Immersive communication evolving through "graph signal processing"

What are the new techniques for improving compression, interpolation, and denoising in images and video?

"Graph signal processing" technology is a relatively new technique used to efficiently analyze and process signals in social networks and sensor networks. We spoke with University of Southern California Professor Antonio Ortega, a pioneer in research on applying this technology to images and video, and NII Associate Professor Gene Cheung, who has been engaged in joint research with Professor Ortega for many years, to learn what kind of innovations graph signal processing technology will bring to "immersive communication" and to the world of images and video.

Antonio Ortega
Visiting Professor, National Institute of Informatics (NII)
Professor, University of Southern California

Gene Cheung
Associate Professor, Digital Content and Media Sciences Research Division, National Institute of Informatics (NII)
Associate Professor, Department of Informatics, School of Multidisciplinary Sciences, the Graduate University for Advanced Studies

Clarifying signal structure through "points" and "lines"

–– Can you provide a simple explanation of "graph signal processing"?

Ortega: A graph is a way to represent the structure of information using nodes and edges. As an example that is easy to understand, let's think of users of a social networking site such as Facebook as "nodes" and the connections between users as "edges".

Individual nodes possess a variety of attributes, such as gender, annual income, favorite music, field of interest, etc. These pieces of information are considered "signals" on each node. Nodes are connected primarily by "friendships", but nodes with common "annual income level", "favorite music", "fields of interest", etc., can also be linked. Edges can be weighted according to the degree of commonality or similarity of attributes. When a company wants to analyze a social networking site for the purpose of marketing, it can efficiently and effectively grasp market trends and customer attributes by appropriately designing the structure of nodes, edges, and edge weights according to a well-defined objective.

This is the same for a network of weather sensors. By arranging weather sensors in a mesh in a certain region, weather changes in that region can be understood by networking adjacent sensors with each other, collecting data such as temperature, humidity, and rainfall amount, and weighting the edges according to measured values and distance between sensors. Efficiently analyzing data by creating a graph made up of nodes and edges in this manner and examining the relationships between generated signals is the basis of graph signal processing.

Cheung: Applying the technique of graph signal processing to image processing is the topic of our joint research. In conventional discrete signal processing, audio is divided into equal time intervals, images are sampled at equal spatial intervals, and video is additionally divided into frames of equal time intervals. This processing is based on, so to speak, a uniformly organized "regular" data structure. In graph signal processing, on the other hand, signals having an "irregular" data structure can be used as the object. By freely creating appropriate graphs as necessary and interpreting the relationships between signal samples, new techniques not imaginable before can be developed.

Fig. 1 Example of depth image transform

Among pixels arranged in a 2D grid, those with high similarity to each other (in this case, those that are roughly the same distance from the camera) are given an edge weight of 1, and those that are very different (those far from the camera) are given an edge weight of 0. In so doing, some of the lines of the graph are disconnected, and the image can be split into two parts (images of the foreground and the background).

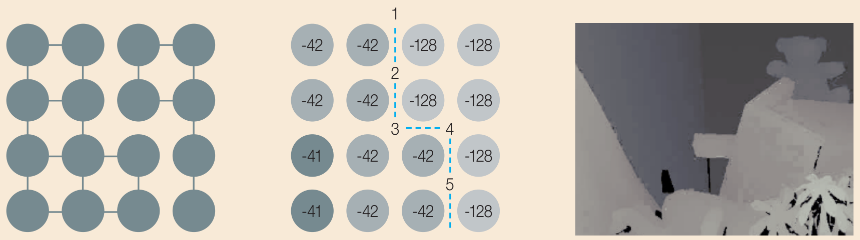
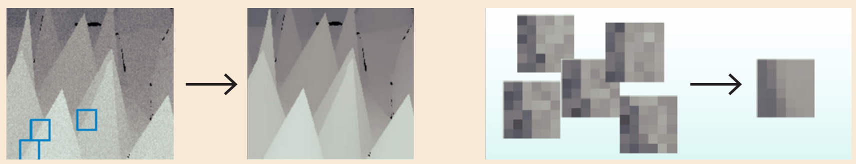


Fig. 2 Example of image denoising via graph signal processing

Noise is removed by averaging similar neighboring pixels. By designing a graph and giving it appropriate edge weights, noise can be removed to produce a clear image without blurring.



Improving indispensable immersive communication components: compression, interpolation, and denoising,

–– How is graph signal processing useful for research in immersive communication?

Cheung: Research in immersive communication aims to produce an interactive visual experience as close to reality as possible. For example, in video conferencing between remote locations, if the person facing the display screen looks over the shoulder of the person being captured, he can see the background behind that person. If we expand upon this capability, we expect that, for example, sports video viewers will be able to freely "change seats" and watch from an angle they like.

To realize this, video must be shot by multiple cameras, and the required viewpoint images must be synthesized instantaneously while predicting changes in the viewer's line of sight and point of view. Since the amount of data will be enormous, data compression is essential to efficiently transmit and process it. Interpolation is also crucial to compensate for missing image pixels due to low resolution or disocclusion. Denoising noise-corrupted images is necessary as well. Graph signal processing is anticipated to be effective in all three areas of compression, interpolation, and denoising.

For example, to separate an object close to the camera and the background of an image, a graph is created by assigning large weights to the edges between nodes (pixels) having high similarity and assigning small weights to the edges between nodes with the greatest differences. When a graph Fourier transform is computed, the foreground and background can be efficiently separated according to the plus/minus sign of the lowest-frequency alternating-current (AC) component.

Applying this method to image compression, if we define a Fourier transform using a graph representing the structure of the image described above for one block, a compact representation of only low-frequency components can be obtained. Due to this sparse representation, the compression ratio can be increased. When we tried it with depth images used in video synthesis to produce realistic video, we demonstrated a high compression improvement of 30 to 40% over state-of-the-art image codecs.

Even if data from part of the image have been dropped, the dropped data can be interpolated by creating a low-frequency signal consistent with a derived graph structure, resulting in a smooth image. Denoising is performed by inserting a weighted average of similar portions in an image, where weights are pre-assigned according to derived image structure, resulting in removal of unwanted noise without blurring object boundaries.

The future of "graph signal processing" in images and video

–– What are your hopes for the future?

Ortega: The application of graph signal processing to images and video is at the research stage, but companies are paying close attention, particularly with regard to compression. Google and Microsoft are both involved in research on this topic. Enterprises will probably begin using their proprietary technology in 3 to 4 years, and may then face the prospect of standardization. Other applications are also emerging one after the other right now, and I anticipate that new applications will be discovered in the future as well.

Cheung: It is interesting that graph signal processing will bring about solutions to age-old problems surrounding image and video processing from a whole new viewpoint. By pushing ahead with research in close cooperation with Professor Ortega, I believe that we can take advantage of graph signal processing as an immensely powerful tool that can be applied to all of the issues involved in immersive communication, such as compression, interpolation, etc.

(Interview/Report by Masahiro Doi)