

Editorial

Patches in Vision

Simon Lucey and Tsuhan Chen

Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Correspondence should be addressed to Simon Lucey, slucey@ieee.org

Received 11 January 2009; Accepted 11 January 2009

Copyright © 2009 S. Lucey and T. Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This special issue contains extended versions of the best papers of the two “Beyond Patches” workshops we ran in 2006 and 2007 IEEE Conferences on Computer Vision and Pattern Recognition (CVPR). In addition, some specially solicited papers have also been included which were not part of these two workshops but do highlight and reinforce the motivation and philosophy of these workshops.

We refer to a “patch” agnostically as an ensemble of spatially adjacent pixels/descriptors which are treated collectively as a single primitive. Patches fall between the two extremes of individual pixels/descriptors and whole objects/images. Analyzing an image or video sequence in terms of patches, rather than individual pixels/descriptors, has some inherent advantages (i.e., computation, generalization, context, etc.) for numerous vision, image, and video content extraction applications (e.g., matching, correspondence, tracking, rendering, etc.). Common descriptors in literature, other than pixels, have been contours, shape, flow, and so forth. Additional novel applications explored in this special issue include image restoration, image compression, pixel motion, and scene recognition.

Our workshops and this special issue have been motivated by the almost ubiquitous employment of “patches” in recent years across the vision community. The papers included in this special issue touch upon many of the benefits of patch-based representations in vision, image, and video processing.

Gupta and Huang proposed a unique approach to image restoration that leverages a multilayer “patch-based” graphical model which unifies the low-level vision task of restoration and the high-level vision task of recognition in a cooperative framework. In their approach, they modeled images as MRFs over a patch-based representation. Through the incorporation of two spatial domain methods, they argue that it is possible to move toward the idea that high-level concepts like recognition can be used to aid low-level

operations like restoration. To validate this argument, they introduce a transformed domain method analogous to the spatial domain patch-based MRF and implement the system for removing compression artifacts from images and videos.

Chandler et al. demonstrate a unique method for measuring the capacity of natural image patches for visual masking. Their central thesis is that the current state-of-the-art models of visual masking have been optimized for artificial targets placed upon unnatural backgrounds. To circumvent this problem, they (i) measure the ability of natural-image patches in masking distortion, (ii) analyze the performance of a widely accepted, standard masking model in predicting these data, and (iii) report optimal model parameters for different patch types (textures, structures, and edges).

A robust algorithm for subpixel motion estimation is proposed by El Mehdi et al. In the work entitled “A Robust Sub-Pixel Motion Estimation Algorithm Using HOS in the Parametric Domain,” a class of algorithms is presented that estimate the displacement vector \mathbf{eld} (DVF) from two successive image frames. It is well understood that in severely corrupted image sequences, second-order statistic (SOS) methods do not work well. Instead, the authors propose using the bispectrum in the parametric domain. The displacement vector of a moving object is estimated by solving linear equations involving third-order hologram and the matrix containing Dirac delta function. Results are presented that demonstrate the utility of this approach on noisy image sequences.

Sluzek in the paper entitled “Building Local Features from Pattern-Based Approximations of Patches: Discussion on Moments and Hough Transform” overviews the concept of using circular patches as local features for image description, matching, and retrieval. The authors base their work on the concept that humans recognize known objects by identifying certain classes of geometric patterns that

are combinations of contour and region properties. Such patterns may have diversified shapes, but all instances of the same pattern have the same structural composition that can be parameterized. The main assumption is that patches of interest correspond to certain geometric patterns that may exist within analyzed images. Even if the image is noised or distorted, the patterns (if prominent enough) are still clearly seen even though their visual appearances are corrupted.

A novel approach to scene classification is described by Monay et al. in the paper entitled “Contextual Classification of Image Patches with Latent Aspect Models” which combines patch-based contextual classification with latent aspect models. In their approach they explore the incorporation of context in two ways: (i) by using the fact that specific learned aspects correlating with the semantic classes, which resolves some cases of visual polysemy often present in patch-based representations, and (ii) by formalizing the notion that scene context is image-specific (i.e., what an individual patch represents depends on what the rest of the patches in the same image are). We demonstrate the validity of our approach on a man-made versus natural patch classification problem.

Finally, Parikh and Chen in the paper entitled “Unsupervised Modeling of Objects and Their Hierarchical Contextual Interactions” outline a method for unsupervised modeling of objects and their hierarchical contextual interaction. They propose a method for analyzing the interactions among patches across a collection of images. They motivate this method by the observation that analyzing the interactions among these objects can allow for a semantically meaningful grouping that characterizes the entire scene. These groupings are typically hierarchical. As a result, hierarchical semantics of objects (hSOs) is introduced to attempt to capture these hierarchical groupings.

To conclude, we would like to thank the authors, reviewers, and the editorial team of the EURASIP Journal on Image and Video Processing for their effort in the preparation of this special issue. It is our hope that this special issue, in some small way, can help open up a dialogue between researchers in the community to answer some of the deeper remaining questions concerning patches in vision.

*Simon Lucey
Tsuhan Chen*