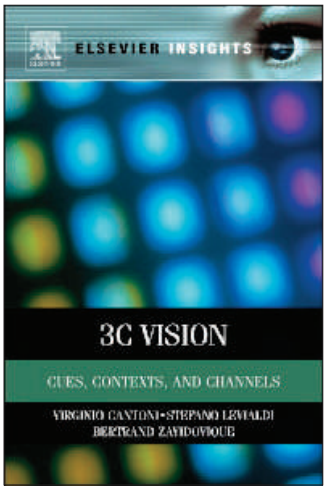
Book Review

3C Vision: Cues, Context, and Channels
Virginio Cantoni, Setfano Levialdi, and Bertrand Zavidovique
Elsevier Inc., Waltham, MA, USA, First Edition 2011
ISBN: 978-0-12-385220-5, $49.95 USD, Hardcover: 254 pages, New

Submitted by Erik Blasch

Erik Blasch

In future *aerospace technology*, there is increasing emphasis on automated systems, such as unmanned aerial vehicles (UAVs) and remotely piloted vehicles (RPVs) that use sensor processing of electro-optical sensors (i.e., visual cameras), operate over varying environments, and support user targeting from multimedia sources. This state-of-the-art comprehensive text in *computer vision* (CV) provides a foundation for numerous aerial and system-level information fusion designs based on contemporary sensing modalities for *cue, context, and channel* (3C) processing. 3C vision and analytical strategies are proposed that give the reader a taxonomy of concepts, mathematical processing strategies, and intuitive illustrations to explain the concepts.



Most existing books in CV follow one of three paradigms: biologically inspired CV algorithms [1], signal and image processing [2], [3], and application-specific algorithms. There is a much larger system-level consideration that has gone unaddressed for many years. Cameras and modern image manipulation tools have made video a more accessible and ubiquitous medium for communication. The explosive multimedia growth has closed the gap between the richness of linguistic communication, once limited to text and print, to video. Therein lies the next set of challenges: How do cultures and the 3C categories factor in the way a message is composed in a visual form? Alternatively, given a set of images of forensic nature, how much can we infer about the context from what is directly seen in the image? Standard physics-based and application-specific assumptions that were factored in a priori models and proven useful in vision will need to be revisited. Those who are practicing in this art will agree that CV has a long way to go before the already impressive machine vision algorithms are capable of matching parity with human vision. The book highlights some of the foundational ideas needed for multimedia algorithmic frameworks to support system-level processing.

The book is well organized, covering vision (including images, text, and multimedia content) considerations over the sensor, target, and environmental operating conditions. The book content is intended for a graduate student but provides fundamentals and concept overviews that only require a general knowledge of physics. The developments come from many aerospace and electronic system (AES) aspects of sensors, tracking, and web-based human-computer interaction (HCI) visualizations that support information fusion management and systems design [4]. The book builds from 1980s CV research in human perception to injection of social contextual (e.g., linguistic processing) for multimedia applications. Within each chapter, a common theme begins from an artistic perception of historical paintings and ends with contemporary issues in 3C vision. Some common themes include 3D analysis, feature extraction, graphical methods, tracking, semantics, and support for user situational awareness, understanding, and assessment. Table 1 organizes the themes in the 3C categories.

Cues (*targets*): Standard methods in image processing for detection, segmentation, and classification are presented with extensions to semantics and situation assessment, as shown in Figure 1. The cues from images and multimedia content provide a basis for 3D geometry, graphical methods of feature processing, and track analysis. Current methods that build on fundamental approaches include dense data, scene analysis, and semantic content for cause determination and decision making. The extraction of cues over locations supports a central theme throughout the book of contextual analysis.

Context (*environment*): Three methods for situational awareness contextual processing are discussed: physical (candidate generation), photogrammetric (candidate evaluation), and computational (consistency determination) from which the variable aspects are assessed for enhanced 3C visioning. For example, context (Q) includes the model set (A), the operator set (O), and the decision policy (I) parameters (variables). Direct, recognition-based, and search strategies are explained for contextual processing. Instantiating a vi-

Table 1.

Techniques Introduced in 3C Vision			
	Cues	**Contexts**	**Channels**
Vision	Concurrent basic information and relative description	Intended action, a priori knowledge, and common culture	Media to convey processed information, including the coding scheme
Analytic strategy	Capturing current scene (geometry and environment)	Discovering relations for scene understanding	Decision making and action planning (interhuman mediation)
3D	Perspective geometry (lighting, local contrast, shape, color, texture)	Image topology (location, foreground/background, pattern, or color)	Information-based, mapping, and visualizers in a web-based virtual reality environment
Features	Feature extraction (region, motion, depth) and decision (computing complexity)	Feature search among appearance of scene features, environment considerations, and processing (e.g., content-based image retrieval)	Linguistic content: document images, textual paragraphs/sentences, and pictorial representations of content
Graphical methods	Bipartite graph matching over feature space	Hypothesis testing using a pattern tree, feature graph, and grammar refinements	Social, cognitive, and sensor networks for information communication extraction and delivery
Tracking	Ego-motion, optical flow, morphological, and spatial tracking	Temporal or circumstantial instances for matching, segmenting, linguistic/syntactic, structuring, and behavior analysis	Information carriers (text, sound, images), multimedia that requires content structuring for dynamic applications
Semantics	Tree-scan grammars	Direct (cue, metadata), search, or recognition-based retrieval for linguistic/syntactic applications	Icons, metaphors, and annotations extraction from web content over images/text representative grammars
Situation	Perspective-based analysis for assessment	Knowledge support for awareness based on patterns, retrieval requests, and terrain content	Interaction through augmented reality, pictorial indexing, and multimedia content for understanding
User	User-directed scene analysis for object and semantic decision making	Attention focusing for perception, action, and cooperative data analysis over locations, activities, and events of interest	Knowledge transfer, social interaction, usability visualizations, annotations to enhance aesthetics, usefulness, and interactivity

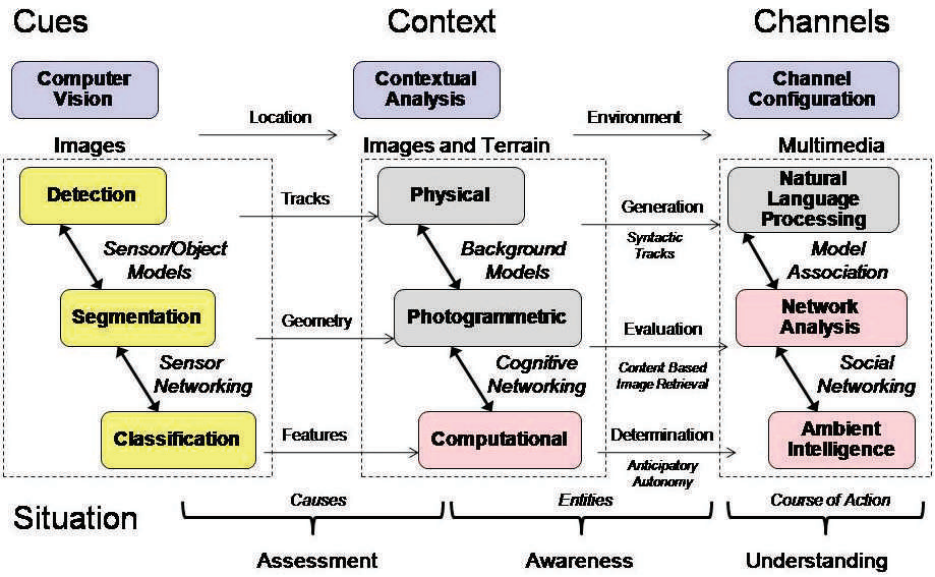


Figure 1.
Concepts introduced in 3C vision.

sual search for context includes a priori probabilities (P), the operator set (O), and the evaluation metrics (E) of which Q = (A = {*p*}, O, I = *f*(E)). Using background models, syntactic tracking, content-based image retrieval (CBIR), similarity metrics, and moment theory, these concepts support anticipatory autonomy over entities in the environment for sensor and information management through channel analysis.

Channels (*sensors*): In natural and artificial vision, a notion of virtual representations is introduced along with other multimedia analysis for course of action and situational understanding. Methods are developed for

1. the use of natural language processing for explanation, comment, question, and integration; and
2. ambient intelligence over embedded, context aware, personalized, adaptive, and anticipatory reasoning.

Adding a new channel captures attention (attentive), supports information understanding (explicative), and improves information retention (mnemonic). Use of representational grammar for grouping, detailing, sequencing, comparing, and directing advances traditional image processing methods over social, cognitive, and sensor networks.

SUMMARY

The book combines contemporary techniques to move from basic image processing techniques of image cues to incorporate context and channels of multimedia content for ambient intelligence. Numerous examples are provided to guide and structure future CV systems based on information fusion of image and nonimaging techniques for a wide domain of users.

REFERENCES

[1] Cantoni, V., Levialdi, S., and Roberto, V. *Artificial Vision: Image Description, Recognition and Communication*. Amsterdam, Netherlands: Academic Press, 1997.

[2] Gonzalez, R. C., Woods, R. E., and Eddins, S. L. *Digital Image Processing* (2nd ed.). Upper Saddle River, NJ: Prentice Hall, 2009.

[3] Young, S. S., Driggers, R. G., and Jacobs, E. L. *Signal Processing and Performance Analysis for Image Systems*. Norwood, MA: Artech House, 2008.

[4] Blasch, E. P., Bosse, E., and Lambert, D. A. *High-Level Information Fusion Management and Systems Design*. Norwood, MA: Artech House, 2012.