# Spatial Compositing of Time

## Techniques for High and Low Speed Cinematography

Research Paper

Bachelor course on Media Technology
at St. Pölten University of Applied Sciences

By:

**Lorenz Pritz**

mt1310261068

Supervising tutor: FH-Prof. Mag. Markus Wintersberger

St. Pölten, 20.02.2016

# Declaration

- The attached research paper is my own, original work

- I have made no use of sources, materials or assistance other than those which have been openly and fully acknowledged in the text. If any part of another person's work has been quoted, this either appears in inverted commas or (if beyond a few lines) is indented.

- Any direct quotation or source of ideas has been identified in the text by author, date, and page number(s) immediately after such an item, and full details are provided in a reference list at the end of the text.

- I understand that any breach of the fair practice regulations may result in a mark of zero for this research paper and that it could also involve other repercussions.

St. Pölten, 20.02.2016 .....

Place, Date Signature

# Abstract

This paper will present an overview on techniques to combine and alter time-lapse as well as slow-motion imagery with the help of digital compositing tools and spatial information. With the spread of possibilities to produce accelerated and decelerated videos for a small budget, new ways have emerged to combine, enhance, and alter these videos in post production. When editing a sequence as well as digitally enhancing a single clip, spatial information is often used to create more credible and enjoyable motion effects. The goal of this text is to sum up possibilities for small video productions to use compositing tools in the process of editing and combining slow-motion and time-lapse videos.

The paper will cover topics such as flow motion time-lapse, smoothed egocentric hyper-lapse, motion estimation for slow motion, and selectively deanimated video. After the overview, there will be a section explaining several means of extracting spatial information from two dimensional footage either via software solutions or subjective, user based methods. This part is followed by a list of different ways the extracted data is used for the desired effects.

A chapter covering tests of the ascertained techniques and the findings of combining multiple different effects will conclude the paper.

# Table of contents

# 1 Introduction

This paper covers a selection of techniques that use spatial information, either estimated by user or extracted from video by software, to create visually interesting accelerated or decelerated videos.

Compositing is the art of combining multiple assets in video or film post production to create a seamless output. It is ubiquitous in film and TV as well as online videos. For completing most tasks, compositing has been a user based and work-intensive job. However, with recent advancements in the field of computer vision in combination with the increasing processing power of modern post production workstations, many former tedious tasks can now be completed in an efficient manner. Computer vision describes an area of research that has the purpose of developing methods for extracting data from digital video, which can then be used for a multitude of tasks such as robotics, medicine, or, as it is of interest in this paper, film and video production.

In addition to the facilitation of common tasks, new creative uses of video processing software algorithms have emerged over the last decade. Selectively de-animated video, automatic cinemagraph portraits, smoothed egocentric hyper-lapse, and flow motion hyper-lapse are four examples of new visual forms of presenting time in an interesting way with the help of spatial information. These four methods/techniques are presented in this paper, following a section covering the general differences between user and software based means of acquiring spatial information. The choice fell on these techniques as they are relatively new and therefore still hold a lot of creative potential for visual experiments.

There is an entanglement between user and software based compositing methods in video post production. This text highlights how manual ways of creating accelerated and decelerated visual media can benefit from the automatic algorithms of computer vision and vice versa.

# 2 Extracting spatial information from two dimensional video

## 2.1 Differentiating between user and software based methods

Video and film post production, as well as filmmaking in general, are often considered to be an art form. This viewpoint does extend even further than creative decision making such as choosing one shot over another in editing, altering the story with innovative juxtapositioning of plotlines, or drawing an animated stop motion sequence for an intro sequence. It also includes a creative and often courageous usage and/or deliberate misusage of technological possibilities in compositing software (Raman & Chaudhuri, 2007, p. 1). Moreover, the utilization of one's own experience is an important aid when merging multiple images to a single scene.

We as humans have multiple ways to perceive the world we live in. To successfully navigate through our environment, it is necessary to be aware of our surroundings near and far. Moving and experiencing the structure of our planet most of our daily lives creates a naturalness and vast amount of experience in estimating distances and other attributes of three dimensional space. This is often necessary when building a new believable scene from scratch or combining different elements with one another spatially.

When looking at a two dimensional video or film, one can quickly imagine how this scene would look like in three dimensional space. Working on the task of extracting spatial information from a two dimensional video, such as trying to separate fore, middle, and background from each other, is therefore a process that comes natural to many compositing artists. However, this user based process of splitting up a scene and creating a new one relies heavily on the understanding and skill of the artist and can never be more than a good estimation.

Moreover, compositing based on user experience is often a lengthy, sometimes inaccurate process, and hence not expedient in every situation. An example for such a scenario would be the task of inserting an artificial digital object in a scene with a moving and shaking hand camera. The user would need to invest a

considerable amount of time and effort to fulfill the task of creating several keyframes for the movement of the object. This is one of various scenarios where an automated process of extracting spatial information can prove useful.

However, software based solutions have limitations as well and can often only be used in a specific context and for a certain task. The current development of new algorithms and applications in the field of computer vision, which promise to be useful in a broader spectrum of tasks, helps to surmount these limitations little by little, but can still not provide accurate results in every case.

Currently, a combination of both user experience and the usage of fast but specific software based means of extracting three dimensional structural and/or movement data from a two dimensional video, is the manner in which most complex compositing tasks are completed.

## 2.2 A selection of software based methods

### 2.2.1 Structure-from-motion

Image reproduction via pixels on a screen or on traditional film faces the obstacle of having to display a three dimensional world on a two dimensional plane. The same problem basically occurs in our human vision system. Light gets reflected from a surface in a certain angle. If the light rays find their way into our eyes, a picture of the world in front of us gets projected on the retina. This image on the retina is a two dimensional image. The brain has then the task to recover depth information with the help of several cues. One example of such a cue would be the stereoscopic view, which is enabled by our pair of eyes and offers humans as well as animals an accurate three dimensional orientation in space, especially in close proximity to an object/person. The fixed distance between the eyes, in combination with the amount of differences in the projection of an object on the two individual retinas, is a set of information that can be processed by the brain to retrieve the three dimensional information.

Another way to recover depth and structural information from a planar image is motion. A good illustration for the human capability to recover three dimensional information from movement is a moving light display, as can be seen in figure 1. All other cues of the human vision system are masked out. Only the movement of the dots unveils that we are looking at a person walking from left to right (Shah, 2012a). In figure 2, a visual representation of the dots on a human body further illustrates the concept of a moving light display.

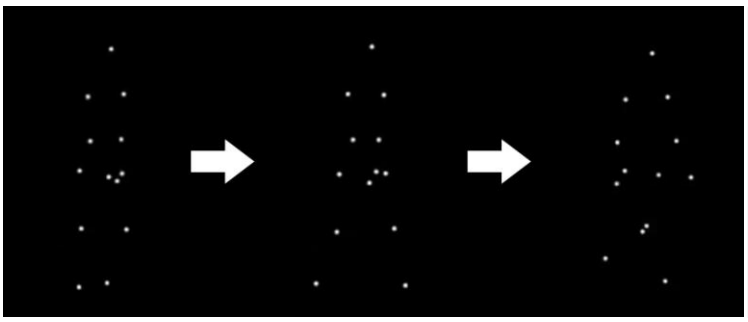

*Figure 1. (left) moving person from left to right, recognizable only by the movement of the dots in a moving light display (Shah, 2012a)*

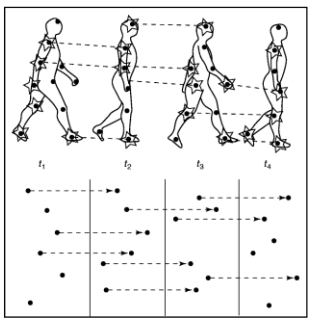

*Figure 2. (right) representation of the dots on a human body (Neri, Morrone, & Burr, 1998)*

The field of computer vision faces the same problems as human vision. The question is how spatial and especially three dimensional information can be recovered from an image that only consists of two dimensional data. Fortunately, the cues and principles humans use to perceive the world in three dimensions can sometimes be applied to the mathematical algorithms of computer vision. The usage of motion in the "Structure-from-Motion" algorithm is a good example of a cue that is both used to retrieve structural information from a planar image by software and human vision (Shah, 2012a). For the algorithm to function properly, the movement of the two dimensional image can either originate from camera or subject (Häming & Peters, 2010, pp. 926–927).

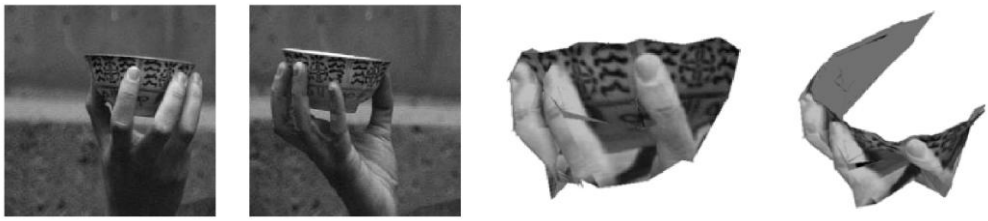

*Figure 3. example of a visualized data output (right) from an orthographic factorization method of the structure from motion algorithm (Tomasi & Kanade, n.d., pp. 152–153)*

Structure from motion and three dimensional data recovery in general can be useful in many applications, such as object recognition, robotics, computer graphics, image retrieval, geo-localization, archeology, and sports. Figure 3 shows the example of object recognition of a small bowl with the algorithm. Microsoft did a lot of research in the area of three dimensional data recovery and this led to developments such as the kinect sensor, which has multiple sensors/cameras to deliver depth information of an object in order to recognize specific gestures. These gestures can then be used to navigate a character in a game for instance (Shah, 2012a). The development of a method for creating smooth "First-person Hyper-lapse Videos" from shaky camera footage can be

considered as a more recent, solely software based, usage of the structure-from-motion algorithm (Kopf, F. Cohen, & Szeliski, 2014, p. 1). This method, which will be discussed in depth in section 4.1 of this paper, has again been developed by a team from Microsoft Research.

### 2.2.2 Kanade-Lucas-Tomasi (KLT) tracking

Tracking in general is the process of choosing a set of target pixels of an object and assigning them to a search area with the goal of following the position of this target area of the object over the time it is visible in the frame (Shah, 2012b). It can be noted that there are essentially two possibilities to choose a group of target pixels to track. The first option is to give a user the means via software tools, such as the camera tracker in After Effects, to mark and select a certain area of the object that should be tracked.

The second and more complicated option is to determine a set of features a tracker should look for in a group of pixels beforehand and let the software choose the tracking area accordingly and automatically. The movement of this cluster of pixels as well as the change over time due to change of perspective or object orientation should be estimated and compensated as good as possible by the algorithm. Robustness as well as accuracy are important features of a high quality tracker (Shah, 2012b).

Kanade-Lucas-Tomasi (KLT) feature tracking is a useful algorithm to track an object and its trajectory automatically, as it is able to compensate rotation, scale and even the perspectival change of an object (Shah, 2012c). To do so it uses an affine motion model (Shi & Tomasi, 1994). The algorithm can be used to track single or multiple objects such as airplanes or cars as well as single or multiple persons. Moreover, specific features of an object can be tracked particularly well with the help of Harris corner feature detection. Harris corner detection is an often used interest point detector. Interest point detection essentially describes the task of finding a set of pixels that is particularly unique in the frame and easy to track. The versatile KLT-tracking algorithm is useful in both tracking a video from a fixed as well as moving camera position. Furthermore, the algorithm can be used for tracking an object/feature which appears in multiple camera views. An example of the visualized output of a KLT track can be seen in figure 4 (Shah, 2012c).

*Figure 4. visualized output of a KLT track (Shah, 2012c)*

The KLT algorithm consists essentially of mathematical operations that can be summed up in the following steps:

1. Detect Harris corners in the first frame
2. For each Harris corner compute motion (translation or affine) between consecutive frames
3. Link motion vectors in successive frames to get a track for each Harris point
4. Introduce new Harris points by applying Harris detector at every *m* (10 or 15) frames
5. Track new and old Harris points using steps 1-3. (Shah, 2012c)

A current example where this algorithm is utilized is the automatic creation of cinemagraph portraits (Bai, Agarwala, Agrawala, & Ramamoorthi, 2013, p. 3). This topic will be covered in section 3.2 of this paper.

# 3 Techniques for decelerated video using spatial information

## 3.1 Selectively de-animated video

Selectively de-animating video can be defined by selecting an area of a video that should be ridded of all movement, in order to shift and focus the attention of the viewer on another area, where the movement is preserved. The selection can either be done manually via user based software tools or automatically with the help of interest point detection algorithms. Traditional means to achieve this technique are time intensive and require considerable skill in compositing software. This chapter will focus on a semi-automatic method, developed by a team at the University of California, Berkeley in collaboration with a team from Adobe, which allows a user to produce good results with little effort. The goal of this "Selectively De-Animating Video" method is to remove large scale motions from a video in order to make fine scale motions visible (Bai, Agarwala, Agrawala, & Ramamoorthi, 2012, p. 1).

As can be seen in figure 5, a man plays a guitar and although the camera is not moving a lot, there are still several areas in the video that contain motion and distract the viewer from the fine movement of the hands. This can be seen in the picture on the left. Since the interest of the viewer lies presumably in observing the hand movement, the rest of the input video is de-animated into a still picture. The output can be seen on the picture on the right (Bai et al., 2012).

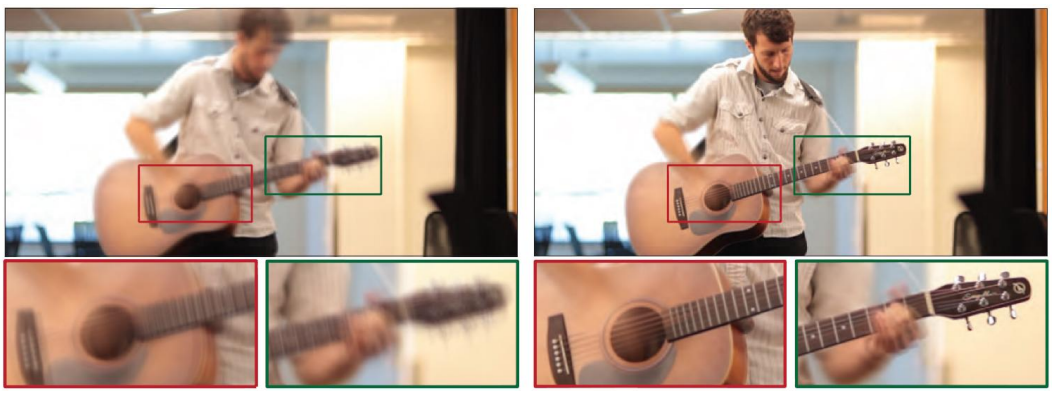

*Figure 5. – average input motion in comparison to the average output motion, which is only preserved in the area of the hands (Bai et al., 2012).*

The method works in a semi-automatic way, which means in this case that the user has the obligation to select the areas in the input video that he wishes to be de-animated. This can be done in a two step process with a digital software tool that can be compared to a brush, by drawing one of three types of strokes on the video. The first step is to draw strokes in the areas of the video that should be de-animated. After that, so called compositing strokes, consisting of static and dynamic strokes, have to be drawn for the purpose of optimizing the output and removing for instance warp in the de-animated areas. Distortions in the de-animated areas can sometimes be induced in step one. Based on these user selections, the algorithm can then track, warp, and composite the final output. A visual representation of the whole process can be seen in figure 6 (Bai et al., 2012, p. 3).

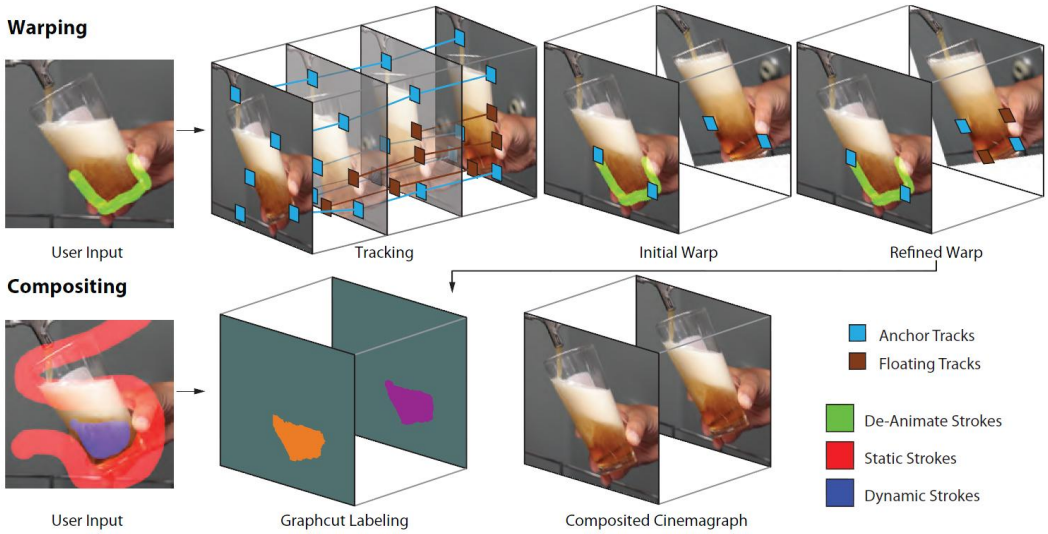


*Figure 6. – representation of the two step process of warping and compositing the video based on user drawn strokes to create a bottomless beer glass (Bai et al., 2012, p. 3)*

Moreover, if specified by the user, the method can create an automatic seamless loop that makes it possible for the video to be played endlessly (Bai et al., 2012, p. 5). This is particularly useful, for instance when exporting to a .gif format.

Possible applications of this technique can be found in tasks such as the creation of artistic cinemagraphs, clear motion visualization, or the analysis of movements in sports and science. Furthermore, in traditional compositing tasks, such as adding a logo to a moving object, this method of selectively de-animating specific areas can prove useful. This can be achieved by creating a video reference frame that enables the user to add an object to the deanimated area. The clip can then be processed back to show the original movement, with the only

difference that the desired object now moves realistically with the movement source (Bai et al., 2012, pp. 8–9).

Nevertheless, as most automatic and semi-automatic techniques this method also has its limitations. The amount of movement in the input video should not be too large. This would make accurate tracks nearly impossible, which are necessary for a high quality output. For example, large scale three dimensional motions can generally not be removed accurately by this method (Bai et al., 2012, p. 9).

## 3.2 Automatic cinemagraph portraits

The difference between cinemagraphs and other, similarly de-animated video clips can be defined by the artistic focus of cinemagraphs compared to a broad field of analytical and scientific purposes of other de-animated video clips.

Cinemagraph portraits are the most frequently produced and also widest known example of de-animated video. The visual effect can be compared to a portrait photograph with the difference that there are fine movements preserved and seamlessly looped in the area of the eyes, mouth, and/or hair. For the viewer, this typically makes a portrait more lifelike and visually interesting. For this specific task the same team that created the method for selectively de-animating video, which has been covered in section 3.1, developed a method for “Automatic Cinemagraph Portraits” creation a year later (Bai et al., 2013, p. 17).

This easy to use approach works by automatically processing a video portrait of a person, which can even be shot handheld. An illustration of the steps of the method can be seen in figure 7. In the first stage, the facial features of the subject of the video are tracked with a KLT feature tracking algorithm. The accuracy of the spatial information of the position of the face and torso are substantial for the following steps. Thereafter, the movement areas are split in two segments. Large scale motions, such as the movement of the torso or the background, are removed and fine scale motions, such as dynamic facial expressions, are retained. This is achieved by a spatially varying warp in stage two. Stage three then produces the final output by compositing the static areas with the dynamic parts of the cinemagraph (Bai et al., 2013, pp. 17–18).

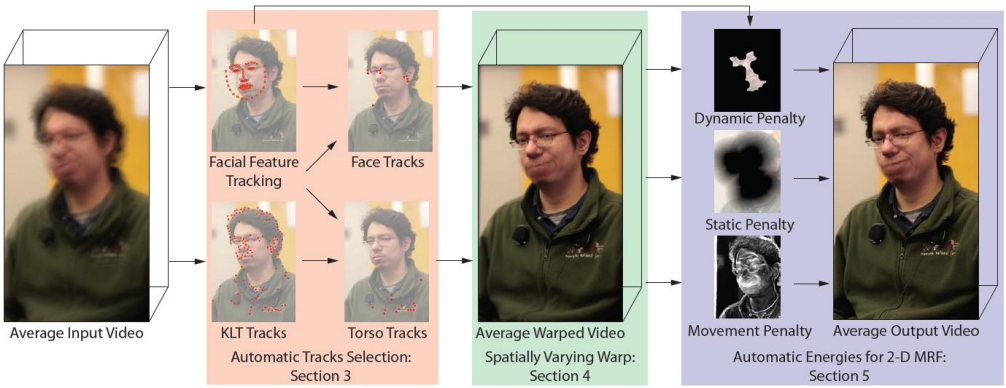


*Figure 7. – process of automatically creating a cinemagraph portrait from a shaky handheld input video (Bai et al., 2013, p. 18)*

However, this technique also has its limitations. As mentioned above, it is heavily dependent on the accuracy of the facial feature tracker. Three dimensional movements of the face for instance, can induce problems in the track and cause flawed results. Another restriction of the method is that it only works for a very specific task and context, namely the automatic creation of cinemagraph portraits. For instance, this process would not work if one wanted to create a different type of artistic cinemagraph, such as the bottomless beer glass example from figure 6 in section 3.1 of this paper. Nevertheless, this method creates great results and is easy to use, as it works fully automatic.

# 4 Techniques for accelerated video using spatial information

## 4.1 Stabilized egocentric hyper-lapse

Before discussing how egocentric hyper-lapse videos can be created, it is helpful to define the term hyper-lapse. This term is an alteration of the better known concept of time-lapse. Whereas time-lapse describes the process of taking fewer pictures (frames) per second than the designated playback framerate of the video is, in order to speed up observable time which is also known as undercranking in film jargon, hyper-lapse emphasizes the movement of the camera through space during the capture of time-lapse footage (Kopf et al., 2014, p. 1). The technique of hyper-lapsing can be achieved in multiple ways and has been a recent trend in online videos (“Video Marketing Trends for 2015,” n.d.).

The difficulty in creating watchable hyper-lapse videos is compensating the unevenness of the designated path of the camera. The camera is typically mounted on a tripod or monopod and the perspective is usually adjusted to keep a specific object/subject in the same position in each frame after every movement step. This is done in order to maintain a reference point for the eye of the viewer to focus on, as the large scale motions of the video can otherwise quickly become tiresome to watch. Creating smooth hyper-lapse footage is therefore a complex task that needs to be well planned beforehand. This task seems especially difficult considering that slower camera movements need a lot of effort to be stabilized as well, with equipment such as dollies, steadycams, or three axis gimbals. All in all, this technique requires knowledge as well as planning and is thus reserved for professionals and enthusiasts. However, this restriction only applies to hyper-lapse that is supposed to be usable out of the camera, with only minor help from stabilization in post production. A new approach, developed by a team from Microsoft Research, promises to make the creation of hyper-lapsed videos possible for a broader range of users and scenarios.

The method that has been proposed by this team for the creation of “First-person Hyper-lapse Videos” introduces a way of creating smooth time-lapsed videos from shaky footage that has not been recorded with the guidelines for usable hyper-lapse creation, as stated above, in mind. Other scenarios include action sports where it is almost impossible to follow an even path with the camera as

well as maintaining an object/subject fixed in approximately the same position in every frame (Kopf et al., 2014, p. 1).

The expectations of a casual user of action cameras, such as the GoPro or Sony action cam, when creating hyper-lapse videos are also a little different from the mainly aesthetic goals of video production professionals. Long and continuous clips, recorded during hiking, climbing, or simply walking around a city with a first person point of view mounted camera, are typically boring to watch and hard to navigate in editing software. Therefore, it is practical to create a type of "summary" of a certain activity by speeding up the footage to a reasonable timespan. Unfortunately, camera shake gets amplified and the results are, most of the time, not pleasing to watch. This is where the hyper-lapse creation method from Microsoft Research, consisting in simplified terms of a collection of software algorithms, comes into play. Its main purpose is to smooth and stabilize the sped up input video to a watchable clip (Kopf et al., 2014, pp. 1–2).

This complex task is accomplished by the iteration of several steps, of which most draw on already known algorithms from the field of computer vision, such as the in section 2.2.1 covered structure-from-motion algorithm. The necessary computations can be categorized in three stages: "1. Scene reconstruction, 2. Path planning, 3. Image-based rendering" (Kopf et al., 2014, p. 2).

In more detail, the first step of stage one is preparing the footage, which means dewarping the individual frames if they are distorted from a wide angle lens, as often found in action cameras such as the GoPro. This is done in order to prepare the video for easier processing in the next steps. Subsequently, reconstructing the scene in which the camera moves is achieved by the creation of dense depth maps as well as the iteration of a structure-from-motion algorithm. The goal is to recover as much information as possible about the three dimensional structure and the shape of the scene from each input frame of the video. The tracking that is necessary to compare frames with one another in these algorithms also enables a first decision about which frames will be kept for further processing and eventual use in the final sped up video. Typically, frames that are texture rich enable a good track and are thus more ideal to be used in the final clip, in comparison to blurred frames and/or frames that are otherwise poor in texture (Kopf et al., 2014, pp. 2–3).

This information about the reconstructed scene is then built upon by the planning of a smoothed path which the camera is supposed to follow. The task of planning an ideal path encompasses four objectives. The first one is to keep the length of the path reasonable (Kopf et al., 2014, p. 4). The second objective is to ensure a smooth path by optimizing orientation and position. The third one is to try to keep the output path as close to the input camera path as possible, in order to ensure a fast and good quality render of the final output frames in stage three. The fourth

goal is similar to the third, as its purpose is once more ensuring high quality of the rendered output video. However, this objective is not directed towards optimizing the proximity of the input camera to the virtual one, but instead towards determining if enough usable frames from stage one are in sight of the virtual camera for further processing (Kopf et al., 2014, p. 4).

In the third and concluding stage the final output frames are composited and rendered. One output frame consists of averagely 3-5 input frames that are stitched and blended together. This creates the possibility to crop less into the video, while maintaining a better stabilized result, than necessary in other available methods, such as the Adobe Warp Stabilizer for instance (Kopf et al., 2014, pp. 6–10).

Three stages of the algorithm from scene reconstruction to the final stitched and blended output can be seen in figure 8.

*Figure 8. from left to right: scene reconstruction, creation of proxy geometry from a virtual camera perspective, composited final output (Kopf et al., 2014)*

In the field of compositing, this is a prime example of how the usage of multiple types of spatial information can benefit and simplify otherwise very complex or, as in this specific case of stabilizing shaky handheld footage and compositing new output frames for a smooth hyper-lapse video, almost impossible tasks. This development that enables stabilizing very difficult videos paves the way to new creative uses of else not usable footage when compositing a new scene. It is both beneficial to the post production workflow of professionals and the need for simple creative solutions of consumers. The proposed method is already available for consumers in the form of a mobile as well as desktop software application.

## 4.2 Flow motion hyper-lapse sequences

### 4.2.1 Definition

Hyper-lapse, as mentioned in the previous chapter, describes a time-lapse clip with camera movement over a significant distance. The term flow-motion, which has been coined by Rob Whitworth, a famous hyper and time-lapse video production artist, focuses instead on a seamless and visually exciting way of transitioning between multiple time or hyper-lapse clips. The words “flow” and “motion” are used to illustrate the way the viewer is supposed to experience his virtual journey through space whilst watching the video. It seems as if the camera would fly from one scene to another with very few visible cuts (Walczak, 2015). In comparison to the automatic hyper-lapse method for egocentric camera movements in section 4.1, the flow-motion technique heavily relies on the creativity, skill and experience of the compositing artist.

Flow-motion is also a collective term that encompasses several often used techniques, such as hyper-lapse and infinite zooms, in the creation of a smooth flowing video with significant camera movement through space and time. It can also be stated that the desired output of a video production of this kind is a camera that explores a scene or city as protagonist, and by doing so, guides the viewer through a destination in a visually exciting way (Whitworth, 2015b).

Time-lapse clips are most of the time comprised of single pictures taken in a certain interval with digital single lens reflex or mirrorless photo cameras over a longer period of time. The shooting duration and interval depend on the movement that the videographer wishes to unveil. This possibility to utilize photo cameras allows the production of videos in a high resolution, often with 4000 (4K) or even 8000 (8K) pixels in a single horizontal line of one frame of the sequence. Moreover, the option of taking raw or high dynamic range (HDR) images, as well as the utilization of long exposure times, opens up possibilities to shoot in extremely difficult lighting conditions. However, despite the fact that the advent of advanced technological camera and post production tools generates a lot of creative flexibility, the big amount of data that has to be processed and stored is something to be considered for finding and maintaining an efficient workflow in post production.

It is therefore crucial to keep the whole workflow in mind while planning, shooting, and editing. Planning in general can be considered the most important stage when realizing a flow-motion project. The ability to visualize how the finished video will eventually look, in combination with documents that support the clarity of the concept, such as a storyboard, facilitates not only the production and post production process, but also helps to present one’s ideas to another person. A

client meeting, for instance, almost always benefits from a well presented concept via storyboard (Whitworth, 2015b).

Rob Whitworth shared one of his workflows for producing a single hyper-lapse clip, which starts with planning a shot, followed by taking the actual raw images, which are then organized in the software Adobe Lightroom. Subsequently, to process and enhance the raw images the application LRTimelapse is used. The compositing work is then done in Adobe After Effects and the final sequence, with the soundtrack and color grading added, is eventually edited and rendered in the non-linear editor Adobe Premiere Pro (Famà, 2014).

### 4.2.2 Flow motion transitions

The real creative task however, is not finding an efficient workflow but instead coming up with innovative and seamless ways of transitioning from one scene to another. There are some techniques that can be found in a number of flow-motion videos. The following section will cover general requisites as well as the steps necessary to create three of these flow-motion techniques and the importance of spatial information in them.

Generally speaking, not more gear is needed for a flow-motion project than for standard time-lapse video. The equipment can consist of a single stills camera with an interval timer and a tri or monopod. Compared to other camera setups with movement this is relatively convenient and cost efficient gear (Walczak, 2015). Besides this flexibility advantage over other production contexts, the additional effort lies in the especially important planning phase. When merging one scene with another seamlessly, parameters such as color temperature, movement speed, direction of movement, perspective, focal length and the resulting fore, middle, and background ratio have to be thoroughly thought through, aside from traditional reflections about interval and duration of the time or hyper-lapse. Due to changing conditions of lighting, shooting time has to be scheduled as well (Walczak, 2015).

The first technique is skipping a certain amount of time between taking two shots and blending them together to emphasize the change that happened on that location. This is a great option for showing the changing color of leaves in different seasons or the progress of a building project for instance. Figure 9 shows the railway station in Bielsko-Biała, Poland, and juxtaposes an old postcard from 1911 to a new time-lapse shot taken in 2014.

*Figure 9. transitioning from 1911 to 2014 via choosing the same camera position (Walczak, 2015)*

For the production of a similar sequence it is crucial to find the approximate position the first picture has been captured from, as well as a resembling altitude of the sun. Also, the focal length has to be known or if there is no information available, as is the case with the postcard in the example, estimated as good as possible. The shot that needs to be recreated is ideally on-hand in some form at the location when creating the neighboring shot. Suitable possibilities would be taking the last frame of the first scene and printing it or transferring the picture on a phone, tablet, or laptop that is also present at the location (Walczak, 2015).

The second technique that is often used to transition from one scene to another requires compositing skill, as it is a zoom in on another video asset that seems to be part of the scene. This asset is framed in some way inside the preceding shot, for example inside a window of the roof of a theatre or inside the back window of a car, as can be seen in Figure 10. The zoom can either be achieved by digitally cropping or manually increasing focal length after each captured frame. In most situations digital zoom is preferred, because it is easier to composite two clips with a consistent framing and apply a zoom effect afterwards. However, the individual frames have to provide enough resolution in order to maintain an acceptable image quality throughout the digital zoom (Walczak, 2015).

*Figure 10. manually zooming in on a tracked video as flow-motion transition (Walczak, 2015)*

The car sequence is also an example for a situation where manually changing the focal length is done to accomplish the zoom. The reason for this is that there was no possibility for the videographer to shoot a time-lapse, as the car is already fast moving. Therefore, traditional video had to be the chosen, which offers less resolution than pictures. This situation only offered the possibility to zoom manually without losing image quality (Walczak, 2015).

To be able to further improve the alignment of two clips and to digitally enhance the smoothness of the clips, stabilizing algorithms such as the Adobe Warp Stabilizer are often used. Also three dimensional camera tracking algorithms can be valuable when compositing an object believably in a scene. Compared to

setting keyframes for every frame by hand, the utilization of three dimensional camera tracking algorithms makes it possible to fulfill this task almost automatically. The development of the concept of flow-motion has been facilitated by the advent of these relatively easy to use single, multiple, and three dimensional point tracking tools in compositing software. Moving shots can be stabilized to look as smooth as if they would have been captured on expensive dolly tracks. Moreover, with the high resolution of modern photo cameras even fake camera movements can be created that are hard to be distinguished from, for instance, a real pan on set that tracks a person. The big amount of data also generates a lot of flexibility in compositing. An example for a digitally optimized track pan can be seen in figure 11. The white lines are indicating that a point tracking based stabilization method must have been used in post production, because of the same fixed position of the head of the ice figure skater in every frame.

*Figure 11. digitally optimized pan with a point tracking stabilization method (Whitworth, 2015a)*

A third transitioning technique, which probably requires the most effort and skill in compositing of the one's discussed in this section, is the illusion of a very fast flight from one location to another. To achieve the illusion of covering very big distances, several shots have to be combined during compositing in a three dimensional coordinate system in which a virtual camera can then move over the separated layers in order to create the seemingly realistic yet digitally created camera flight. This can again be seen in several films by Rob Whitworth, such as "Barcelona GO" or "Dubai Flow Motion". This is also the technique that requires the biggest amount of user estimated spatial information, as the individual layers need to be given coordinates in three dimensional space manually according to the compositing artist's opinion. Figure 12 shows an example of the virtual movement over the individual layers of such a transition.

*Figure 12. Virtual camera movement over multiple individual composited layers (Whitworth, 2015a)*

All in all, flow-motion is a combination of both user estimated spatial information and the help of digital algorithms. At this point it should be stated that several other techniques can be considered a part of the flow-motion concept as well. Also, flow-motion sequences are especially interesting to the viewer when they incorporate a vivid sound-design that supplements the movement of the visuals.

# 5 Practical evaluation of selected techniques

## 5.1 Workflow evaluation: stabilized egocentric hyper-lapse

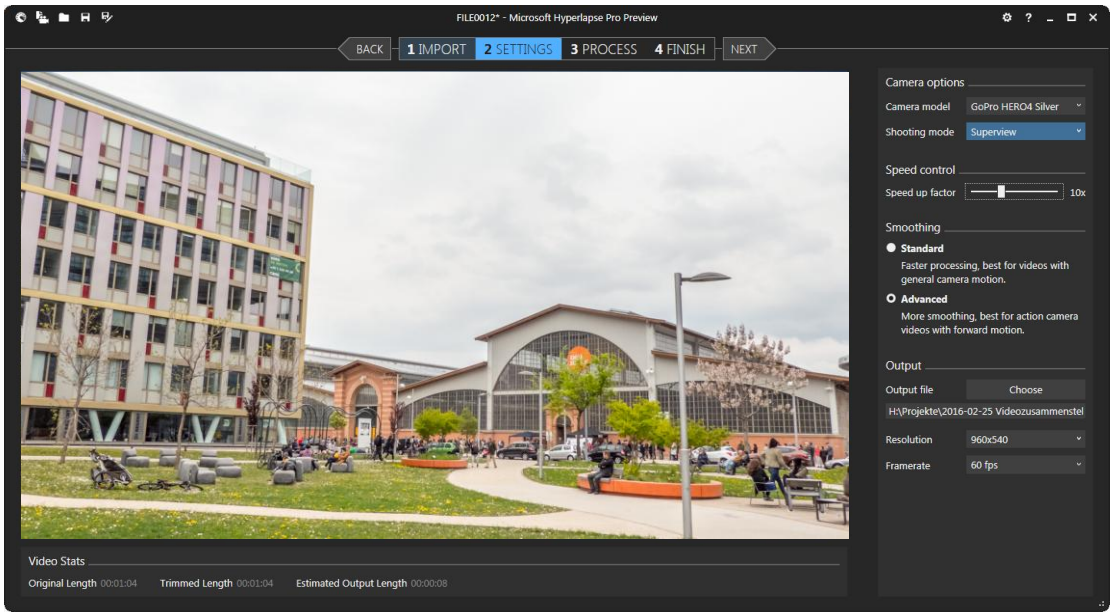


*Figure 13. Graphical User Interface Microsoft Hyperlapse v.1.1.56, settings tab*

This chapter covers the practical evaluation of my workflow for the creation of an egocentric hyper-lapse video in late 2015. I want to compare the effort necessary as well as the visible result of creating a stabilized video file with both Adobe's Warp stabilizer and Microsoft's Hyperlapse software. For this test I chose an exhibition hall in Vienna and took my photo camera in order to expose one picture every two steps on a direct walking path towards the hall. The camera was set up on a monopod.

I then took those raw images and edited them for maximum visual data in the application Adobe Camera Raw. By maximum visual data I refer to taking the information of the raw images and compressing the dynamic range by putting all the luminance information that is hidden in the raw data of the image in a, for my human vision, visible area. I accomplished this with Adobe Camera Raw. Although I did not have any knowledge of the code of Adobe's or Microsoft's video editing software I assumed that I will increase the quality of the stabilization if the software's code is written for looking at lower information dense video

information rather than working with raw photographic image data. I then imported this image sequence, together with the xml data created by Camera Raw, into Adobe After Effects for encoding into a new file. This step was necessary in my case; however it does not have to be if the processing power of the designated workstation is high enough. As my workstation is not powerful enough to handle a video comprised of raw images of twelve megapixels each, I exported the image sequence into an .avi file with the same dimensions but easier to process data; i.e. less data. Moreover, at the time of my test only Adobe After Effects accepted raw image sequences. Therefore, using a raw image sequence would have limited my stabilization test solely to Adobe's warp stabilizer. The resulting .avi file enabled me to get great video quality with high software compatibility.

The second step in the creation of my stabilized hyper-lapse video was to apply the Microsoft and Adobe algorithms to the footage. First I took my .avi file and imported it into both Adobe After Effects and Microsoft Hyperlapse Pro (trial version). In Adobe After Effects I then selected the effects tab and did a search for warp stabilizer. I drag and dropped the effect on the designated video layer in the layers tab. The video was then analyzed by the algorithm and stabilized afterwards. For the analysis there are two options: regular and detailed analysis. As I wanted to test the most efficient workflow, I chose the regular depth for the analysis. For the stabilization, which uses the extracted spatial information from the analysis, there are multiple options. These options can be combined into two main options: type of stabilization (warp, position, position scale and rotation, perspective) and amount of smoothness. The standard for type is subspace warp, and the standard value for smoothness is 50 percent. An increased value for smoothness correlates with an increased crop factor of the final video if there should be no black edges appearing around the video after the stabilization. The video was now stabilized and I could watch the smoothed result after a little pre-render. Subsequently I exported the clip into a new file. My second approach to stabilize the footage was with the software Microsoft Hyperlapse Pro (trial version). The process was fairly straightforward. The software prompted me to create a new project, import a video, apply settings such as speed up factor and used camera in the capturing of the video. The next step has been the processing step where the algorithm was applied. At the end of this step the video was exported and saved to a new file.

In my particular test, I was more impressed by the algorithm and the quality of Microsoft's Hyperlapse stabilization approach than with the results of the multipurpose stabilization algorithm inside Adobe's video editing software. However, there are some downsides in my particular case that come with the usage of Microsoft Hyperlapse Pro. The main one is that I need to go to extra software outside of my non linear editing system (NLE, i.e. Premiere, Avid etc.).

This results in an extra conversion step for the video, which is not ideal when trying to maintain maximum quality. However, the biggest upside I saw in my tests with Microsoft Hyperlapse Pro was that, when handed long shaky videos, Microsoft Hyperlapse still managed to create respectable results, whereas Adobe's warp stabilizer cropped the image to its maximum of 150 percent and showed a clearly overchallenged stabilization.

In the next figure one can observe the different crop factors of the two algorithms. In the fourth panel, I took the video exported from Microsoft Hyperlapse and applied Adobe's warp stabilizer to combine both. Despite the quality of the stabilization being the smoothest, the video quality suffered the most.


*Figure 14. Crop factors in different stabilization workflows*

The videos of this test can be found on my YouTube channel via this Link: https://youtu.be/HjzPK4leowo
(0:00 – 0:40 Hyper-lapse stabilization and Flow Motion Hyper-Lapse)

## 5.2 Workflow evaluation: flow-motion hyper-lapse sequence

In this chapter I will review my experience trying to create a flow motion hyper-lapse sequence and the tools I utilized.

The shoot of the raw image sequences did not differ from other time-lapse shoots I completed. However, one additional thing needed to be taken care of before going on location and capturing the footage. The most important thing was planning the designated “path” in the pre-production phase. Knowing where the sequence started in shot number one and via which perspectives it would eventually move to shot x at the end of the sequence has not just sped up the shooting time on location, as it often turns out with extra planning in other projects as well, it was simply necessary. Without planning a designated path the whole finished sequence would not fit in the category of a flow motion video. After I had captured the raw images I went on to process them for maximum visual data a subjective appeal in Adobe Camera Raw and subsequently exported a temporary .avi file, with which I could efficiently work on my workstation. The next part was the one I experienced as the most enjoyable. Creating a flowing "high speed flying montage" through a specific area was very enjoyable, as the time and hyper-lapse clips did not only portray each focal element in front of the lens, they also put them into a spatial context of the specific area.

Flow motion is a collective term used to describe a multitude of possible transitions and effects. However, a very common transition is to place the focal element of the next shot in the first one. At the end of clip one, the video artist zooms **into** the video in a fast and blurred manner until the size of the focal element of the next frame is the same size as the zoomed in element. A fast crossfade, similar to a match cut\crossfade used in some instances in high budget movie production contexts, is used to transition to the next shot. The video artist then zooms **out** of the second clip in order to create the illusion of a seamless transition without a visible cut. I utilized this way of transitioning in the test sequence as shown in figure 15. Moreover, sound effects are often heavily used to enhance the transitions and therefore I added some to my test as well. As tool to create this sequence I chose Adobe’s After Effects, which can be seen in figure 16. This application proved as the optimal tool for me as it enabled me to also use stabilization algorithms for some shaky time and hyper-lapse shots that I wanted to improve before compositing.

The final video can also be found online on my YouTube channel via this link: https://youtu.be/HjzPK4leowo
(0:40 – 1:24 Hyper-lapse stabilization and Flow Motion Hyper-Lapse)

*Figure 15. Zoom transition in the test video flow motion sequence*

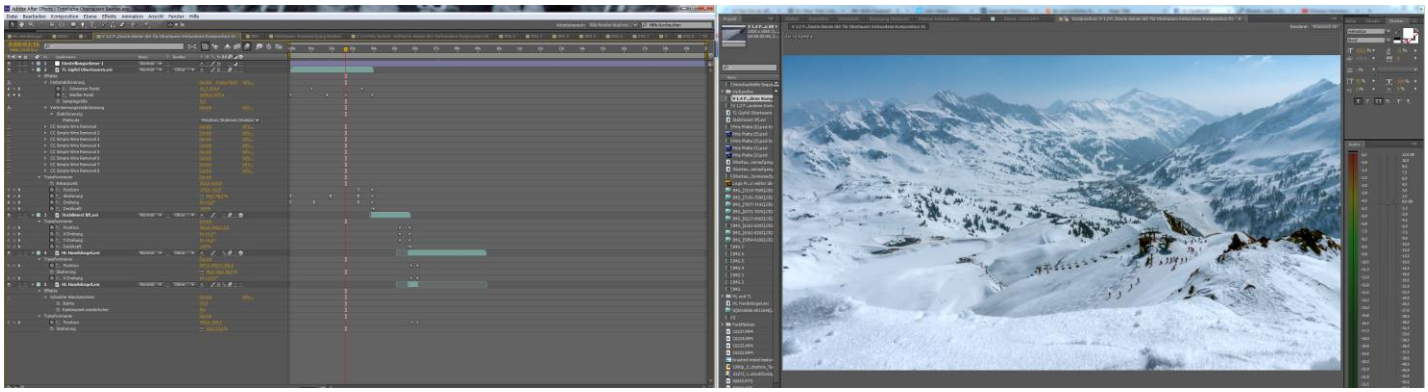

*Figure 16. Flow motion project opened in dual monitor After Effects interface*

# 6 Conclusion and future outlook

This paper presented four techniques that illustrate new forms of accelerated and decelerated video.

Selectively de-animated video and automatic cinemagraphs help to focus the attention of the viewer on details and fine movements. The output of these techniques is a hybrid between video and photo. The user, as well as the software algorithm, “reflects” on which movement is important and which is expendable to create an interesting clip.

A method for the creation of stabilized egocentric hyper-lapse shows a new way of automatic stabilization of extremely shaky footage, which would not have been usable in video post production unstabilized. This method also opens up new possibilities to create an interesting type of summary of long and otherwise boring videos captured from a first person view, such as a continuous clip recorded during hiking with an action camera.

Flow motion hyper-lapse sequences allow for a visually exciting seamless and fast trip through a destination. The viewer can experience accelerated video in combination with camera movement, edited to a seamless, almost flight like, experience. This is achieved by the usage of different possible flowing transitions. These transitions can be done in a multitude of ways and still hold a lot of creative potential to experiment.

Moreover, many computer vision solutions found an application in video compositing tasks. Video stabilization, three dimensional camera tracking and structure from motion are few examples of this development. It can be safely presumed that many more innovations in this area will pave the way to new, unseen creative accelerated and decelerated videos. However, software algorithms have to find a way to be useful in a broader spectrum of tasks, as they are today mainly expedient in specific contexts.

Everything considered, the future outlook for fast and slow-motion video post production promises a visually exciting and user friendly experience.

# References


Bai, J., Agarwala, A., Agrawala, M., & Ramamoorthi, R. (2012). Selectively de-animating video. *ACM Trans. Graph.*, *31*(4), 66.

Bai, J., Agarwala, A., Agrawala, M., & Ramamoorthi, R. (2013). Automatic Cinemagraph Portraits. *Computer Graphics Forum*, *32*(4), 17–25. http://doi.org/10.1111/cgf.12147

Famà, M. (2014, 06). Talking to Rob Whitworth, the Master of time-lapse Post-Production. Retrieved from http://timelapsenetwork.com/interviews/talking-rob-whitworth-master-timelapse-postproduction/

Häming, K., & Peters, G. (2010). The structure-from-motion reconstruction pipeline – a survey with focus on short image sequences. *Kybernetika*, *46*(5), 926–937.

Kopf, J., F. Cohen, M., & Szeliski, R. (2014, August). First-person Hyper-lapse Videos. ACM Transactions on Graphics (Proc. SIGGRAPH 2014)(Web).

Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature*, *395*(6705), 894–896. http://doi.org/10.1038/27661

Raman, S., & Chaudhuri, S. (2007). A Matte-less, Variational Approach to Automatic Scene Compositing. In *IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007* (pp. 1–6). http://doi.org/10.1109/ICCV.2007.4408901

Shah, M. (2012a). ▸*Lecture 15: Structure from Motion - YouTube*. Retrieved from https://www.youtube.com/watch?v=zdKX7Xo3Cb8

Shah, M. (2012b). *Lecture 04 - Interest Point Detection - YouTube*. Retrieved from https://www.youtube.com/watch?v=_qgKQGsuKeQ

Shah, M. (2012c). *Microsoft PowerPoint - Lecture-10Allignment - Lecture-10-KLT.pdf*. Retrieved from http://crcv.ucf.edu/courses/CAP5415/Fall2012/Lecture-10-KLT.pdf

Shi, J., & Tomasi, C. (1994). Good features to track. In *, 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94* (pp. 593–600). http://doi.org/10.1109/CVPR.1994.323794

Tomasi, C., & Kanade, T. (n.d.). Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, *9*(2), 137–154. http://doi.org/10.1007/BF00129684

Video Marketing Trends for 2015. (n.d.). Retrieved July 12, 2015, from http://www.rockadove.co.uk/blog/2015/01/05/video-marketing-trends-for-2015/

Walczak, T. (2015, March 22). Flow Motion tutorial: how to create a flowing hyper-lapse video. Retrieved from http://timelapsenetwork.com/mini-tutorials/flow-motion-tutorial-create-flowing-hyperlapse-video/

Whitworth, R. (2015a). *Dubai Flow Motion*. Retrieved from https://vimeo.com/117770305

Whitworth, R. (2015b, March 22). Flow Motion tutorial: how to create a flowing hyper-lapse video. Retrieved from http://timelapsenetwork.com/mini-tutorials/flow-motion-tutorial-create-flowing-hyperlapse-video/

# List of figures