IMINT
IMAGE INTELLIGENCE

Imint Vidhance for Video Stabilization and Compression

1 Introduction

1.1 Imint

Imint AB, based in Uppsala, Sweden, is a recognized innovation leader providing next generation, software-based solutions that optimize real-time video enhancement and analytics. Imint provides solutions for the current, acute need to fix image distortions caused by either ambient motion or environmental factors that disrupt image data from one frame to the next in video flow. Initially, the company focused on enhancing live aerial surveillance video streams for military applications by removing obstacles to human perception caused by (a) disturbances (such as uncontrollable movements, vibration, and turbulence), and (b) low contrast visibility (due to fog, mist, and rain). The resulting core software package is Vidhance. Building upon its initial success, Imint serves a broader range of customers in a variety of fields, with new products, and continues to focus on constantly improving and expanding the functionality of Vidhance.

1.2 Vidhance

Vidhance is unique, proprietary software, built partly on patent-pending methods. At its core, Vidhance (a) captures essential image data, (b) pre-processes/analyzes such data and then (c) processes this data to fix the image distortion. Vidhance is a modular and highly optimized suite of filters, algorithms and functions for real-time processing of live video streams. Vidhance brings capabilities to a standard PC platform that would otherwise need dedicated hardware implementations. Imint's proprietary algorithms are the result of advanced university research, which Imint has continually refined through its R&D to address the specific problems associated with image enhancement. Vidhance can briefly be described as two separate meta-analysis engines that extract information from any video stream, regardless of cameras or formats used. These analysis engines are derived from research methods, but clever shortcuts and optimizations are performed to meet real-time requirements and keep latency low. The first analysis engine is motion analysis, which provides features such as stabilization (planar, rotational, and z-axis), object tracking, detection of moving foreground objects on a moving background, video mosaicking, etc. The second analysis engine extracts information about contrast and saturation, and provides features such as scene-adaptive local contrast enhancement, colorization and color correction. A third, optional module in Vidhance brings in other system metadata, such as camera location and orientation, which can be used for geo-referencing pixels, automatic horizon alignment of video, etc. The software is originally written as Windows PC software, and makes use of both the CPU and GPU. It is however designed with modularity and portability in mind, and does not rely on any semiconductor specific hardware acceleration features.

Imint's feasibility studies conclude sufficient processor performance in current, widely used portable consumer device cores, such as ARM Cortex A15 and Qualcomm Krait, with dedicated graphics acceleration. The Vidhance algorithms have been trained over the years with a large data set of authentic full motion video from aerial, land, sea, and subsurface camera platforms. Vidhance brings radical visibility improvements to any vision challenged video, including High Definition (HD) video, with a delay of less than one frame. The Vidhance algorithms are constantly trained with new video data sets, and new features are being added. The unique Vidhance sweet spot is providing meaningful visual enhancements to image environments facing a combination of harsh filming conditions, real-time requirements, and resource constrained computing platforms.

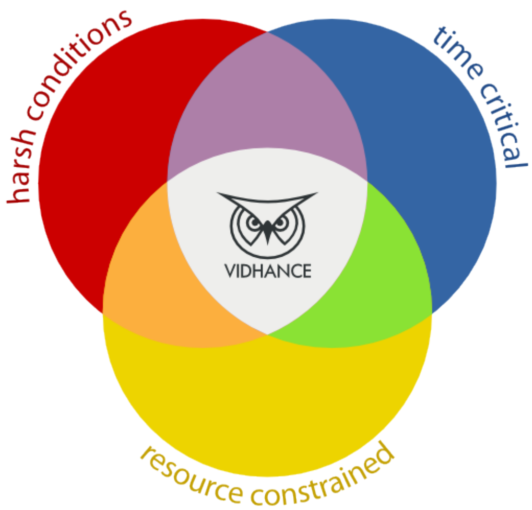

Figure 1: The unique Vidhance sweet spot.

1.3 Video Compression using Vidhance

Video compression usually works by computing and compressing deltas between frames, i.e. the difference between one frame and the next, rather than storing each individual frame as a compressed image. The more similar a frame is to the previous one, the less data is required in order to describe the delta. Stabilization works by compensating for unwanted camera movements, effectively making each frame in the video more similar to the one before it. Imint conducted a recent study investigating the potential gain in video compression by applying the Vidhance stabilization to video from handheld devices. This report outlines the method, findings, and other resultant associated benefits.

2 Theoretical Background

2.1 The difference between two images

The difference (diff) between two pixels can be computed by taking the absolute value of the difference between the color of those pixels. An image diff can be created by computing the pixel diff for every pixel in two different images. If a pixel has the exact same color in both images, that pixel will be black in the diff (black being RGB (0,0,0)), and the more different the color is in the two images, the brighter the pixel will be in the diff. The more similar the images are, the darker the diff will be.

Figure 1 shows a single frame from the unstabilized video. Figure 2 shows an image diff between that one frame and the next. Because the two frames are mostly very similar images, the diff is mostly black. The outlines of the tree trunk can be seen as green lines, as the camera has panned horizontally, causing these pixels to go from a dark grey in the first frame to a bright green in the second, or vice versa. Similarly, the outlines of the girl's hair can be seen as purple lines where pixels have gone from green to white (adding purple (255,0,255) to green (0,255,0) results in white (255,255,255) in the RGB color space).

Figure 2: A frame from the unstabilized video.

Figure 3: The diff between figure 1 and the following frame in the unstabilized video.

2.2 Stabilization

Most modern video encoders have some form of motion compensation built-in. H.264 employs local (but not global) motion compensation, and the same goes for its future successor, HEVC. This compensation can often reduce the delta between two frames to a few vectors describing how elements in the frame have moved in the next. The encoders do not, however, compensate for unwanted movement in the video. Professional video is typically very stable to begin with, but the same cannot be said for amateur video, captured using a camcorder or a cell phone. Imint's Vidhance stabilization keeps the image stable by compensating for unwanted movement moving the *frame* of the video, keeping the object being viewed in place. This makes for a more comfortable viewer experience and takes some of the workload off the video encoder. It does, however, require a little headroom for a cropping, which in turn requires the sensor of the camera to have a higher resolution than the resulting video.

Figure 3 is the result of superimposing the same two frames as in figure 2, but translating one of them by (23,-2) to compensate for unwanted camera movement. The diff is almost completely black aside from the girl's hand. This is not due to unwanted camera movement; the girl moved her hand. The area where the two frames overlap is now only 1224x716, which explains why some overhead is necessary in order for the stabilized video to be kept at a certain resolution.

Figure 4: *The diff between the same two frames, manually "stabilized" by translating one of the frames.*

2.3 x264

x264 is an H.264 encoder. It allows for 51 different quality settings where 0 is lossless and 51 is the most lossy, and where increasing the quality by 6 doubles the bitrate. Encoding a video at quality 23 (default) thus results in a file twice as large as encoding one at quality 29. The FFmpeg and x264 encoding guide[1] suggests qualities 18-28 to be a "sane range" and that anything below 18 should be considered visually lossless. There are also 8 different encoding presets, which give the encoder an idea of how long you're willing to wait for the video to be encoded. Forcing the encoder to work faster means less visual quality and/or larger files.

1. http://trac.ffmpeg.org/wiki/x264EncodingGuide

2.4 Constant bitrate

When encoding video at a constant bitrate, the quality of the video is limited by the amount of data allowed for each frame, so the delta must be reduced to that amount of data, even if this means losing a lot of information. By stabilizing the video before compressing it, the delta will contain less information because the stabilized frame is more similar to the preceding one. This translates into higher video quality with stabilization than without for a given bitrate.

2.5 Variable bitrate

When encoding video at a variable bitrate, the bitrate of the video is governed by the quality setting, so in order to maintain a given quality score, the delta may become very complex. A complex delta requires more data to store it without losing information. By stabilizing the video before compressing it, the delta will be smaller in size because the stabilized frame is more similar to the previous frame, which translates into a smaller video with stabilization than without for a given quality setting.

3 Procedure

Several video clips were shot using a handheld Sony HD camcorder at 1440x1080i. After deinterlacing, one stabilized and one non-stabilized version of each clip, upscaled to 1920x1080 and then cropped to 1280x720, were converted into H.264 using 4 of the 8 encoder presets available (ultrafast, faster, medium, and veryslow), and each of the 52 available quality settings.

4 Results and analysis

In all of our experiments, the stabilized video results in smaller files than does the unstabilized video for the lower quality settings, especially for very low quality settings. As a rule, the lower the quality setting, the greater the benefit of stabilization. In the sane range of qualities, the benefit is rarely less than 5%. Figures 4 and 5 show the file size reduction for 2 different video files at all 4 encoder presets, while figure 6 shows the file size reduction with the "veryslow" encoder preset for all videos in the experiment.

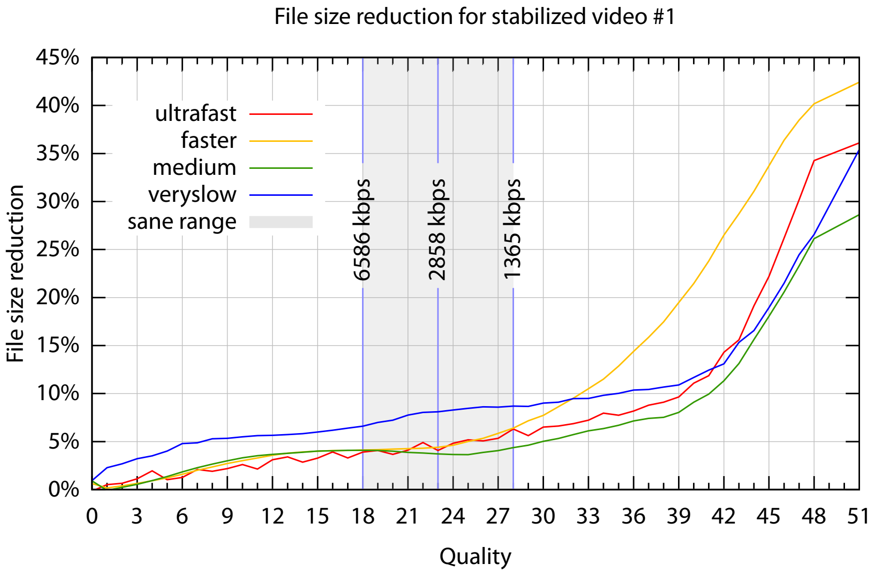

Figure 5: A plot demonstrating how file size reduction generally increases as video quality decreases for the video in figures 1, 2, and 3. The bitrates indicated are of the stabilized video for qualities 18, 23 and 28 at the "veryslow" preset.

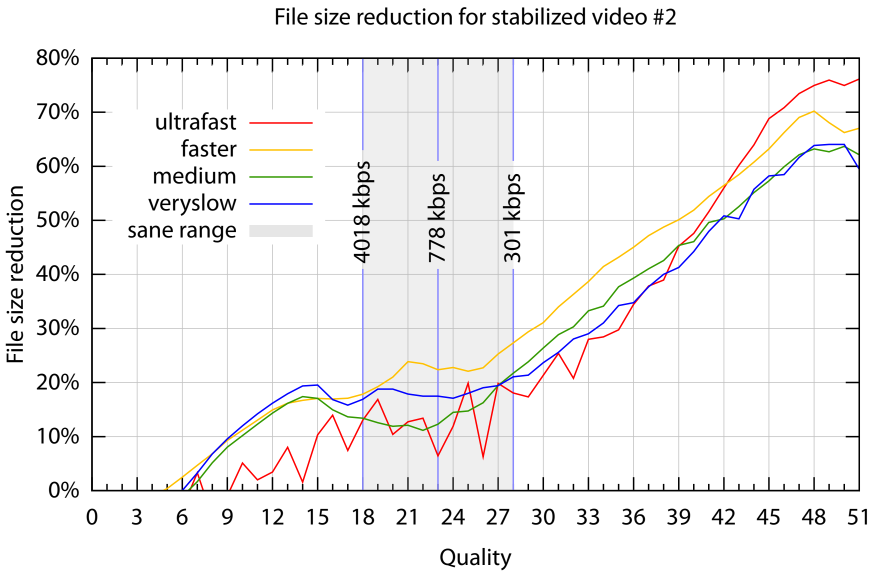

Figure 6: A similar plot demonstrating the same tendency in another video. In the sane range of qualities, the file size for this video is reduced by 16.9%-21.1%.

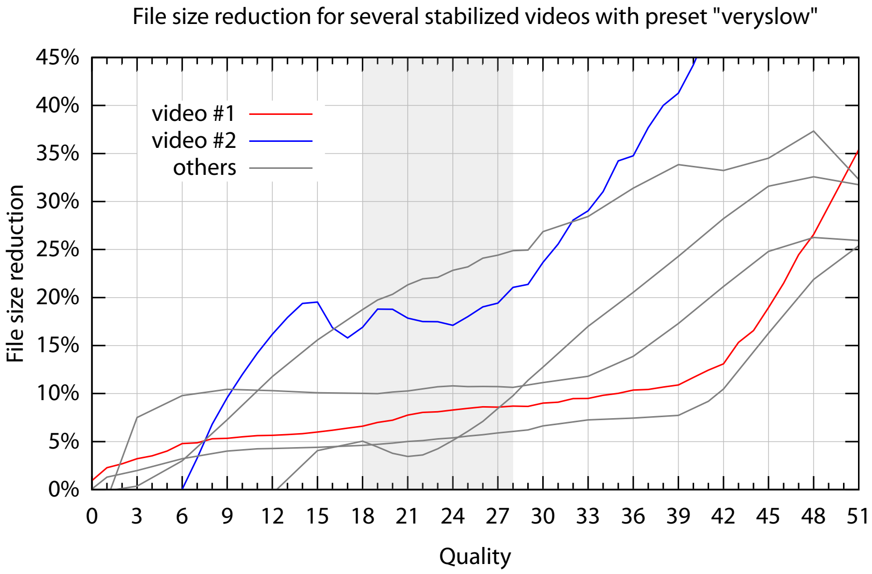

Figure 7: A plot demonstrating the same tendency in all of the videos used in the experiment, for only the "veryslow" encoder preset.

For very high quality video, the benefit of stabilized video is sometimes insignificant, and in rare cases, even negative – the stabilized video in these extreme cases is actually larger than the original video. In our experiment, this was only observed at extremely high quality settings, well outside the "sane range" and with no practical use. Part of the explanation could be that stabilization causes a misalignment of the blocky compression artifact pattern already in the video. These blocks can hardly be spotted without extreme saturation and brightness/contrast adjustments (as in figure 7), but at very high (above visually lossless) quality settings, they will still be "seen" by the encoder. Without stabilization, this blocky pattern stays in the same place throughout the video. If this misalignment can explain the contradictory results, uncompressed video should not suffer from the same problem, and should see a greater improvement across the full range.

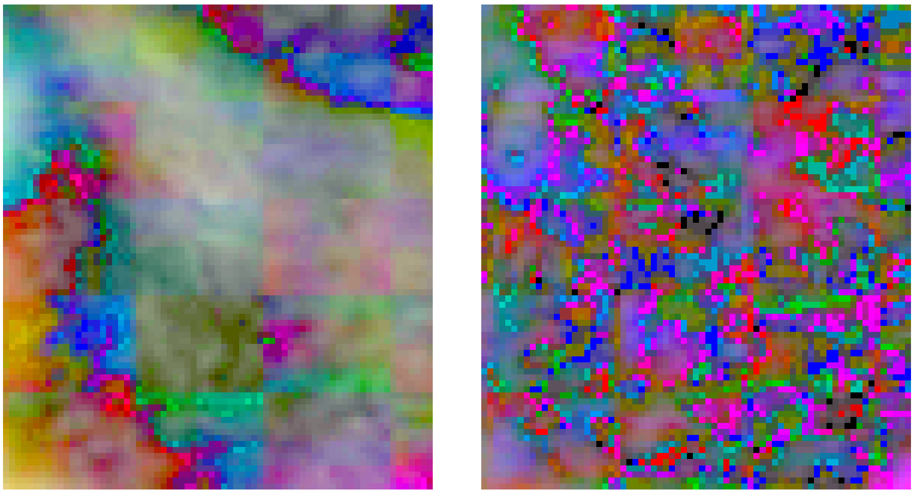

Figure 8: The images in figure 2 (left) and figure 3 (right), auto-leveled, saturated, cropped from (901, 444), and zoomed to 400%. In the unstabilized video (left), the compression artifact blocks are clearly visible after extreme saturation and brightness/contrast adjustments. The block size, $21.\overline{3}$x16 pixels, is the result of upscaling the 16x16 pixel blocks in the input video from 4:3 to 16:9. In the stabilized video (right), the blocks are still there, but barely visible, because the compression artifact patterns in the two overlaid frames are no longer aligned.

5 Conclusions

- By utilizing Imint's technology, stabilizing video before compressing it can dramatically reduce file size for medium and low quality video. This translates into less bandwidth usage, ultimately increasing network performance.
- Imint's initial study shows size reduction reliable above 5% in the range of standard encoding qualities, and typically in the range of 5-20%.
- Stabilization requires the camera sensor to have a resolution that is slightly higher than the resulting video.
- The tests have been made with the compressed video that could be obtained as output from the camera. It is a reasonable assumption that the size reduction could be even greater if implemented earlier in the video capture and encoding chain.
- The benefit of utilizing Imint's technology to stabilize uncompressed video is assumed to be significant enough to motivate further investigation. These future experiments should also be extended to include additional types of video content in order to find more conclusive and detailed results.