2011 IEEE/RSJ International Conference on
Intelligent Robots and Systems
September 25-30, 2011. San Francisco, CA, USA

Multilayer real-time video image stabilization

Jens Windau, Laurent Itti

Abstract— **In many camera-based robotics applications, stabilizing video images in real-time is often critical for successful performance. In particular vision-based navigation, localization and tracking tasks cannot be performed reliably when landmarks are blurry, poorly focused or disappear from the camera view due to strong vibrations. Thus a reliable video image stabilization system would be invaluable for these applications. This paper presents a real-time video image stabilization system (VISS) primarily developed for aerial robots. Its unique architecture combines four independent stabilization layers. Layer 1 detects vibrations via an inertial measurement unit (IMU) and performs external counter-movements with a motorized gimbal. Layer 2 damps vibrations by using mechanical devices. The internal optical image stabilization of the camera represents Layer 3, while Layer 4 filters remaining vibrations using software. VISS is low-cost and robust. It has been implemented on a "Photoship One" gimbal, using GUMBOT hardware for processing Sparkfun-IMU data (Layer 1). Lord Mount vibration isolators damp vibrations (Layer 2). Video images of Panasonic's Lumix DMC-TZ5 camera are optically stabilized with Panasonic's "Mega O.I.S." technique (Layer 3) and digitally stabilized with "Deshaker" software (Layer 4). VISS significantly improved the stability of shaky video images in a series of experiments.**

I. Introduction

Many robotics applications use camera systems to interact with their environment. However, it is critical that camera systems filter video images from vibrations in real-time for the robot's successful performance. Especially aerial vehicles (UAVs) use video images for surveillance, target tracking, navigation and localization tasks. Also human operators often track and observe environments via live video images from UAVs. Due to the significance of UAVs in military applications, reliably capturing clear video images is invaluable. Computers or humans often have to make a decision within seconds based on real-time camera footage. Recognizing targets or landmarks is likely to fail when vibrations cause defocusing and blurriness in images. The challenge of this work was to find an effective solution to stabilize video images in real-time.

The UAV sector is considered the "most dynamic growth sector of the world aerospace industry" with spending that will "more than triple over the next decade" [1]. Due to the significance of UAVs in the future, stabilization systems will remain an important part of this industry: Although most applications are military related, UAVs are also used in a small, but growing number of civil applications, e.g., firefighting or pipeline surveillance. Future UAVs will change in size, shape and configuration and require adaptable image stabilization systems. The motivation for this project was to develop a reliable stabilization system for a future UAV that had no predetermined specifications in type, size, shape, and weight yet. The stabilization system was required to eliminate vibrations onboard of the UAV for a vision-based surveillance application. For this purpose, a low-cost, light-weight and reliable stabilization system with an efficient adaptable architecture was needed. VISS aims to minimize horizontal, vertical and angular displacement of shaky images, reduce blurriness and avoid defocusing.

Vibrations in aerial robotics vary in amplitude and frequency [7][8][10][11]. Optimal stabilization is achieved when vibrations of any type in the entire amplitude and frequency spectrum can be compensated for. To achieve this, most stabilization systems use only one compensation technique (single-layer systems). Existing single-layer systems are hardware-based (e.g. gyro-stabilized gimbals by MicroPilot), software-based (e.g. Sarnoff, IMINT, Ovation Systems), or mechanic-based (e.g. Anti Vibration Camera Mount by DraganFly Innovations Inc or Vibration Isolators & Mounts by Lord Corporation). Various tests show excellent compensation results within certain amplitude and frequency ranges, but stabilizing vibrations over the entire amplitude and frequency spectrum usually fails with single layer systems [7] [12]. In fact, compensating all types of vibration in one system is still a challenging research problem [8] [9]. A major reason for failure is the time delay between detecting vibrations and performing compensation actions such as counter movements or digital image shift/rotate operations. One state-of-the-art example is the image stabilization system of "GE Intelligent Platforms". The real-time "SceneLock" algorithm detects horizontal, vertical, and angular displacement in images and compensates it via a pan-tilt-gimbal camera platform. However, "in some situations, the camera platform cannot take out all of the motion due to data transmission, processing latencies, or frequency response of the platform" [12]. GE suggests employing a different technique in these situations. Another challenging problem is the limited operational range for counter movements. While some approaches stabilize the camera body externally via rotating movements along pitch, roll, and yaw axis, many stabilization systems only focus on digital stabilization of captured camera images. In this case, the camera body is static and not externally stabilized. However, problems occur when e.g. critical landmarks do not stay in the camera image due to strong shakes. One approach is to construct a stable view of a scene by aligning

Jens Windau, Laurent Itti are with the iLab and Computer Science Department at the University of Southern California, Los Angeles CA 900089-2520, USA. (website: http://ilab.usc.edu phone: +1(213)740-3527 email: jenswindau@gmail.com, itti@pollux.usc.edu)

previous camera images and pasting them into an image mosaic. Image mosaic data can then be used to filter vibrations and let images look motionless [13]. However, images that were taken in unlucky moments in which e.g. dynamic landmarks are not visible might lead to problems when the final image mosaic is constructed. A different approach claims that only non-smooth rotational motions lower the quality of camera images. The algorithm estimates the camera body's rotation along pitch, roll, and yaw axis and corrects the images digitally via a smoothing algorithm. However, this approach is also limited to the field-of-view of the camera. The system is likely to lose track of a landmark when it disappears from the field-of-view due to a strong shake with large amplitude [14]. To overcome the limitations of single-layer systems, one approach to improve stabilization performance is to combine multiple, self-contained vibration compensation techniques in one system (multi-layer approach) [Fig. 1].

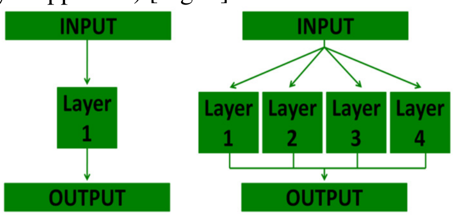

Fig. 1: A single-layer approach using one stabilization technique (left). VISS is a multilayer system with four layers working in parallel (right). Each layer uses a self-contained stabilization technique (right).

When adequate stabilization techniques simultaneously work together, they complement each other's operational range and overcome time delay problems. Parallel operating techniques could cover each other's area of poor performance in the amplitude and frequency spectrum: E.g. one layer compensates low frequency vibrations, a second layer damps medium frequency vibrations and a third layer isolates high frequency vibrations. The amplitude spectrum is split accordingly. A multilayer solution could thus achieve better compensation results when faced with all types of vibrations. Existing multilayer approaches are highly expensive. The US-manufacturer "Cloud Cap Technology" specializes in multilayer stabilization systems. Their state-of-the-art camera gimbal series for UAVs (TASE gimbals) uses both hardware-based gimbal counter movements and software-based image stabilization (two-layer system). In comparison to existing multilayer systems, VISS aims to lower the high costs ($14k-$95k, Pricelist Cloud Cap Tech, Feb. 2010) by providing a similar quality of stabilization performance. Furthermore, in case of a total power supply breakdown, the VISS architecture still provides a lower level of vibration stabilization by using a mechanical layer to damp vibrations. The VISS approach presented in this paper has a set of unique contributions: (1) a frequency-amplitude diagram to display and classify common vibration sources in aerial robotics (2) a unique multilayer architecture design derived from this diagram (3) a series of experiments showing performance results of the VISS architecture including the interaction among all four layers.

II. THE VISS APPROACH

(1) Displaying sources of vibrations in aerial robotics

Before designing the architecture of a stabilization system, vibrations in robotic systems should first be analyzed. Depending on a robot's environment, different types of vibrations can occur: Amplitude and frequency vary for rough atmospheric conditions (UAVs), rugged terrain surface (UGVs), wave motions (USVs), or underwater currents (AUVs) [3]. The rough vibration characteristics of a particular robot application should be listed in a frequency-amplitude diagram [Fig. 2]. In general, low frequency vibrations are more likely to affect equipment than high frequency vibrations [9]. This fact and simple performance tests of Layer 1 helped to determine the frequency axis scaling in the frequency-amplitude diagram. The amplitude axis scaling was chosen arbitrarily. Once all possible vibrations of aerial robots were displayed in the diagram, adequate compensation techniques had to be derived from it. In detail, vibration hot spot areas had to be covered with techniques that are expected to perform well in these frequency and amplitude range.

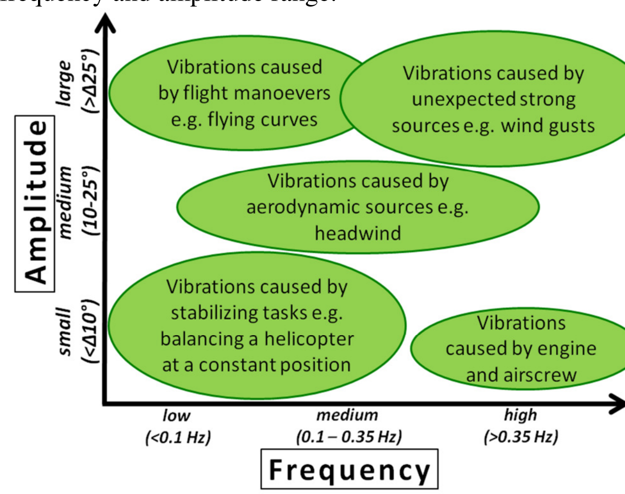

Fig. 2: Sources of vibrations in aerial robotics [7].

(2) VISS' multilayer architecture

A variety of compensation techniques exists to either partially damp or fully compensate vibrations: inertia-based [5], mechanic, optical, digital, GPS-based (Racelogic Ltd), and even temperature [6] and magnetic-based (Tyndall National Institute) sensor systems are available. After researching and testing various stabilization techniques separately from each other, four promising stabilization techniques have been carefully chosen. When combined in the VISS' architecture, they cover the entire frequency/amplitude spectrum of vibrations with either active compensation or passive damping [Fig. 3]. All layers differ in their physical approach on how to counteract vibrations: one inertia-based, one mechanical, one optical and one digital stabilization layer were implemented in the VISS architecture [Fig. 4]. Although each layer represents a self-contained stabilization technique, all four layers work in parallel and interact with each other to improve the overall stabilization result.

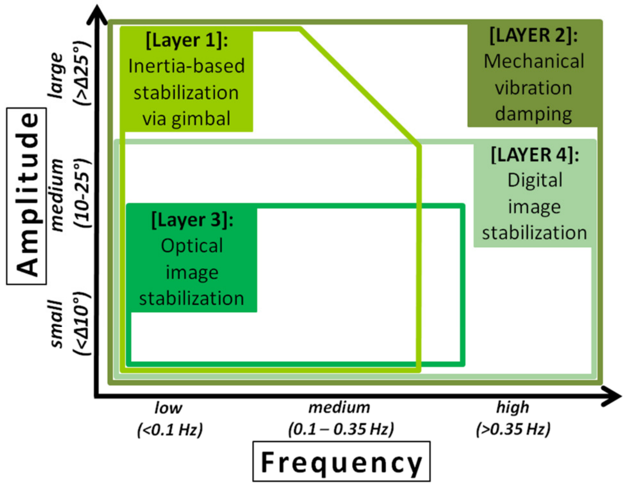

Fig. 3: The combination of four stabilization techniques in the VISS architecture covers the entire frequency-amplitude spectrum of vibrations.

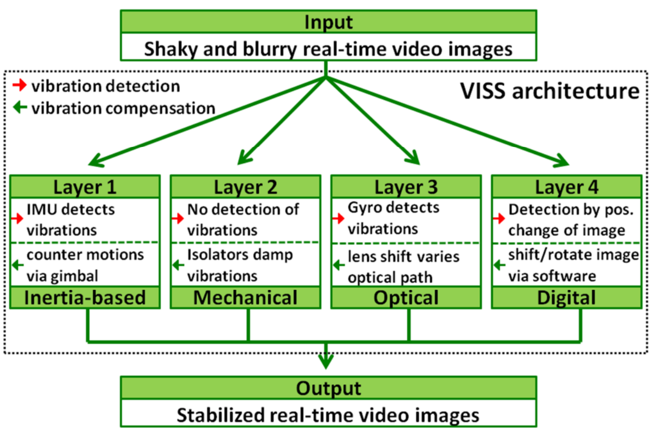

Fig. 4: The VISS architecture combines four layers. Each layer represents a self-contained stabilization technique which covers vibrations of Fig. 2.

(3) Interactions between layers

The idea of VISS is to run multiple stabilization techniques in parallel [Fig. 5]. In many experiments the combination of two and more concurrent running stabilization techniques is analyzed. By combining multiple stabilization techniques, disruptive interference of a layer could lead to a decline of the stabilization performance of another layer or even the entire system. However, almost all experiments showed a positive interaction between the layers [see Chapter IV].

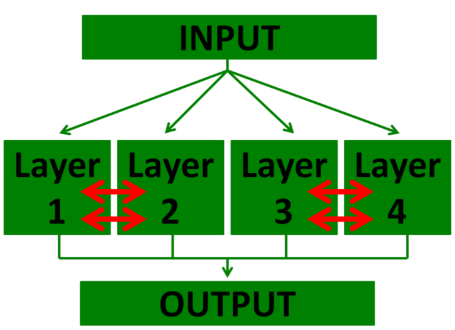

Fig. 5: Interactions between layers improve the overall stabilization result.

III. COMPENSATION PROCEDURE

A. Overview of VISS' four layer architecture

All layers and their operational ranges in frequency and amplitude are now explained in more detail [Fig. 6]:

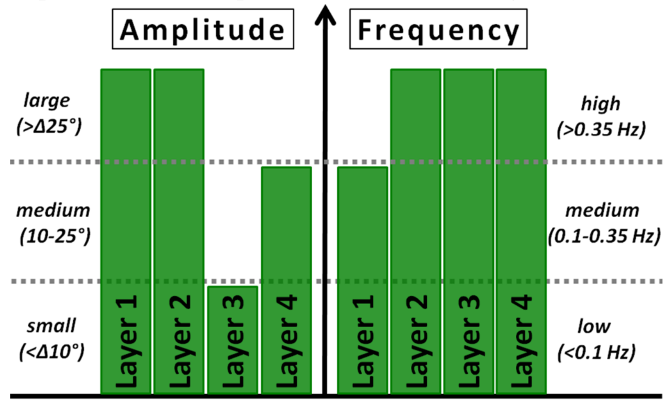

Fig. 6: Performance of each layer varies in amplitude and frequency ranges.

Layer 1: This layer uses an Inertial-Measurement-Unit (IMU) to measure vibrations and compensate the camera body externally via gimbal counter motions. Inertia-based gimbal stabilization is one of the only efficient techniques (besides mechanical gyros [4]) to compensate large amplitude vibrations. However, processing latencies and the gimbal's response time are unavoidable trade-offs using this technique. Thus, layer 1 only performs well for low/medium frequency vibrations.

Layer 2: Mechanical devices such as shock absorbers or Lord Mount isolators passively damp any kind of vibrations. Layer 2 is required to overcome the time delay problem of Layer 1, because it can instantly smooth high frequency vibrations. These vibrations can then further be stabilized by Layer 1. However, Layer 2 only passively damps vibrations of any frequency and amplitude, but does not actively and fully compensates them. Therefore, this layer needs to be combined with further stabilization techniques.

Layer 3: Optical Image Stabilization is a mechanism integrated in video cameras to stabilize the recorded video images by varying the optical path to the CCD-sensor. Video cameras are equipped with gyroscopes to measure horizontal and vertical movements. This stabilization technology moves either the lens or the CCD-sensor to perform compensation for small amplitude vibrations which cause blurriness and defocusing.

Layer 4: Digital video stabilization is another layer in the VISS architecture to eliminate camera shakiness. Most algorithms detect vibrations by position change of two consecutive video images. Vibration compensation is done by simply shifting and rotating images back to their original position. Software-based stabilization techniques compensate vibrations of small and medium amplitude. However, real-time results of software-based solutions are dependent on the quality of raw video images. Therefore, VISS uses Layer 3 to improve the image quality to guarantee a more reliable performance of Layer 4.

B. Technical set up

Layer 1 – IMU-based stabilization:

Layer 1 uses the PS1-3X Camera Gimbal (Photoship One) to perform counter movements along the pitch and roll axes [Fig. 8]. This gimbal is a light-weight product (1.3 lb), equipped with three high torque servos (PS1360-MG) and is able to carry cameras up to 3.5 lb. It is equipped with a 6-DOF Sparkfun IMU (3-axis accelerometer, 3-axis gyroscope) that is placed parallel next to the video camera. It measures vibrations that change the camera's pitch and roll angle. Counter motion performed with the gimbal are calculated on the GUMBOT hardware [2]. For PID-control purposes, the servos were manipulated to run as DC-motors.

Layer 1 consists of multiple filter sublayers [Fig. 7]. The goal of the filter sublayers is to convert noisy IMU data step by step into smooth motor speed commands to perform gimbal counter motions. In detail, accelerometer data is used to calculate pitch and roll angles. In order to guarantee reliable results, it first gets band-pass filtered to eliminate error measurements. Next, a Kalman Filter is used along with gyroscope data as a second input to come closer to the real pitch and roll angles. After this step, the data gets averaged to minimize noise. Finally, the filtered accelerometer data is converted into pitch and roll angle. With this data, PID motor commands are calculated to stabilize the gimbal for a desired pitch and roll angle.

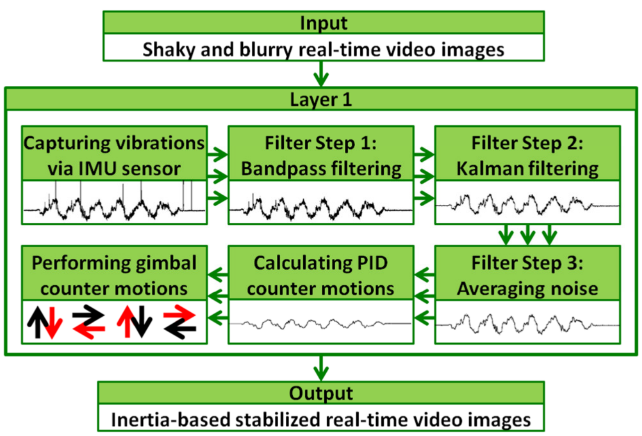

Fig. 7: A detailed look into Layer 1 with all its sublayers.

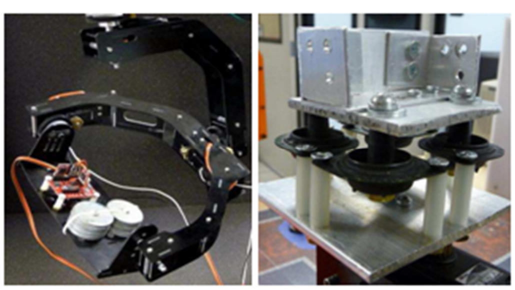

Fig. 8: PS1-3X Camera Gimbal with IMU (left), cluster of 4 vibration isolators to damp vibrations (right).

Layer 2 – Mechanical Vibration Damping

The VISS gimbal is connected to the UAV via a mechanical cluster of 4 Lord Mount Isolators built in square shape [Fig. 8]. This configuration enables vibrations to be damped along pitch, roll, and even yaw axes. Lord Mount Isolators are available in different strengths depending on camera weight. VISS uses hard strength for all experiments. When using too soft isolators, vibrations cannot be efficiently damped and might even cause further oscillations.

Layer 3 – Optical Image Stabilization

The VISS prototype uses the Panasonic DMC-TZ5 camera with an integrated "Mega O.I.S." stabilization technique. Two gyroscopes detect vibrations with a sampling rate of 4 kHz. Once a vibration is measured, the necessary compensation movements of the optical lens will be calculated. "Mega O.I.S" uses a linear motor to shift the optical image stabilizer lens of the camera.

Layer 4 – Digital video image stabilization

Real-time software enables stabilization without any significant time delay. However, VISS used the software "Virtual Dub" with the "Deshaker"-Plugin for test-purposes. Although its algorithm is not operating in real-time, configurations were made to simulate a performance similar to real-time stabilization software: Parameters of "Deshaker" were set to achieve fastest processing speed. For a 640x480pixel video clip with 29 frames per second, "Deshaker" was calculating all motion vectors and performing compensation motions with no noticeable time delay (Deshaker's image matching parameters were set on quarter scale with all pixels used in RGB mode). Tests were performed on a 2.4 GHz i5 processor with 8 GB RAM. "Deshaker" uses an area matching algorithm to calculate motion vectors. Based on these motion vectors, panning (yaw axis) and rotation (roll axis) of the video images are calculated and compensation motions are performed accordingly. However, compensation motions create empty areas in edges and borders. These areas are filled with image data of previous video frames.

IV. TEST RESULTS

A. Test equipment and test procedure

Throughout most stabilization tests, a Scorbot robot arm was used to perform reproducible sequences of vibrations that typically occur during UAV flights. The camera gimbal attached to the vibration isolator cluster was mounted on the robot arm to perform compensation motions. Software was programmed to measure horizontal, vertical and angular displacements of video images recorded by the camera. The final stabilized camera footage shows the quality of video images regarding blurriness and defocusing.

In an initial series of experiments, the basic functionality of each VISS layer was tested separately. Next, Layers 1+2 and Layers 3+4 were combined. Scorbot performed sets of vibrations with constant frequency and amplitude. Results show the importance of interaction between layers in the VISS architecture.

Eventually, the entire VISS system with all four layers was tested. Scorbot was simulating two UAV flights with vibrations of any amplitude. The first test flight included vibrations of low/medium frequency. The second test flight focused on medium/high frequency vibration compensation.

B. Basic functionality and interaction test

First, the angular displacement of the camera body along the pitch and roll axis was measured in Layer 1+2 – first individually and then by combining both layers. In a next step, the horizontal/vertical/angular displacement of video images was tested in Layer 3+4. During the basic functionality test series, the Scorbot was performing vibrations with constant large amplitude (±35°) at one selected frequency level (low, medium, or high) [Fig. 9].

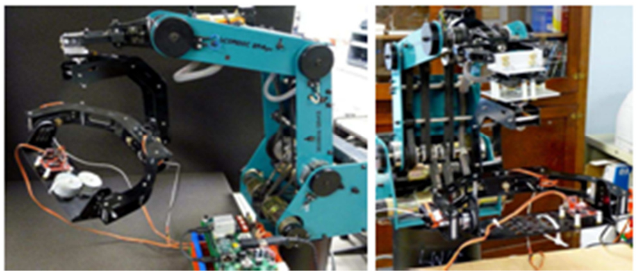

Fig. 9: Scorbot is testing basic functionality of Layer 1 (left) and 2 (right).

Angular displacement (Layer 1+2)

Layer 1 compensated for the entire amplitude spectrum. Low/medium frequencies were compensated with good results. As expected, Layer 1 fails to successfully compensate high frequencies due to time delays in performing counter movements [Fig. 10]. **Layer 2** damps the entire frequency spectrum and achieved good damping results for small/medium amplitudes. However, large amplitudes were only partially damped by Layer 2 due to its limited physical operational range of ±30° amplitude size.

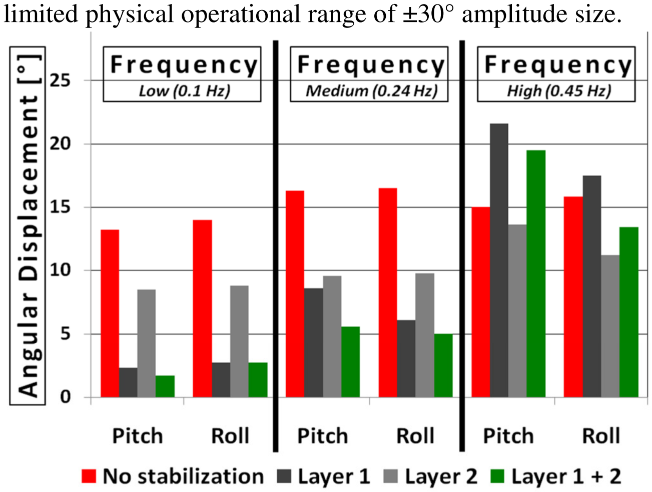

Fig. 10: Results of basic functionality test: Performance of Layer 1 and 2 along pitch and roll axis was tested separately and in combination.

Interaction of Layers 1+2: Layer 2 overcame the lack of Layer 1 to compensate for high frequencies. Results show an improvement of roll axis performance. However, the overall damping result showed its physical stabilization limits for the pitch axes. Damping high frequencies with large amplitude failed [Fig. 11]. However, these extreme types of vibrations rarely occur during UAV flights. In summary, test results show that the combination of Layer 1 and 2 improves the overall result when vibrations of the full frequency spectrum need to be stabilized. Except for high frequencies, the interaction of Layer 1 and 2 showed even better results compared to only using one of both layers. Amplitudes were damped to 1.7° (pitch), 2.7° (roll) for low frequencies and 5.6° (pitch), 5.0° (roll) for medium frequencies.

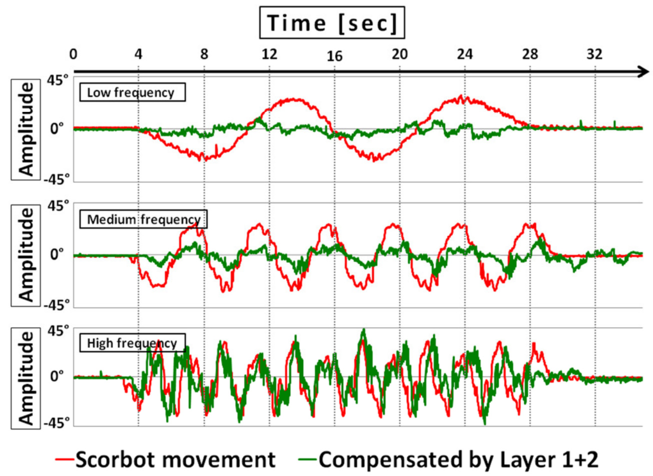

Fig. 11: Layer 1+2 achieve good results for low/medium frequency vibrations. No improvement for high frequency vibrations.

Horizontal/Vertical/Angular displacement (Layer 3+4)

Layer 3 solely showed qualitative results (reduced blurriness and defocusing) rather than quantitative improvements (reduced pixel displacements). The overall result was a smoother video image flow. **Layer 4** compensated for the entire frequency spectrum. However, Layer 4 was limited to eliminate only small amplitude vibrations because of the problem of creating empty borders and edges when rotating/translating the image. Medium/large amplitudes should be filtered by Layer 1 + 2.

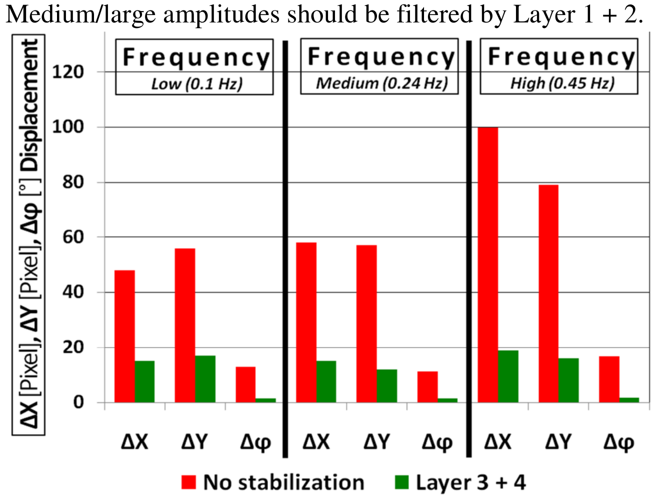

Fig. 12: Results of basic functionality test: Performance of Layer 3+4 was determined based on horizontal, vertical, and angular displacement.

Interaction of Layer 3 + 4 in the VISS architecture: The stabilization algorithms used in Layer 4 (Deshaker's block matching technique) performs slightly more reliable when blurriness and defocusing in video images was removed by Layer 3. The experiments showed positive results, both for horizontal, vertical, and angular displacement values [Fig. 12]. Horizontal and vertical displacements declined to less than 20 pixels for all frequencies. Angular displacements reached 1.5° for low and medium frequencies and 1.7° for high frequencies.

C. Simulating UAV flights- complex interaction test

The overall performance of VISS was tested in two simulated UAV flights. Video images were recorded and stabilized in real-time in Layer 1-3. Layer 4 simulated digital real-time stabilization with Virtual Dub's Deshaker Plugin.

The Scorbot arm was programmed to follow two different flight paths by varying its pitch and roll angle for 28 seconds. The first flight produced low/medium frequencies, while the second one contained medium/high frequencies. Vibrations of the entire amplitude spectrum were performed in both flights. The stabilization improvement was measured layer by layer [Fig. 11].

The overall result showed the importance of interaction between all VISS layers: Layer 1+2 significantly reduced the amplitude size of vibrations. Layers 3+4 further eliminated these vibrations with good results. The VISS architecture performed well throughout the entire frequency and amplitude spectrum. For the exceptional case of high frequencies and large amplitudes, stabilization only worked with partial success. Experiment results show that adding more layers will lead to a decreasing displacement in horizontal, vertical and angular direction. In addition, the image quality gets noticeably better regarding smoothness, blurriness and defocusing. Layer 1 significantly improves the horizontal, vertical and angular displacement. Layer 2 clearly contributes with overall damping. Layer 3 primarily avoids image blurriness and defocusing, but does not show major displacement improvements. Layer 4 almost fully eliminates the displacement along all axes. Eventually, demonstration videos were recorded and show that tracking objects felt significantly easier in VISS stabilized video images compared to raw video images [See attached video].

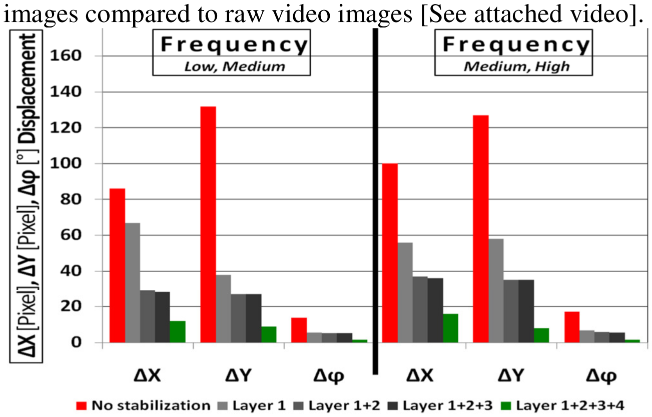

Fig. 13: Two UAV-flights with displacements measured layer by layer: Left: (low/medium freq.) Final error: 12 pixel (X), 9 pixel (Y), 1.5° (φ). Right: (medium/high freq.) Final error: 16 pixel (X), 8 pixel (Y), 1.6° (φ).

V. Future Improvements And Conclusion

VISS showed valuable results for improving camera footage that was recorded and filtered during the Scorbot test flights. The interaction between all four layers performed well in general, layers can still be varied, removed or substituted through further stabilization techniques if needed. Further research can be done to extend the VISS architecture: Layers should be disabled when their performance leads to a worse overall result. Exceptional cases e.g. high frequency/large amplitude vibrations should temporarily disable Layer 1 which could otherwise worsen the overall result [Fig. 10]. Further research should be done layer-wise: **Layer 1** only compensates along pitch/roll axis, but not yaw axis. Gyroscope drift prevents long-term stability when yaw axis is measured by the IMU. Non-inertia sensors have to be tested to stabilize the yaw axis via gimbal. **Layer 2** should be tested with more effective anti-vibration materials to show better results for damping high frequencies with large amplitudes. **Layer 4** was tested in static environments with a non-moving background. Implementing real-time stabilization software and testing its performance in dynamic environments would be necessary.

With an increasing number of UAVs being used for military, civil, and private purposes, the fields of applications are also increasing. Considering the growing number of vision-based UAV applications, reliable and real-time stabilization systems will always be critical for a UAV's successful performance. In future, it is likely that cameras will record images with even more optical zoom that will come along with higher vibration sensitivity and further stabilization challenges. By providing valuable image stabilization, we hope that VISS made a contribution to this development.

Acknowledgment

Supported by DARPA. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressly or implied, of the Defense Advanced Research Projects Agency or the U.S. Government. We thank the students at the USC iLab for their assistance with this work.

References

[1] Kimon P. Valavanis, “Advances in Unmanned Aerial Vehicles”, vol 33, Springer Netherlands, pp. 3-13, 407-430, 2007

[2] R. C. Voorhies, C. Siagian, L. Elazary, L. Itti, “Centralized Server Environment for Educational Robotics“, Proceedings of the IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS), 2009

[3] Eddy Vermeulen, “Real-time Video Stabilization For Moving Platforms”, 21st Bristol UAV Systems Conference, p. 3, April 2007

[4] Yasutaka Murata, Press Release of Murata Manufacturing Co., Ltd. “Bicycling Robot ‘Murata Boy’”, September 2005

[5] Marcelo C. Algrain, “Accelerometer-based platform stabilization”, Proceedings of SPIE, Vol. 1482, 367, May 2005

[6] B. Taylor, C. Bil, S. Watkins “Horizon Sensing Attitude Stabilization: A VMC Autopilot”, 18th Bristol UAV Systems Conference, 2003

[7] Vladislav Gavrilets, “Autonomous Aerobatic Maneuvering of Miniature Helicopters”, PhD Thesis, MIT, p. 53, May 2003

[8] “Understanding and Isolating Vibrations”, MicroPilot Newsletter Vol 1 Issue 2, pp. 1, 6, 2010

[9] Henri Eisenbeiss, “A mini unmanned aerial vehicle (UAV): System overview and image acquisition”, Int. Workshop on “Processing and visualization using high-resolution imagery”, p. 6, 2004

[10] RAMA UAV Control System, Article “Control System Architecture” http://rtime.felk.cvut.cz/helicopter/control_system_architecture, retrieved Mar. 08, 2011

[11] Vladislav Gavrilets, “Avionics Systems Development for Small Unmanned Aircraft”, Master Thesis, MIT, May 1998

[12] GE Intelligent Platforms, “Applied image processing” catalogue, Article “Image Stabilization”, p. 6, 2010

[13] M. Hansen, “Real-time Scene Stabilization and Mosaic Construction”, Proc. 2nd IEEE Workshop on App. Comp. Vision, p. 2, 1994

[14] Z. Duric, A. Rosenfeld, “Image Sequence Stabilization in Real Time”, Real Time Imaging 2, Academic Press Limited, pp. 12-14, 1996