \tag{11}$$

Therefore we have a probability for the transition from $\mathbf{S}_k$ to $\mathbf{S}_{k+1}$,

$$p(\mathbf{R}_k) = p(\mathbf{S}_{k+1}|\mathbf{S}_k) = \prod_{i=1}^{m(k)} p(\gamma_{k,i}) \tag{12}$$

These probabilities $p(\gamma_{k,i})$ were obtained by maximum liklihood estimate. Seven graduate students with and without computer vision background labeled graph transitions in 50 images from the Corel database. Some examples of learning stochastic graph grammars can be found in [12]. We now briefly discuss the three types of graph transitions, as Figs. 2 and 5 illustrate.

(i) Sharpening of image primitives without structural changes $|g_{k,i}| = |g'_{k,i}|$. Only replace the image primitives from a blurred dictionary Δ_k to a dictionary Δ_{k+1}. Fig. 3 illustrates some examples of the continuous sharpening of primitives (edges, bars, junctions). This could be used for image enhancement and super-resolution with limited scale changes. Fig.5.c also shows the continuous increase in contrast of the step edges in the 1D profile.

(ii) Graph grammars for mild changes in the graph topology. Fig. 2 shows the examples of switching from a blob to a cross, then the split of a bar to double step-edges, and terminators to two L-junctions, etc. Each time the expansion reveals more details. This part is crucial for formulating a robust super-resolution framework that moves beyond simple sharpening to hallucinating generic topological structures.

(iii) Catastrophic changes from texture to texton with explosive births of image primitives. Fig. 4 shows two examples where increasing the scale evokes the perception of many blobs/stripes which cannot be seen at a lower resolution.

In Fig.5.c the four bars in the toaster are born at one scale and each bar is further split into two step edges. The tree structure remains unchanged afterwards with only edge sharpening with increasing resolution.

In summary, our goal is to infer the sketch pyramid together with the optimal path of transitions by maximizing a Bayesian posterior probability,

$$\begin{aligned} & (\mathbf{S}[0,n], \mathbf{R}[0,n-1])^* \\ = \quad & \arg\max p(\mathbf{S}[0,n], \mathbf{R}[0,n-1]|\mathbf{I}[0,n]) \end{aligned}$$

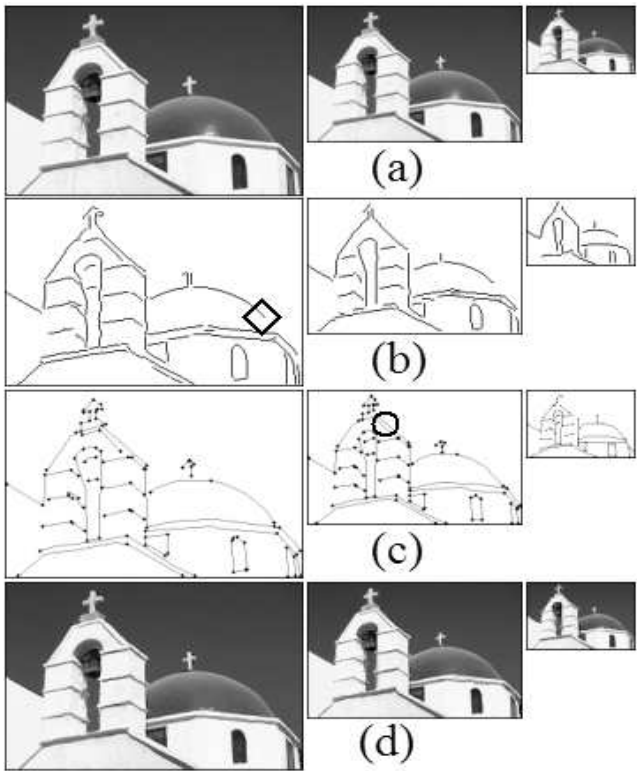

Figure 6: A church image in scale-space. (a) Original images over scales. The largest image size is 241×261. (b) Initial sketches computed independently at each level by algorithm[2]. (c) Improved sketches across scales. The dark dots indicate end points, corners and junctions. (d) Synthesized images by the sketches in (c). The symbols mark the perceptual transitions denoted in Section 3.

$$= \arg\max \prod_{k=0}^{n} p(\mathbf{I}_k|\mathbf{S}_k;\Delta_k) \cdot p(\mathbf{S}_0) \prod_{k=1}^{n} \prod_{j=1}^{m(k)} p(\gamma_{k,j}).$$

Figs. 6.c and 7.c show examples of the inferred sketch pyramids, and we compare them with the initial sketches (b) where each level is computed independently. The improved results show consistent graph matching over scales. We discuss the algorithm shortly.

2.3 Sketch transition as model comparison

A central issue for computing the sketch pyramid and the perceptual transitions is to decide which structure should appear at which scale. In other words, we should study the criterion or mechanism for the transitions. This is a typical model comparison problem, and can be handled in the Bayesian framework.

By induction, suppose $\mathbf{S}_k$ is the optimal sketch from $\mathbf{I}_k$ computed from level 0 to k. At the next level, image $\mathbf{I}_{k+1}$ has increased resolution due to the addition of the Laplacian band image $\mathbf{I}_k^+$. Let $\mathbf{S}_k^+$ be the new structures introduced (including the three types of transitions). Therefore we compare the ratio of the posterior probabilities over $\mathbf{S}_k$ and $(\mathbf{S}_k, \mathbf{S}_k^+)$.

$$\begin{aligned}\delta(\mathbf{S}_k^+|\mathbf{I}_{k+1}) &= \log \frac{p(\mathbf{S}_k, \mathbf{S}_k^+|\mathbf{I}_{k+1})}{p(\mathbf{S}_k|\mathbf{I}_{k+1})} \\ &= \log \frac{p(\mathbf{I}_{k+1}|\mathbf{S}_k, \mathbf{S}_k^+)}{p(\mathbf{I}_{k+1}|\mathbf{S}_k)} + \log p(\mathbf{S}_k^+|\mathbf{S}_k).\end{aligned}$$

The first term above (log-likelihood ratio) is usually positive for a good choice of $\mathbf{S}_k^+$ because an augmented generative model will fit the image better, and the prior term $\log p(\mathbf{S}_k^+|\mathbf{S}_k)$ is negative to penalize complex models. Therefore $\mathbf{S}_k^+$ is accepted if $\delta(\mathbf{S}_k^+|\mathbf{I}_{k+1}) > 0$. Thus a new feature is introduced at level $k+1$ if and only if the following is true

$$\delta(\mathbf{S}_k^+|\mathbf{I}_{k+1}) > 0 \text{ and } \delta(\mathbf{S}_k^+|\mathbf{I}_k) < 0. \quad (13)$$

At the top level $k = 0$, each pixel summarizes thousands of pixels at the bottom, by the central limit theorem, we assume $\mathbf{I}_0$ to be an iid Gaussian model and $\mathbf{S}_0$ has an empty sketch.

3. Upwards-downwards Inference

In this section, we briefly introduce the algorithm that infers the sketch pyramid upwards-downwards across levels.

The original sketch pursuit algorithm in [2] is not designed to generate sketches across a wide range of scales, so the consistency of the sketch pyramid was not considered. In the results shown in Figs. 6.(b) and 7.(b), we observe some inconsistency in the sketch graph over scales. We use the original sketch pursuit algorithm to obtain initial sketch graphs, and then adopt the MCMC reversible jumps[1] to track and edit the sketch graphs both upwards and downwards iteratively in scale-space across scales.

Our Markov chain consists of six pairs of reversible jumps as follows. They correspond to the grammar rules in Σ_{gram}.

1. $\mathcal{T}_{bn}/\mathcal{T}_{dn}$: birth/death of a node (denoted by ∇),
2. $\mathcal{T}_{bj}/\mathcal{T}_{dj}$: birth/death of a junction (denoted by ∇),
3. $\mathcal{T}_{ext}/\mathcal{T}_{shr}$: extending/shrinking a node (denoted by $\Diamond$),
4. $\mathcal{T}_{rec}/\mathcal{T}_{ecr}$: splitting a ridge terminator into a pair of step-edges with a set of corners, and its reverse operator (denoted by $\Box$),
5. $\mathcal{T}_{spt}/\mathcal{T}_{mrg}$: splitting a ridge into a pair of step-edges, and its reverse operator (denoted by $\bigcirc$),
6. $\mathcal{T}_{bcata}/\mathcal{T}_{dcata}$: catastrophic birth/death a large number of nodes (denoted by $\triangle$).

Each pair of reversible jumps is selected probabilistically and they observe the detailed balance equations. These steps simulate a Markov chain with invariant probability $p(\mathbf{S}[0,n], \mathbf{R}[0,n-1]|\mathbf{I}[0,n])$ in Eqn.(8).

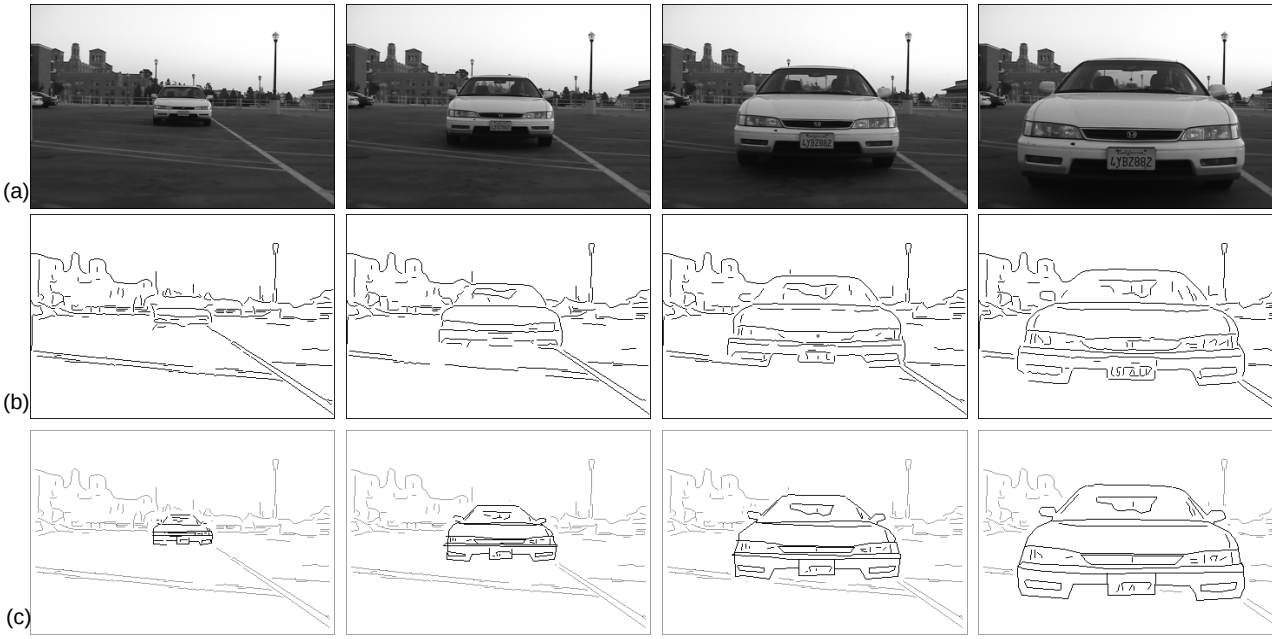

Figure 7: Car tracking sequence. (a) Sample frames from the observed sequence. The largest image size is 352×240. (b) Corresponding initial sketches from bottom-up algorithm[2]. (c) Corresponding tracked car sketches. The tracked car sketches are in black. The background sketches are in grey.

4. Two applications

We show two applications of the perceptual scale space with the sketch pyramid.

1. Multi-scale object tracking. Most work in motion tracking assumes certain object structures (like a contour[3] or small Markov graphs [15]) appear in a narrow range of scales. When the object motion occurs in a wide range of scales, we observe significant structural changes in the graph representation. This has always been considered a challenging problem in the motion tracking literature. Here we choose the example of tracking a car driving towards the camera. The results are shown in Fig. 7. The sketches for the car and background are shown in two different colors.

Scaling is one type of motion different from the traditional tracking task, as it involves a lot of photometric and topological changes. Additionally, since our sketch graphs are inferred upwards-downwards in the pyramid to maintain consistency, we can even "hallucinate" the detailed sketches of the car at a far distance. Furthermore, the background sketches are also stabilized through frames.

Before computing, the car sketches are manually labeled in the first frame of the video sequence. Then the tracking is performed by estimating the scale change of the foreground and, in the mean while, inferring the perceptual transition grammar rules. We assume the camera is at a fixed position, thus the background is still.

2. Adaptive image display. This task has recently emerged from the growing need to display large digital images (say lattice $\Lambda = 2048 \times 2048$ pixels) using a small screen (say $\Lambda_o = 128 \times 128$ pixels), such as in PDAs, cellular phones, and digital cameras[14]. Normally a user has to manually browse through an image by selecting a location and zooming to the desired level. This is inconvenient for very large images especially with today's shrinking screens. It is desirable then to present the user with a "tour" of the image that summarizes its informational content in as few frames as possible. Each frame would then be at a different location and resolution.

The problem formulation naturally leads to the sketch pyramid if we interpret informational content as sketch content. The solution then is to associate each subregion of an image with a scale such that any further zooming would not expand its sketch graph, i.e. there is no perceptual gain to further zooming. A tour would then consist of visiting these subregions at their identified scales.

For this task, we adopt a quad-tree representation with the root node representing the top level image of the sketch pyramid. As shown in Fig.8, when a node in the quad-tree is split, its 4 children represent 4 sub-images displayed at the next higher resolution. A node should not be split if no more perceptual information is available at finer scales. For example, in Fig.5.c the tree for the toaster does not change after seeing the step edges, and therefore zooming-in this part of the image is less desirable.

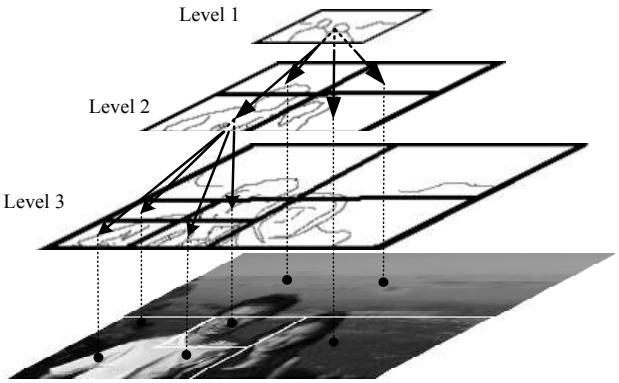

Figure 8: The image is partitioned using a quad-tree. A quad-tree node is expanded to the next level in the sketch pyramid if and only if the sketch graph expands, i.e. additional semantic/structural information appears at the next higher resolution.

A key to these tasks is to measure the "information gain" when we split a node. In the perceptual pyramid, a node v at level k corresponds to a sub-graph $\mathbf{S}_k(v)$ of the sketch, and its children at the next level correspond to $\mathbf{S}_{k+1}(v+)$. The information gain for this split is measured by

$$\delta(v) = -\log_2 p(\mathbf{S}_{k+1}(v+)|\mathbf{S}_k(v)). \quad (14)$$

That is the number of new bits needed to describe the graph expanding. As each node in the quad-tree has an information measure, we can expand the node in a sequential order until a threshold τ (or maximum number M) is reached,

$$\delta(v_1) \geq \delta(v_1) \geq \cdots \geq \delta(v_M) \geq \tau. \quad (15)$$

Fig.10 shows results of the quad-tree decomposition and multi-scale image reconstruction. The reconstructed images show that there is little perceptual loss of information if each region is viewed at its determined scale.

The information gain measure in Eqn.(14) is more meaningful than calculating the power of the bandpass Laplacian images. For example, as shown in Fig.11, a long sharp edge in an image will spread across all levels of the Laplacian pyramid, and thus demands continuous refining in the display if we use the absolute value of the Laplacian image patches. In contrast, in a sketch pyramid, it is a single edge and will stop at certain high level. In future work, we plan to integrate more meaningful information measures on subgraphs. For example, faces and texts could be emphasized more since human vision pays more attention to them.

5. Summary and Future Work

In this paper, we formulate the perceptual scale space representation and develop inference algorithms with two applications. In future research, we will further study two issues: (i) Learning a large set of graph grammars and dictionaries for natural images and objects in addition to making a richer connection to the grammar work in the 1980s literature[11]. Such grammars and dictionaries are needed for applications such as super-resolution and defining semantic image metrics, and (ii) studying the hierarchic representation of object models for multi-scale feature detection and object recognition[7, 4].

Figure 9: Visiting the nodes of a quad-tree decomposition of a sketch pyramid is an efficient way to automatically convey a large image's informational content.

Acknowledgments

This work was supported by grant NSF-IIS-0222967 and a National Science Foundation Graduate Research Fellowship. The authors also thank Ziqiang Liu for sharing part of his programming code and Dr. Xing Xie and Xin Fan for helpful discussions.

References

[1] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination", *Biometrika*, vol.82, 711-732, 1995.

[2] C.E. Guo, S.C. Zhu, and Y.N. Wu "A Mathematical Theory of Primal Sketch and Sketchability," *ICCV*, 2003. An extended version is referred to S.C.Zhu's website)

[3] M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density", *ECCV*, 1996.

[4] T. Kadir and M. Brady, "Saliency, Scale and Image Description," *IJCV*, 2001.

[5] J.J. Koenderink, "The Structure of Images," *Biological Cybernetics,*, 1984.

[6] T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers, 1994.

[7] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints,"*IJCV*, 2004.

[8] S. Mallat, "A Theory of Multi-resolution Signal Decomposition: the Wavelet Representation," *PAMI*, 11, 1989.

[9] B. M. ter Haar Romeny, *Front-End Vision and Multiscale Image Analysis: Imtroduction to Scale-Space Theory*, Dordrecht, Kluwer Academic Publishers, 1997.

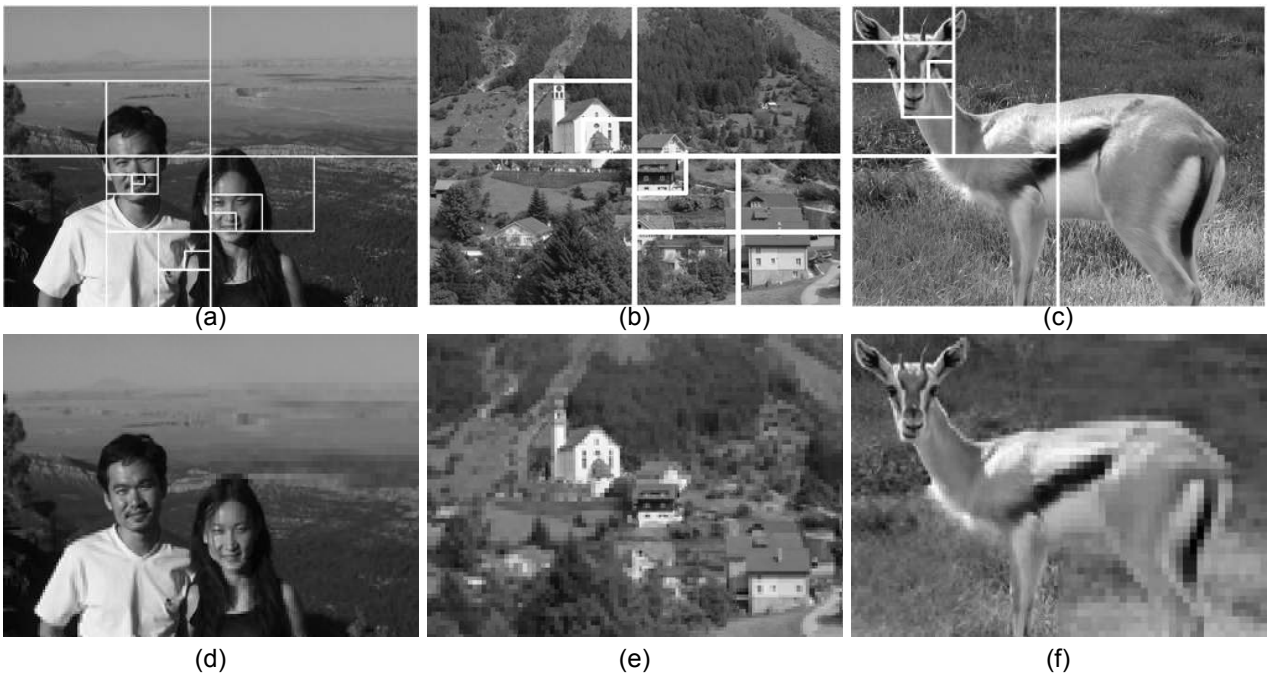

Figure 10: Images (a-c) show three sample quad-tree decompositions. The partitions correspond to regions in the sketch pyramid that experience sketch graph expansions. If the graph expands in a given partition, then we need to increase the resolution of the corresponding image region to capture the added structural information. Images (d-f) represent the replacement of each sketch partition with an image region from the corresponding level in the Gaussian pyramid. Note how the areas of higher structural content are in higher resolution (e.g. face, house), and areas of little structure are in lower resolution (e.g. landscape, grass). A limited view screen can thus view each partition at its appropriate scale without loss of structural information.

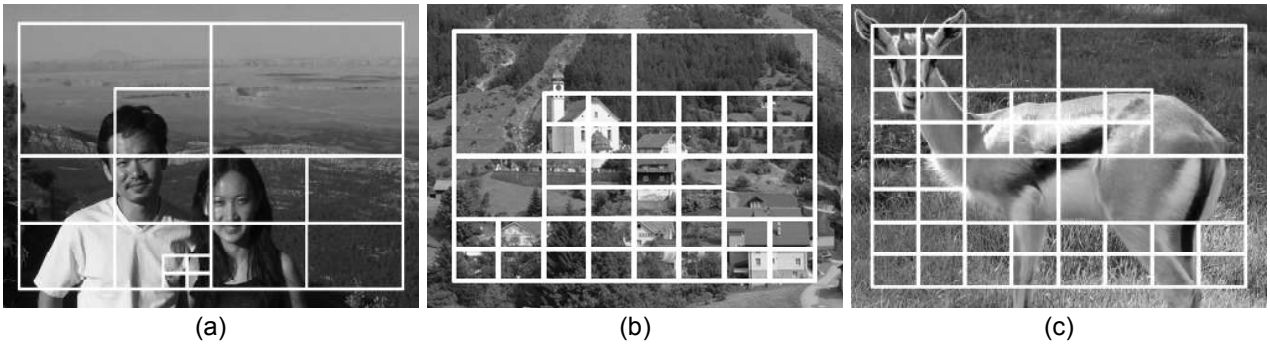

Figure 11: For comparison with the corresponding partitions in the sketch pyramid (Fig.10), a Laplacian decomposition of the test images are shown. The outer frame is set smaller than the image size to avoid Gaussian filtering boundary effects. The algorithm greedily splits the leaf nodes bearing the most power (sum of squared pixel values in the node of the Laplacian pyramid I_k^+). As clearly evident, the Laplacian decomposition does not exploit the perceptually important image regions in its greedy search (e.g. facial features) instead focusing more on the high frequency areas.

[10] E. P. Simoncelli, W. T. Freeman, E. H. Adelson and D J Heeger, "Shiftable multi-scale transforms," *IEEE Trans. Info. Theory*, 38(2):587-607, March 1992.

[11] W. H. Tsai and K. S. Fu, "Attributed grammar–A tool for combining syntactic and statistical approaches to pattern recognition," *IEEE Trans. Syst. Man Cybern.*, 1980.

[12] Y.Z Wang and S.C. Zhu, "Modeling complex motion by tracking and editing hidden Markov graphs," *CVPR*, 2004.

[13] A.P. Witkin, "Scale Space Filtering," *Proc. 8th Intl. Joint Conf. Art. Intell.*, Karlsruhe, West Germany, 1983.

[14] X. Xie et al. "Browsing large pictures under limited display sizes", *IEEE Trans. on Multimedia*, to appear.

[15] T. Yu and Y. Wu, "Collaborative Tracking of Multiple Targets," *CVPR*, 2004.