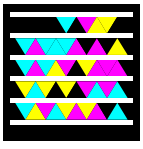
almost been synonymous with machine vision, which suggests processing of data from cameras operating in the visible range, the trend is to use multiple modalities of inputs that include sensors such as radar and microphone arrays. The trend is to tackle the problem using a layered model starting with the sensor outputs and working through other layers such as front-end processing, object localization, object recognition, context recognition, and spatiotemporal perception. Key technical challenges include: 1) developing robust real-world algorithms for midlevel tasks, 2) generating "complete" ontologies of scenes/scenarios of interest, and 3) discerning and describing events beyond the trained set. With respect to autonomous systems, there are further challenges such as computation, scalability over robot platforms, interfacing with machine intelligence, and human robot interface components. The field offers a rich variety of signal processing problems in areas such as multichannel processing involving multimodal sensor outputs, cue and behavior inference, and symbolic representations.

Slides



CONCLUSIONS

This article has outlined major emerging signal processing applications and industry technology in areas of consumer electronics and military systems. As a matter of fact, signal processing is also finding a very wide use in many other areas such as e-health and e-learning. It is worth mentioning that one trend is that various different applications are converging to one single device, and smart phones are an excellent representative of such a converged device, including autostereoscopic 3-D, gesture recognition, immersive gaming and video experience, and augmented reality. Finally, we wish to mention that the latest information and development on related topics can be found in the IEEE SPS Series Conferences on Emerging Signal Processing Applications (ESPA) whose first edition will be held 12–14 January 2012 in Las Vegas together with one of the annual largest events: the Consumer Electronics Show.

AUTHORS

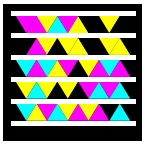
Fa-Long Luo (f.luo@ieee.org) is the chief scientist of two high-tech international companies headquartered in San Jose, California.

Ward Williams (ward.williams@elementcxi.com) is vice president of Element CXI, San Jose, California.

Raghuveer M. Rao (raghuveer.rao@us.army.mil) is the chief of the Image Processing Branch, Army Research Laboratory.

Rajesh Narasimha (rajeshn@ti.com) is a member of technical staff, Texas Instruments Inc.

Video



Marie-José Montpetit (mariejo@mit.edu) is a research scientist with the Massachusetts Institute of Technology Research Laboratory of Electronics.

Gaurav Sharma, Lina Karam, and Patrick Wolfe

Select Trends in Image, Video, and Multidimensional Signal Processing

The Image, Video, and Multidimensional Signal Processing (IVMSP) title combines within it a broad scope of technical areas dealing with the representation, processing, analysis, classification, compression, and transmission of still images, video, and general multidimensional data. Within the IEEE Signal Processing Society, the IVMSP Technical Committee (IVMSP-TC) has the charter of promoting and guiding the advancement of these technical areas. The authors, as members of the IVMSP TC, presented a summary of selected trends within the IVMSP arena at ICASSP 2011; this column is an outgrowth of the eponymous presentation. We begin with vignettes of the global high-level trends affecting the IVMSP area and then focus our attention on two specific topics. No attempt is made at exhaustive coverage and at including detailed bibliographic references, both being infeasible within the space and time devoted to the column because of the extremely broad scope of IVMSP.

We argue that, faced with ever-increasing data rates across a range of scientific, engineering, and consumer applications, the signal processing community is forced to consider new paradigms for high-dimensional data analysis and understanding. As illustrative examples, we discuss current trends in 3-D video processing, the analysis of signals that can be represented as graphs, and the mathematics of non-Euclidean data analysis that will be required to drive future developments in signal processing for "big data."

Digital Object Identifier 10.1109/MSP.2011.943130
Date of publication: 8 December 2011

GLOBAL TRENDS

The IVMSP area has seen explosive growth in recent years with contributions coming from developments in new applications/devices and in novel theory/algorithms. On the applications side, the predominant modes of interaction with images have become mobile and distributed—phone cameras are by far

1053-5888/12/$31.00©2012IEEE

the leading class of image capture and display devices today, and social networks are the main mode for sharing these images. Among network traffic sources, streaming video now represents the largest single source of data, outstripping the peer-to-peer networks that dominated for the last several years. Stereo and 3-D imaging are gaining acceptance and are becoming commonplace in movie theaters as well as in consumer devices. The traditional approach wherein (almost) every pixel within an image is captured by a sensor is being replaced by a computational imaging framework where pixels are computed from multiple captures of either images or other observations.

The new applications and devices in the IVMSP area are drawing upon advances in theory and algorithms, where we also see some overarching trends. Multidimensional, multimodal large data sets are becoming the norm rather than the exception in problems in science and engineering. The most promising problems and algorithmic approaches tend to increasingly cross boundaries of what were traditionally separate disciplines. The IVMSP area is increasingly defined by overlaps of conventional signal processing with other disciplines. The traditional disciplines of mathematics, physics, and statistics are well represented in this overlap but there are also increasing overlaps with computer vision, psychology, communications, machine learning, and artificial intelligence.

GRAPH REPRESENTATIONS FOR MULTIDIMENSIONAL DATA

A clear metatrend across signal processing and related disciplines is the way in which increasing data rates are driving researchers to extract low-dimensional structure from high-dimensional data. Of course, this is the essence of parametric statistical modeling, but now more than ever before, convergence of formerly separate disciplines is also driving us toward a common framework for the analysis of large, multimodal data sets. An important trend in the area of signal processing for "big data" is the emergence of graph representations in multidimensional signal processing. Graphs provide a common framework for treating diverse types of sparse, high-dimensional data that range all the way from very structured (e.g., social networks) to very unstructured (e.g., free text). Modern-day examples include everything from relational databases, to social media data, to pairwise cooccurrences and other similarity measures on a set of objects such as documents. Graphs are often the actual data objects of interest themselves (as in the case of social networking data), or alternatively they may be induced in a number of ways from traditional vector-valued Euclidean data (as in the case of the range space of a set of image pixel values).

The metatrend identified in the preceding paragraph (itself the subject of much attention at recent ICASSP and other IEEE Signal Processing Society conferences) provides strong evidence that the requirement to bring together structured and unstructured data, at increasing computational scales, is set to become a central component of signal processing in the future. Indeed, for the reasons cited above, the various communities of interest in these areas have begun to coalesce around network representations as a common framework for this type of signal processing. Network structure can provide a type of prior contextual information that informs traditional signal processing of content (sounds, images, etc.), or alternatively such structure can itself be inferred on the basis of this same content. This opens up a new vista for multidimensional signal processing theory, methods, and applications.

We close with a brief discussion of challenges in this area. Perhaps the central challenge is to determine what a general theory of signal processing for graphs would, could, or should look like. Many challenges and opportunities lie in uncovering the basic phenomenology and driving principles governing network formation and dynamics, developing fundamental theories of detection and estimation that define graph-based signal processing, and fully integrating graph-based structural information into the processing of traditional signal types such as images, speech, and text. Each of these opportunities is set to drive important future developments in multidimensional signal processing for "big data."

TRENDS IN 3-D VIDEO PROCESSING

Recently, the amount of data that is acquired and processed by multimedia applications has been increasing and expanding in dimensionality at a rapid pace. For example, in the consumer electronics area, TVs were introduced with black-and-white motion video in the 1930s. Since then, the dimensionality of the data expanded in the spectral dimension, from black-and-white to color, in the spatial dimension, from standard definition (SD) to HD, to full HD, to 3-D, to multiview, and in the temporal dimension with higher frame rates to accommodate high action video and a more natural 3-D/multiview video playback. On the consumer's end, in addition to the large TV sets at home, there is a growing demand for smaller, portable/mobile multimedia gadgets capable of capturing and processing this large amount of high-dimensional data. Due to the constrained nature of these devices in terms of power, computational performance, and bandwidth, this would not be possible without the design and development of efficient multidimensional signal processing technologies. For example, these signal processing technologies would enable 3-D multimedia applications on low-power mobile devices through a selective acquisition and processing of only a portion of the data. On the 3-D content producers' and broadcasters' end, even when resources (e.g., bandwidth and power) are available, ensuring a good quality of experience (QoE) requires producing good-quality 3-D content, which in turn requires the acquisition and/or production of a large amount of high-dimensional data (e.g., multiview, full HD, 3-D color video at 240 frames per second). This process can be very costly, especially for 3-D video as the cameras need to be set up and calibrated to minimize mismatches between the right and left views. Any remaining mismatches would need to be corrected through time-consuming manual editing as improper 3-D content can be visually very annoying and can cause severe visual discomfort and headaches. So, there is

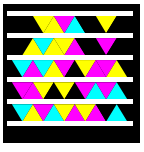
also a need here for efficient multidimensional signal processing methods for the acquisition and production of good quality content in a timely manner and at a reasonable cost. These methods would allow the automatic generation of the desired content from few captured data samples, reducing thus significantly the acquisition and production time and cost.

In addition to consumer electronics applications, there are various 3-D video applications that would benefit from signal processing methods for enabling real-time deployment; these include medical, automated manufacturing; remote sensing; localization/tracking; surveillance; automated navigation; and situational awareness applications, to name a few. To this end, there is a need for improved signal processing methods dealing with two-dimensional (2-D) to 3-D conversion; virtual view synthesis; disparity estimation and depth map generation; 3-D video compression and transmission; 3-D video representation and reformatting; 3-D video content quality assessment; and 3-D video pre- and postprocessing (e.g., distortion correction, color matching, enhancement); and perceptual-based 3-D video processing,

Disparity estimation is at the core of most of the enabling solutions. The disparity is estimated using a stereo image pair or from multiple views of the same scene. The estimated disparity can be used for computing a depth map for the scene, for compression by exploiting inter-view redundancy or for reconstructing virtual views as discussed above. Global disparity estimation techniques estimate the disparity field by minimizing a global cost function over the entire image, while local disparity estimation techniques consider only on a local region surrounding the current pixel being processed. In computing disparity/depth maps, there is a need for techniques that would ensure smoothness, temporal consistency (to avoid flicker), and the reliable handling of occlusions.

In 2-D to 3-D conversion, the objective is to reconstruct the depth information of the 3-D scene from 2-D images or 2-D video. These techniques exploit the relative motion between object and camera. When only a single static 2-D image is available, monocular cues and/or machine learning techniques are used in estimating the depth information. If two or more views are used, disparity estimation and camera calibration techniques can be employed to recover the depth information.

Slides



In virtual view synthesis, the objective is to reconstruct missing views using a limited number of available views. To this end, improved methods for inter-view disparity estimation and accurate dense depth map computation are needed. Virtual view synthesis is key for 3-D post-production and processing to ensure a pleasant viewing experience (e.g., depth adaptation depending on application, adaptation to screen size and resolution of 3-D display, content creation for autostereoscopic multiview displays, and free viewpoint video functionality). Common virtual view synthesis techniques include depth image-based rendering (DIBR), layered depth images (LDI), and image-based warping. In DIBR, the depth maps are estimated from existing views, and these depth maps are then used to synthesize virtual views. Issues considered include handling depth discontinuities to reduce artifacts along object borders, checking the consistency of the disparity maps of different views, handling occlusions, and reducing blur due to errors in camera parameters. In LDI, several color and depth values are stored for each pixel to compensate for possible occlusions. Image-based warping techniques do not require reprojection to the 3-D space, and they operate directly in the image space by defining feature correspondences between source and target images, which are used to generate virtual views through interpolation and morphing.

While existing 2-D video coding standards can be used to encode 3-D video content after reformatting it into a compatible format, there is a need to develop efficient 3-D video compression and transmission schemes that can exploit the 3-D video characteristics. One effort in this direction is the International Telecommunication Union-Telecommunication Standardization Sector (ITU-T) Multiview Video Coding (MVC) extension of the H.264/Advanced Video Coding (AVC) standard, which exploits the inter-view redundancies in addition to the temporal and spatial redundancies for coding the 3-D video. More work is needed to ensure that the perceptual characteristics of the 3-D video are taken into account. One emerging area of interest is asymmetric 3-D video coding and processing, which is based on the observation that the human visual system (HVS) can compensate for some distortions if these are present in only one of the stereo views.

Assessing the 3-D video QoE is crucial to ensure the production of good 3-D content. Despite its importance, this area is still in its infancy. Before developing objective QoE metrics, there is first a need to develop suitable 3-D subjective quality assessment methodologies that would shed light on the perceptual characteristics of 3-D video and how these affect our viewing experience. Factors that might be of interest include depth perception, naturalness, immersiveness, and visual comfort in addition to the overall 3-D video quality. There is also a shortage of good 3-D video databases. These are needed for conducting the subjective assessment and for evaluating the performance of the objective QoE metrics. Automatic 3-D video pre- and postprocessing techniques need to be developed for "fixing" any quality issues in the 3-D video content, such as 3-D restoration and enhancement techniques to reduce mismatches between stereo views and reduce annoying distortions. Reconditioning techniques are also needed to optimize the 3-D content to the available viewing environment. Perceptual-based processing techniques that exploit the HVS characteristics are certainly desirable and are expected to improve efficiency and quality.

AUTHORS

Gaurav Sharma (gsharma@ece.rochester.edu) is an associate professor at the University of Rochester.

Lina Karam (karam@asu.edu) is a professor at Arizona State University.

Patrick Wolfe (patrick@seas.harvard.edu) is an associate professor at Harvard University.

Video