Video Image Reconstruction and Enhancement

A Tera-scale Computing Application

Abstract: We describe advanced video technologies for image reconstruction and enhancement. Sophisticated image reconstruction methodologies can enhance and improve the image quality of lower-resolution video.

Introduction

Harnessing the Power of Tera-scale Systems

The performance of future tera-scale computers, based on 10s to 100s of integrated processor cores, will empower technology with more human-like capabilities. In the future, our devices will be able to comprehend the meaning of bits and bytes of data streaming past, and use this knowledge to act on our behalf. To do so, computers must have the ability to think in terms of models—digital concepts of the people, places, and things found within streams of bits and bytes.

By forming ever better models of what a person looks like in a photo, how virtual objects and characters should move, which sounds define speech, or even what data makes up a financial portfolio, future computers will be able to do much more than those of today. They will be able to recognize these models within rich data, such as photos or streaming video. They will be able to mine a desktop PC, corporate server, or the Internet to find similar models. They will even be able to generate new models–for instance, synthesizing new virtual characters or testing out financial strategies for a user.

Engineers and researchers in Intel's advanced research labs have been investigating ideas associated with tera-scale (10^{12}) computing power (trillions of operations per second) to help solve highly complex problems, perform critical mathematical analysis, and run computationally intensive workloads more efficiently and in real time. This research is based on the need for intensive computation, visualization, and manipulation and management of the massive amounts of data associated with these complex scientific and commercial applications. **We anticipate that trillions of calculations per second (teraflops) will be required to achieve the level of performance needed to run these applications.**

Intel researchers believe that a small set mathematical algorithms underpins a broad class of numerical, high-performance workloads—workloads that are typical of future applications that will utilize tera-scale computing performance. They are exploring the development of these algorithms, focusing on workloads in the areas of financial analytics, medical imaging, visualization, energy, and information search and manipulation. We anticipate that Intel Architecture will be broadly used to run tera-scale applications in these domains over the next ten years.

Video Image Reconstruction

In this white paper we describe an advanced set of algorithms for video image reconstruction and enhancement. Our methodology can be used to improve still images and to enhance the quality of low-resolution video without restricting the motion of objects within a video sequence. We describe a multiple-frame capability that leverages and builds on the information in successive frames to dramatically improve video resolution.

There are four primary techniques for improving image resolution: video frame rate conversion, motion stabilization, video artifacts removal, and super-resolution. This paper describes the details of these techniques, which can be used individually or in combination. We also discuss example applications or usage models of the techniques, which can be embedded in, or used to enhance the output of, a variety of consumer electronics products, such as to cell phones, camcorders, set-top boxes, media players, and media PCs. For instance, the techniques described in this paper can dramatically improve the quality of home videos taken using older, low-resolution camcorders, often achieving a level of clarity that surpasses the capability of state-of-the-art digital camcorders.

Video image reconstruction is just one example of emerging intelligent applications that will require advanced server computing capabilities and the use of parallelism to allow applications to be processed in real time. Real-time processing represents a massive shift in consumers' ability to search in video content, and it requires an equally massive shift in hardware and software. For instance, today the process of searching through a home video to find, say, all the scenes that include a dog, is 10 to 1,000 times slower than what most users would find acceptable, according to our calculations. Offline processing can be used in many situations, but time-critical detection of content requires real-time processing, and that, in turn, requires tera-scale systems and parallelization of software threads.

Motion Stabilization	**Super-resolution**
Frame Rate interpolation & Conversion	**Camera and MPEG Artifacts Removal**

Figure 1: Four Advanced Media Functions, for Video Image Reconstruction

Intel researchers are exploring a variety of usage models that will depend on tera-scale processors and parallel code, including more intelligent and intuitive user interfaces that are more proactive in anticipating a user's needs and intent. Our research will enable us to understand how to optimize these applications for tera-scale architectures. The research also will provide guidance to the designers of future Intel architectures, and a preview of the benefits these new applications will bring to consumers over the next 5-10 years. Figure 2 shows the iterative process that Intel engineers and researchers are using to advance Intel's tera-scale computing research.

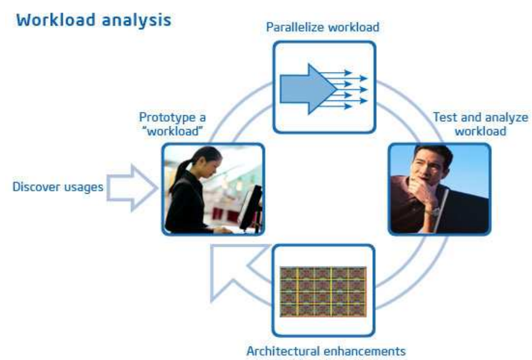

Figure 2: Iterative process used to advance tera-scale computing research

Advanced Digital Media Algorithms

Introduction

Image sequences in a video usually contain highly redundant information about the same scene. This information can be used to obtain a higher resolution image sequence from the lower resolution sequence. We apply a two-step process to achieve this.

The first step involves motion estimation. We calculate motion vectors to estimate the direction of moving objects. For instance, if a low-resolution video shows a boy riding a bicycle, we can compare the location of the bike in two successive or more frames, using motion estimation techniques to interpolate between the frames and calculate the probable location of the bike.

In the second step, we apply robust Bayesian statistical techniques to determine which motion estimates produced in the first step are the most probable or "best fit," and eliminate those motion vectors that are unlikely or statistically impossible (for instance, a frame showing two bikes, or a boy with two heads).

Figure 3 provides an overview of this two-step methodology for enhancing image resolution, which we describe in more detail below.

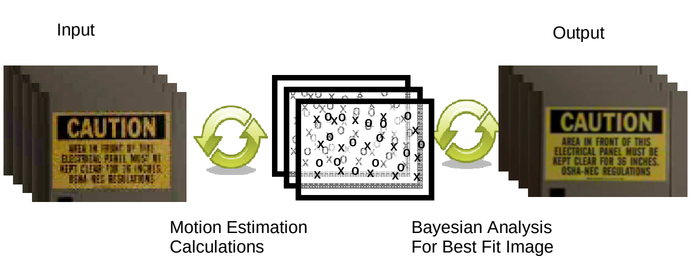

Figure 3: Block diagram shows the iterative combination of motion estimation and robust Bayesian Methods

Combining Motion Estimates with Robust Bayesian Inference

Figure 4 shows the details of how motion estimate techniques are used to estimate the movement between two pairs of images—the first step in our image enhancement methodology. Specifically, pixel-by-pixel precise alignments are calculated for the lower resolution video frames. Figure 4 shows the alignment of the pixels.

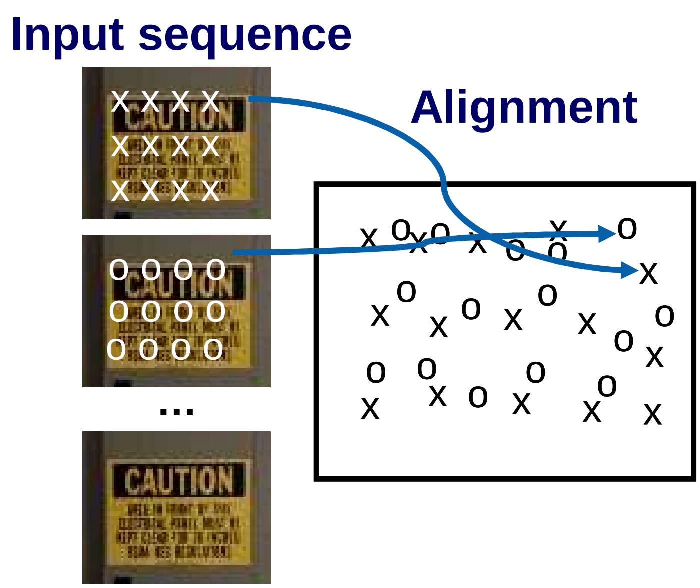

Figure 4: Motion estimation between image pairs

Once the pixels are aligned, a higher resolution image is calculated using robust statistical Bayesian inference of missing spatial information (step two). We explain the fundamentals of the Bayesian Inference model used. The Bayesian

Maximum a Posteriori (MAP) probability can be expressed as the following product:

$$p(g|F) = p(F|g) \cdot p(g) \quad \text{(Eq 1)}$$

where $p(g|F)$ is the Bayesian posterior probability, $p(F|g)$ is the likelihood function, and $p(g)$ is the prior probability, which acts as a regularizer. The goal is to find the best posterior probability with a likelihood function that best fits the observed images.

Described in terms of the images, the equation is:

$$p(\text{parameters}|\text{images}) = p(\text{images}|\text{parameters}) \cdot p(\text{parameters}) \quad \text{(Eq 2)}$$

where p(parameters|images) is the posteriori probability, p(images|parameters) is the likelihood, and p (parameters) is the a priori knowledge.

Maximizing the posterior probability gives the most likely set of parameters that best fit the observed motion estimation. The likelihood function is given by the high resolution to low resolution image with a geometric transformation. Robustness is essential in the likelihood function because some contributions to the noise come from incorrectly aligned regions, which are considered as outliers and are discarded.

Figure 5 shows the original 80x80 pixel image (top) and the image after super-resolution with 5X enlargement in both dimensions.

Figure 5: Original (top) and after (bottom) image using super-resolution

Image Reconstruction Techniques

The previous section described the overall two-step methodology we use to enhance image resolution. In this section we describe the four classes of image reconstruction techniques that are used within this two-step framework: **frame rate interpolation and conversion**; **motion stabilization**; **artifacts removal**; and **super-resolution**. For each technique, we suggest potential usage models.

Frame Rate Interpolation and Conversion

The frame rate interpolation and conversion technique is based on calculating image motion for each pixel. Motion estimation is used to calculate the probable motion of objects within a digital image sequence. Typically, the motion estimates are represented as vectors originating or terminating at pixels in a digital image sequence. For our implementation, a motion flow warped image is used to calculate frames at arbitrary time steps. Then robust Bayesian estimation statistics are applied to account for errors.

The frame rate interpolation and conversion technique has many useful applications. Following are three potential usage models:

- **Converting videos that use different international standards**. There are different and incompatible video formats in use around the world. For instance, video broadcasts and recorded material in the US will not play on equipment in Europe, and vice versa.

  Globally, there are three main video standards used: PAL, NTSC, and SECAM, each with different numbers of lines and different frequencies of fields/frames. Converting videos from one international standard to another is not a trivial task. Consider that the PAL standard is 625 lines at 50 frames/second, while NTSC specifies 525 lines at 60 frames/second. This means that in order to convert an NTSC video to the PAL standard, for every second, an additional 10 frames must be generated and the frame rate interpolation must be done at non-regular time increments to adhere to the required standard. Using our interpolation and conversion technique, this function can be carried out on today's computers that are based on Intel[*] Architecture. However, it requires offline processing; it cannot be done in real time.

- **Precision slow motion.** With the ability to create additional frames, an action can be slowed down without diluting image quality. Slow motion can accentuate a desired action view. It often is used in sports playback, for analysis of a player's physical actions or to highlight significant details that would be missed at normal speeds.

  Creating slow motion sequences requires more frames per second. By adding frames, when the video is played back at normal speed, scenes seem to go slowly, while the activity being shown appears to be "smooth."

- **Enhancing low-frame-rate videos.** This technique also can be used to fill in or interpolate between frames of low-frame-rate videos, to enhance the quality of the videos.

Motion Stabilization

The objective of the motion stabilization technique is to remove shaky motion from videos produced by hand-held devices such as cell phones, digital cameras and consumer camcorders. Motion stabilization uses robust statistics to calculate the "dominant motion" across an entire image. Then a plausible motion path is calculated, using statistical algorithms. Each image is warped according to the stabilized motion calculations. Following are three potential consumer applications of this technique:

- **Enhancing camera and cell phone videos**. Motion stabilization can be combined with the super-resolution technique (described later in this paper) to produce full-frame videos from the smaller size videos produced by cell phones and digital cameras.
- **Video Player Feature**. Motion stabilization could be embedded into consumer electronics video players as a playback feature.

[*] Intel is a trademark of Intel Corporation in the U.S. and other countries.

- **Media Center PC Feature.** This capability also could be included as a feature in a media center PC offering, and used for offline post-processing of home videos that have shaky motion.

Removal of Noise and Video Compression Artifacts

The third technique we are investigating is designed to remove noise and compression artifacts in videos, which are the result of low light conditions or “noise” artifacts produced by MPEG compressions. This artifact removal technique can be easily performed on today’s computers that are based on Intel architecture. Figure 6 shows a sample output of the removal technique.

Figure 6: Noise removal: left side of split screen is before, and right side is after artifacts removal

Following are two potential uses of these artifact removal techniques:

- **Correction of graininess due to low light level.** Camcorders and cameras with low signal-to-noise ratio lenses can produce images that appear grainy and smeared when filmed in low light conditions. Our techniques can improve these grainy and blurred images, transforming them into images that are comparable in quality to those produced by higher sensitivity cameras.

- **Video compression artifacts and MPEG noise removal.** There are visible artifacts associated with the most common video compression schemes, such as MPEG. Two primary artifacts are produced when pushing the limits of MPEG-2 compression: mosquito noise and blocking artifacts. Both types of artifacts can be corrected through our artifact removal techniques.

Video Super-resolution

Video-super-resolution capability, the final image reconstruction and enhancement technique, is one of the most important aspects of our research. Super-resolution is the process of reconstructing a high-resolution image from low-resolution input components. This technique generates high-resolution image and video content by combining *multiple* low-resolution images. This is done by calculating the precise alignment of multiple low-resolution images for each pixel. We use robust statistics to account for alignment errors (as explained previously, in “Frame Rate Interpolation and Conversion.”)

Our research in this area focuses on super-resolution of frames from dynamic video sequences, which may contain significant object occlusion or scene changes. Because the quality of super-resolved images relies heavily on the correctness of image alignment between consecutive frames, we use the motion estimation method described earlier to accurately estimate motion between the image pair. A scheme was designed to detect and discard incorrect matching that may degrade the output quality. This scheme requires multiple iterations to check and compare images. Computation convergence is one parameter, which can be adjusted for the desired performance, quality, and robustness of the output.

Figure 7 shows a super-resolution of text, which was unreadable prior to processing. The output image (bottom) is a four-times enlargement of the original, which was produced using a cell phone camera.

Figure 7: The top photo is a 160x190 pixel image produced by a cell phone camera. The middle image is a simple 4X resizing of the original photo, highlighting the poor quality of the input. The bottom image is a 4X enlargement of the original in which the text is readable. This image was created by using the super-resolution technique to fill in missing information in the original 160x190 pixel image.

While the super-resolution technique can produce impressive results, there is a tradeoff between the timeliness and quality of the output. If time is not a concern and the image can be processed overnight, this technique can produce high-quality results. However, if a result must be generated in real time, quality could suffer; without sufficient processing, the output could be corrupted—an unacceptable outcome in many cases. For instance, a doctor wanting to review enhanced images of a tumor may not have time to wait for overnight processing of the original image. However, a quality output is also mandatory in this situation. Insufficient processing time processing could produce a poor-quality or corrupted result, such as an image showing multiple tumors when in fact there was only one on the original image.

In general, the goal when applying the super-resolution technique is to "do no harm" to the original image—to ensure that above all, the output is no worse than the input because of insufficient processing. Typical computation times from our research show a range of nearly real-time to 10-100X less than real-time. This lag can be reduced with the higher processing power of multi-core computers.

Following are four potential applications of the video super-resolution technique:

- **Creating Still Images from Video.** The super-resolution technique enables users to select and capture a specific photo image in high resolution from a video. The technique can be implemented in a few seconds on today's PCs. It could be incorporated as a feature in video editing software or in home media PCs.
- **Improving the Quality of Mobile Phone Videos.** Cellular networks and video-sharing sites such as YouTube* could utilize super-resolution capabilities to improve the quality of video clips taken by cell phones, as illustrated in Figure 8.

- **Quarter CIF to CIF conversion.** Low-resolution or quarter-CIF video clips are common. These videos can be converted to CIF and even standard VGA resolution using our super-resolution technique. This capability can be added to video editing software or incorporated in home media PCs.

  Quarter CIF to CIF conversions can be performed in real time. Typical computation times from our work show a lag of 10-100X less than real time for converting from standard definition (SD) to high definition (HD). This lag can be reduced with the higher processing power of multi-core computers. For today's PCs, this feature can be run in the background or at night, when usage is low.

Figure 8: Mobile cellular phone camera video with 3X enlargement; on the left is the original image, and on the right is the image with the noise removed.

- **Converting legacy content to HDTV dimensions.** The aspect ratio for HDTV is 16:9 as opposed to 4:3 for traditional TV formats (NTSC, PAL, and SECAM). This represents a 33% increase in horizontal dimension. As high-definition televisions become common in consumers' homes, legacy content can be converted to high-definition quality, using our super-resolution technique. Typical computation times from our work show a lag of 10-1000X less than for real time processing. This lag can be reduced with dedicated hardware used by television and movie studios.

Conclusion

This white paper presented a technique for enhancing still images and computing high-resolution videos from low-resolution input without restricting the motion of objects within a video sequence. We described four advanced media capabilities for image reconstruction and enhancement—video frame rate conversion, motion stabilization, video artifacts removal, and super-resolution—and described the advanced algorithms underlying these techniques. We also discussed example usage models of these capabilities in consumer electronics, such as cell phones, set-top boxes, video recorders, and media PCs. For instance, the techniques described in this paper can be used to dramatically improve the quality of home videos taken using older, low-resolution camcorders, often achieving a level of clarity that surpasses the capability of state-of-the-art digital camcorders. Our investigations will continue in the areas of advanced 3D realism, immersive gaming and simplified natural gesture-based interfaces.

In this paper we have only hinted at the many industries that could benefit from video image reconstruction and enhancement techniques that tera-scale computing power will enable. We have shown that the consumer electronics industry can use these capabilities to enhance the quality of video images produced by a broad range of consumer devices. There are many other potential applications in a variety of industries. For instance:

- Video content providers, such as the television industry, could convert legacy video footage into high-definition video, or to up-scale legacy video footage.
- Companies in the field of security and surveillance could enhance recognition of faces, license plates and other identifying objects.
- The biomedical industry could apply the techniques to image analysis—for instance, to enhance ultrasound or MRI images.
- The semiconductor industry could apply the techniques during the manufacturing process, to enhance scanning electron microscope images.

Many other market opportunities are likely to emerge in the coming years, as our research advances and as tera-scale systems become widely adopted.

References

Haussecker, H. W.; Fleet, D. J., "Computing optical flow with physical models of brightness variation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, June; 23, 2001, (6): 661-673.

Nestares, O. Flee, D. J., "Error-in-variables likelihood functions for motion estimation," *ICIP* (3) 2003: 77-80.

Liang, B. and Zhang, Y., "Tera-scale Applications," *ITRS003, Intel Developers Conference*, Beijing, China, 2007, April 17-18.

Held, J., Bautista, J. and Koehl, S., "From A Few Cores to Many: A Tera-scale Computing Research Overview," White Paper, at http://www.intel.com/technology/magazine/research/tera-scale-1006.htm, 2006 December.

To learn more about tera-scale computing, visit
http://www.intel.com/go/terascale