# Abstract

This project experiments with algorithms which could address problems in Time – lapse photography, where image processing for an image sequence might be required to obtain better visual effects. Specifically, image registration is essential when the images are not aligned decently, which could result in flickering at play back stage, a motion – estimation based algorithm is applied to align images after searching for the best reference image. In addition, when a large motion is presented between two frames, a new frame could be created and this indicates adjusting the play back timescale to be non - linear, as a consequence, the movements become relatively slower and more details are shown to the audiences.

# Acknowledgements

I would like to express my thanks to Mr. Colin Dalton, for his constant encouragement, invaluable advice and guidance throughout the project.

Thanks also to Fan Zhang, for his help on technical problems, and to Tao Shi, Bingqin Chen and Zonglin Du for providing photographic equipment.

Sincere thanks to my family, friends for their support, help and encouragement during this project.

# Contents

# Introduction

## Non – Linear Time – Lapse Photography

Time – lapse photography, by definition, is a process containing long – time image sequence capturing and then playing back at some rate (usually 24 frames per second), as a piece of animation, for example, a sunrise and sunset procedure can be recorded by digital camera every five minutes during an entire day, and software can be used for the short film construction.

Since the capturing stage covers a long time, a tripod could ensure the camera in the same position, however, if capture work lasts for weeks, it could be extremely difficult to keep the tripod in the exactly same position, therefore, image registration is required to avoid flickering when they are playing back.

Time appears to be moving fast and thus lapsing when the image sequence is played at a certain frequency, which is often 24 frames per second, to put another way, 24 images are played in sequence on screen within a second. If such frequency is modified, from the audiences’ perspective, objects in the scene would move faster (frequency increased) or slower (reduced), which, indicates a non – linear playing back rate.

In this project, non – linear time is implemented by image interpolation techniques, if new images containing intermediate positions between two neighboring frames, it would be tantamount to lessen the frame rate, for a smaller number of images will be presented per second and thus, the speed of moving parts in the scene would be felt slower.

## Aims and Objectives

This project explores algorithms that address problems in capture and playback phases of time – lapse photography. When capturing images, a tripod is indispensable to fix the position of a camera, however, usually it could be massively time – consuming for this process which could cover longer than weeks, thus a practical problem is that it is difficult to keep the camera exactly in the same position, and the first aim of this project is to find a proper method to implement registration for image sequence, including an algorithm to align two images as well as deciding the best image to act as the reference.

* Objectives for image registration
* Implementation of an algorithm to align two images.
* Apply the algorithm to image sequence.
* Experiment with the algorithm for acceleration and robustness.

In the playback stage, the main aim is the non – linear time, on the one hand, to increase the speed, removal could be easily applied in an certain software, thus the project primarily focus on the opposite aspect – to reduce the playing back frequency. Image interpolation is utilized to achieve this aim.

* Objectives for image interpolation
* Implement an algorithm which is able to create a new image between two neighboring frames.
* Apply the algorithm to image sequence.
* Experiment with the algorithm to enhance the plausibility of the new frame.
* Find methods to automatically decide where to create new frames.
* Algorithm acceleration

## Structure

This report consists of five main sections, firstly, background of relative methods and techniques in conjunction of previous research, are described in the following section, where a number of registration and interpolation methods are included. In section 3, there are methods designed for this project, including those are unsuccessful after experiments. Then the implementation and results section mainly explains thresholds selection and parameter configuration for relatively better segmentation, and thus obtaining a plausible result. The next section is a discussion of the results from experiments, followed by an evaluation of the project to determine if it has obtained the objectives. Finally future work is discussed in different aspects, including software developing and extensions of present methods.

# Background

## Image Segmentation Methods

In this part I described two fundamental image segmentation techniques that are relative to this project, one is to divide an image into two parts by a threshold and the other is based on sub – divisions.

### Thresholding

Thresholding requires gray – level images and is the relatively simpler method segmentation process, it could works well when the image owns a decent contrast, which is suitable for a constant threshold to distinguish the background and objects in image. A basic thresholding algorithm can be described as follows:

**Algorithm 1 Basic Thresholding**

Search all pixels f(x, y) of the image f. An image element g(x, y) of the segmented image will be categorized as object if f(x, y) > T, where T denotes the threshold, otherwise, it belongs to the background.

There are a number of methods to determine T, usually detecting the mean or median gray value could be suitable for segmentation with brightness. Mean gray value is calculated from the mean value of all pixels, while median is often the gray value of the pixel at the median position after sorting all pixels.

### Region – Based Segmentation

A natural region – based method for segmentation is region growing. Beginning from the raw image data after sub – division, two adjacent regions will be merged if they satisfy some merging criterion, then try the next region until no pair matches the condition. On the opposite, segmentation could also be implemented by splitting to small regions from the entire image, while furthermore, splitting and merge techniques could be used simultaneously.

When merging and splitting are combined, it works in a hierarchical data structure (M. Sonka, et al, 1999), images are sufficiently subdivided when not matching a homogeneity criterion, then merge adjacent regions when then could be in a homogeneous region.

### Summary

Thresholding and region – based methods are basic but significant techniques for image segmentation, in this project, segmentation is utilized before processing images and some thoughts are adopted from these fundamental algorithms.

## Optical Flow and Motion Estimation

### Optical Flow

Optical flow indicates the motions in images during a small time interval, including movements from objects, edges, surfaces and even the entire scene (camera movement). Basic optical flow estimation is on the basis of the Optical Flow Equation (OFE)

(1)

Where is the image sequence, denote the 2 – D velocity in the image plane respectively, while are partial derivates of intensity within a frame and between frames. Equation (1) indicates that intensity along motion trajectory remains constant.

### Motion Estimation in H.264 / AVC

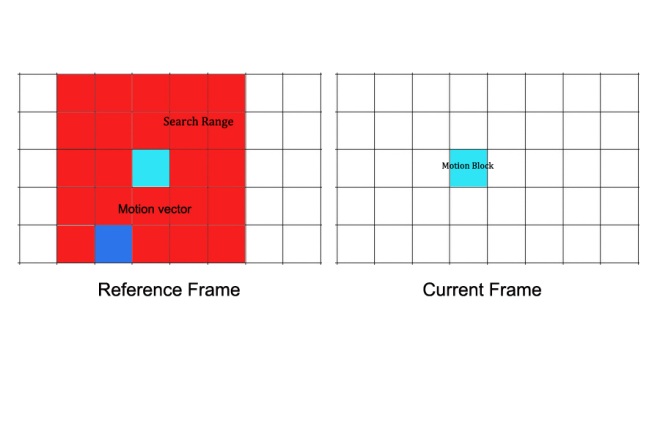


Figure 1 Basic concept of BMA

H. 264 / AVC is a commonly used compression, recording and distribution standard for high definition video, and motion estimation techniques form the core of such applications (Iain Richardson, 2010). A motion vector is used for representing motion information from one frame to a reference frame, and an image could be sub – divided into small motion blocks (usually square - shaped), each motion block contains a motion vector when compare to the reference.

Block Matching Algorithm (BMA) (I.E.G Richardson, 2003) is a popular method in motion estimation, as shown in Figure 1, the algorithm is described as follows:

**Algorithm 2 Block Matching**

1. For one motion block in the current frame, we compare it within the search range (a larger region containing same size blocks around the corresponding block) in the reference frame, for each comparison.
2. A motion vector is recorded along with some figures that denote similarity.
3. Finally, the motion vector with the largest similarity will be selected, and all pixels in the block are regarded as having the same motion vector.
4. Repeat from 1) until all the motion blocks are checked in the current frame.

### Summary

As researching with motion problems, the basic concept is the principle of optical flow – constant intensity along motion trajectory, and the BMA in motion estimation could implement segmentation to distinguish the background and the moving objects between frames, this is fairly significant to create new images with the moving part in the correct position.

## Image Registration

### Introduction

Image registration is the process of transformation of coordination systems, from different ones into one. In this project, registration mainly is represented by image alignment operations, which could avoid flickering when the image sequence is played back, it is significant since the photographer cannot ensure that the tripod are installed at exactly the same position throughout the entire capturing phase, which could last longer than weeks or even months. In addition, as I experimented in practice, even if the tripod position are carefully marked, it is difficult to capture the a consistent sequence after moving away from original place, thus, registration is indispensable for generating a more reasonable sequence.

Another factor in time – lapse photography that requires consideration is the registration should only be applied to the background rather than the motion parts, which should actually be intentionally neglected when implementing the registration algorithm, therefore, segmentation is essential before alignment.

### Previous Work

Image registration has been applied to different regions, for example, virtual reality (A. Aseem, et al, 2006), analysis of medical images (K. Loewke, et al, 2008) or for military purposes (N. Heinze, et al, 2008). There are also a number of methods to implement alignment process, the method introduced in (Q. Zhang, et al, 2011) is a popular feature matching algorithm, with the combination of Fourier Transform and SIFT feature point extraction for efficiency and robustness purposes, while when applied to time – lapse photography, features from motion objects might also be detected and thus affects the results. (F. L. Seixas, et al 2008) proposed an Genetic Algorithm (GA) – image registration method, creating a new perspective for matching and mapping point sets, but it is difficult to learn images for such a substantial quantity.

An automatic and robust algorithm introduced (Q. Zhang, et al, 2011) which is implemented by matching features of two images, Fourier Transform (FT) is employed for relative points calculation and Scale-invariant Feature Transform (SIFT) can be decent algorithm boils down to retaining features like illumination, rotation, or geometric distortion. Their experiments have shown that to combine FT and SIFT will be less expensive in extracting feature points. In FT, image transformation pair is assumed as *f(x, y)* and *F(ξ, η)*, with displacement *(Δx, Δy)*, thus the transformation can be expressed as:

(2)

For two images, *f1(x, y)* and *f2(x, y)*, their relationship in time domain can be assumed as

(3)

Therefore, FT of each image can be obtained from (1) and (2):

(4)

At this stage, phase correlation function can be calculated by cross power spectrum of inverse FT:

(5)

where *(Δx, Δy)* is a non - zero position of phase correlation function which values in zero at most positions, and it is the shift that expected for image overlap region segmentation.

The algorithm preferred at early stage is (G. Ward, 2003), which is original designed for High Dynamic Range (HDR) images composition, during which capturing of several source images is similar time – lapse photography. This method thresholds images after transforming into grayscale color space, then applies an exclusive – or (XOR) operator to show misaligned coordinates, before adjusting x and y offsets. This method neglects the variations in intensity in different images, while it could be true in time lapse photography, it is implemented in this project and results are discussed in section 4.

### Summary

In conclusion, as it is time – consuming for capturing images, and from practical experiments, registration is a first step in addressing problems in such image sequence, specifically, as a compensation for slightly movements of the camera to ensure a smoothly changes between frames, a number of algorithms could be adopted and then applied to the entire sequence.

## Image Interpolation

### Introduction

Current films are usually projected at 24 frames per second, which means 24 images appear in each second (linear time), an interesting challenge in this project is to adjust the playing back speed when necessary.

Image interpolation is the method that is employed in this project to implement a non – linear timescale, by definition, image interpolation can also be regarded as image scaling and resampling, which will increase the image size and resampling to a newly – created image plane. Particularly, in time – lapse, the resampled image need a motion prediction algorithm, which could create a new image that depicting the scene time – related but not captured by a digital camera, but rescaling is not included in the scope.

Implementation of non – linear time results in a more expressive film because when something actually occurs between frames, the story could be displayed more in detail to the audience when new frames are created, objects appear to move slower than before while on the opposite, the process could be accelerated when nothing moves or only slightly movements take place.

### Previous Work

A path – based algorithm is introduced in (D. Mahajan, et al, 2009), they constructed a path between two images where new images need to be inserted, and along this path, pixel gradients could be easily copy and move to implement the interpolation, and the frequency contents of intermediate images could be reserved without ghosting and blurring, with temporal coherence maintained. (T. Stich, et al, 2011) proposed a method attempting to adapt human visual system for such image sequence, which included the indispensable motion prediction process. The two methods seem to perform in their own characters and I will describe them in detail respectively.

The basic idea of (D. Mahajan, et al, 2009) is to create some ‘motion path’ to lead the variation of pixels between two input images, the basic notion of such path is described in Figure 2, (a) defined two simple images constructed by single lines (blue background not included), and at t = 1 the image is a left – shifted version of the one at t = 0, they are together regarded as the input images in this example, Figure 2(b) shows that for images could be interpolated between the inputs, and each time, the image is actually shift to the left at a step of 1 pixel, character ‘f’ is the information on pixel p0, which will be interpolated at pixel p0 – 1, in the next image, so the algorithm copies ‘f’, from the input t = 0 image, and then copies ‘g’, ‘h’, respectively, after that, the copy operation will be moved to t1 image, which is indicated by the orange line in Figure 2(b), and this process comprise the path. In the t = 0 image, the p0 + 2 pixel (recording ‘h’) is the point that the path changes direction, this is defined as transition point, where the intensity values of two images match, and hence, the interpolation could be plausibly implemented. In addition, the path could be represented by two general motion vectors mA and mB, thus path could be described as Equation (21).

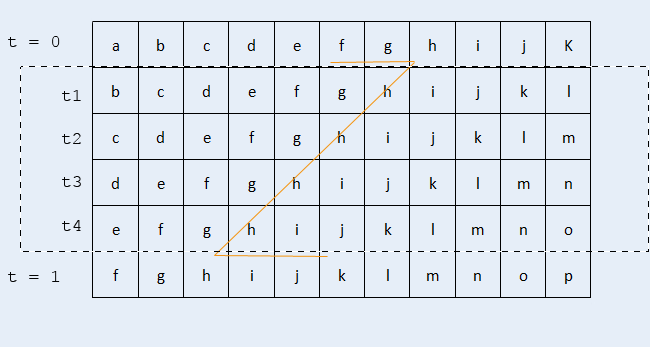
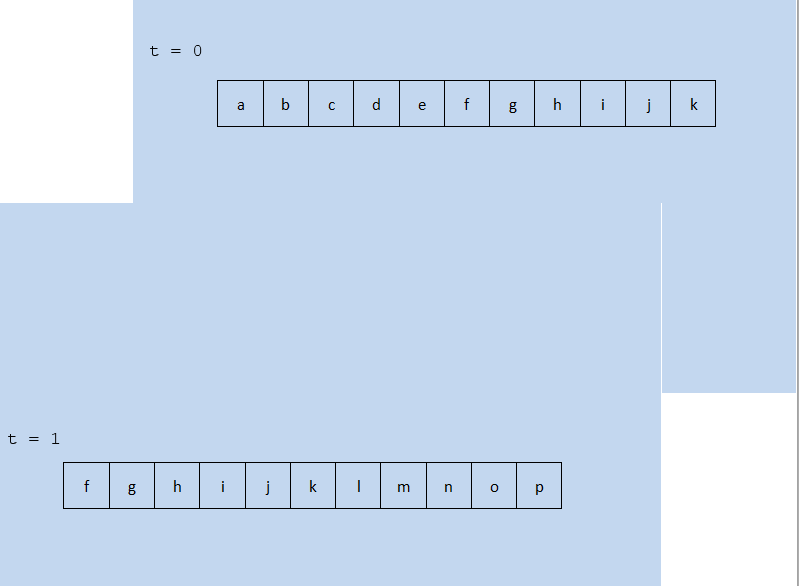


Figure 2 Basic notion of path

(a)

(b)

*p0*

Relative to interpolation, computing has been applied to optical flow as well as video segmentation (C. L. Zitnick, et al, 2005), but due to implicit occlusion reasoning, it is difficult to present some occlusions. Graph cuts are also employed for global optimization (V. Kolmogorov, et al, 2001) and (J. Sun, et al, 2005), however, they are difficult for explicitly occlusions considerations. With regard to image warping and morphing, an image – based rendering method is proposed by (A. W. Fitzgibbon, et al, 2003) utilizes manual correspondence based on stero and optical flow. In addition, interpolation is also applied to video processing, (H. Wang, et al, 2004) proposed an algorithm for video editing based on 3D Poisson equation, while similarly in the gradient domain, with Poisson equation, there are applications for image editing(A. Agarwala, et al. 2004) and (J. Jia, et al, 2008). However these methods are not compatible to the interpolation in time – lapse photography because their inputs and outputs are both images or videos, which could be regarded as static methods.

### Summary

The key factor for image interpolation is trying to create the most plausible images, from the previous work, it could be seen that the new frame data should be copy from the original ones, in which way the consistency in intensity and color could be kept to the largest extent, a robust algorithm for new position computation is the key element during this process, in addition, occlusion handling should be carried out as well for the completeness of the newly created frame.

## Background Summary

This section has explained the reason for the registration and interpolation in this project, in conjunction of relative methods and algorithms of this project, from basic concepts to previous research work, algorithms that might be adopted are the project is described in detail.

# Methods

In this section all the methods and algorithms that are employed in this project are described, they are primarily divided into two parts as per the operations of registration and interpolation, firstly, the method from (G. Ward, 2003) are described in detail with modifications, then motion estimation based algorithm is found and applied to both operations.

## Registration for Image Sequence

This part contains methods for image registration, firstly, the thresholding and XOR algorithm proposed by (G. Ward, 2003) is adopted and explained in 3.1.1, then, inspired by image interpolation, I experimented with motion estimation to align images, finally, in section 3.1.3, I describe the methods that benefits the processing speed, which is fairly significant in such image sequence solution.

### Extension of a previous Framework

**Thresholding and XOR method overview**

The fast and robust method proposed by G. Ward was taken into consideration in the early, the registration for HDR image construction might be relevant to some extent, since HDR images, similarly, requires several images to be overlapped with aligned perfectly before enhancement in color, intensity and contrast. A hardware solution will be a tripod to capture multi – images, however, this algorithm is designed for hand – held exposures, which could cause similar camera movement when creating time – lapse image sequence, hence, it should be worth experimenting with.

The input to the algorithm is two grayscale images, which could be converted by the following formula[[1]](#footnote-1).

(6)

Then one of the two is selected as the reference, both images will be segmented by a median threshold, which is selected from sorting all the gray values, the segmentation algorithm is displayed below.

This algorithm could segment image according to the exposure, brighter part will be completely set to white while darker part will be black, an example from images captured in Edinburgh, Scotland are shown in Figure 3 (with original image, grayscale image and segmented result), the result contains only two values and could be easy to implement the following XOR operation.



Figure 3 Results of grayscale and thresholding segmented images

Left: Original Image Middle: Greyscale image Right: Segmented image

**Algorithm 3 Median Thresholding**

1. Compute grey values for all pixels in the image and then accumulating frequencies to form a histogram.
2. Begin from 0, count pixel numbers located in each grey value until they are add up to half of the pixel amount (image height multiplied by image width).
3. At this point, the value is selected as the threshold .
4. Segment image, if pixel value , set pixel value as 255 (black), otherwise, set values to 0 (white).

The identity of two images is measured by an exclusive – or (XOR) operator, which could get a value of 0 under the condition that two pixels own the same values, while non – zero if they are different (set to 255 for showing a recognizable result), an example of two unaligned images is shown in figure 4.



Figure 4 Result of XOR operation

Left and Middle: Segmented images Right: XOR result

In spite of the cloud in the sky, the outline of buildings indicates the difference, and it is noticeable that computing one image with itself will result in a completely black scene, thus to find the offsets for a pair, we select one as the reference, then apply XOR in four directions respectively, and calculate the total non – zero pixel quantity, the smallest direction will be regarded as the correct movement.

**Acceleration**



Figure 5 Image pyramid

When applied in practices, the basic method could encounter difficulties, naturally images from time – lapse photographer may have a High Definition (HD) resolution at 1920 \* 1080, this could result in a time – consuming pixel comparison, or worse, a larger range to detect the displacement in contrast to smaller – sized images. A common solution is to create an image pyramid, which resizes images down by a quarter (halves height and width respectively), with such technique, computation cost will be massively reduced. Figure 6 gives a concept of image pyramid.

Now that the source images are shrunk to smaller sizes, we could limit a pixel range for XOR comparison in horizontal (x) and vertical (y) directions respectively. Finally, if the image pyramid is created to levels, returning consequence from the accelerated algorithm will be linearly expanded by a factor of 2 in each level, to put another way, the final shifting magnitude is times that of the bottom level.

In conclusion, to align images via threshold and XOR operation is less complicated comparing to others (SIFT, GA, etc.), the key factor is to compare differences in two images, in conjunction with acceleration method, the results are shown in section 4.2.1.

### A Motion Estimation – Based Method

Due to the block matching algorithm could segment background and moving parts in an image, with motion information (motion vector) in each motion block, it is deducible that the motion vectors of the background parts could indicate the displacements between two images, simultaneously, for aligned images, that of moving objects could be a reference for interpolation, I will explain the registration in this section, while the interpolation in section 3.2.

As described in section 2.2.2, if two images are input to an algorithm based on the BMA and one is selected as reference, blocks in the other image could find similarities by an search range in the reference, and thus obtaining a motion vector encoding as motion information, if images are divided into a number of motion vectors, after computation, motion information are recorded for registration and interpolation.

**Motion block and search range**

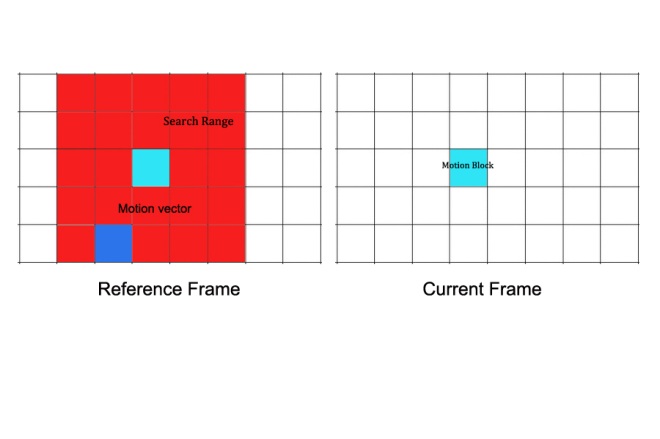


Figure 6 Sample Motion Block and search range demonstration

Block size: 8 \* 8, search range: 40 \* 40

The size of motion blocks is first set to 8 \* 8 pixels, with a - pixel distance search range, this is illustrated in Figure 6, which means that, each block will compare in the context up to two blocks shifting all directions, the comparison is to traversal in the search range with a step of one pixel, after that, we could get a reasonable motion vector indicates the smallest differences for this block. However, as experimented with HD images (1920 \* 1080), such size configurations are relatively small, which shoots up computation, simultaneously, another element to consider in interpolation is accuracy, which could be affected by size as well, so I left the size changeable, and results are discussed in section 4.4.1.

With respect to the measurement of block differences, in each block comparing pair, I compute the squared value of total pixel pair differences, this algorithm is demonstrated below.

**Algorithm 4 Block difference computation.**

1. For one motion block in the input image, start with the first pixel in the block, assume the input image has blocks, each block has pixels.
2. Subtract the pixel grey value with corresponding pixel in the reference image, square the result .
3. Add up to of all pixels in the block as an illustration of block difference, denoting by .
4. Traversal in the search range of this block from step 2) and 3), with step of 1 pixel, obtain to .
5. Find the smallest from to , record the motion vector for this block.
6. Repeat from 1) to 5) until all blocks in the input images are computed.

**Segmentation**

At this stage, motion vectors have been computed and then they should be categorized into two parts, background and foreground. In time – lapse photography, the foreground parts usually contain moving objects, and for photographer, camera will be carefully installed to avoid shifting, so an assumption here is that the background will move slower (own smaller vector norms) than the foreground, so that there will be a threshold for the motions.

It is noticeable that with algorithm 4, consider the background part, even if two images are aligned perfectly, the slight intensity difference might not result in a motion vector of (0, 0), hence, a key element for detecting the threshold is that it should be somehow large to avoid such incorrect election probability, a mean value is less reliable since zero vectors could still account for a certain proportion. As a consequence, there vectors are neglected when selecting the threshold, below listed the algorithm.

**Algorithm 5 Detection of motion vector threshold.**

1. Input all the motion vectors in algorithm 4, form a sequence , where is the quantity of motion blocks.
2. For , , compute , denotes the moving distance of a block.
3. Exclude all vectors when , and sum up other non – zero distances, with whose quantity accumulated, results in .
4. Threshold .

Now the threshold could be larger because zeros are intentionally disregarded, and then remaining data will result in a mean that could implement the segmentation, the vector norms of foreground blocks will be larger than , while the remaining belong to the background, an example result of segmentation is shown in Figure 7.It could be seen that the segmentation could roughly recognize the background (yellow part), but via data selection before shift, such errors are under tolerance.

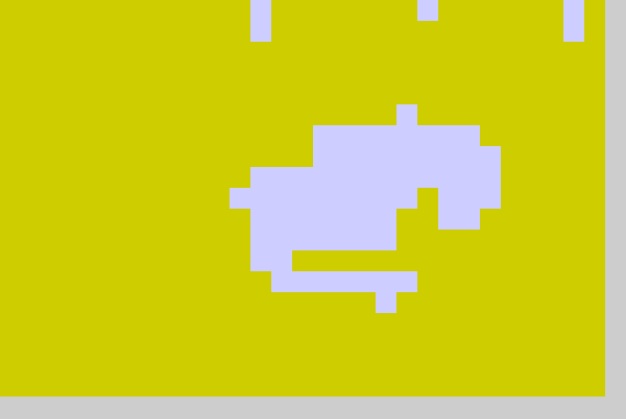


Figure 7 Sample segmentation by motion estimation

Left: Source image Right: Segmented result

**Shift the image**

In the practice of time – lapse photography, the moving parts should not be considered when applying alignment operation, and this is the reason I did the previous segmentation part, thus to obtain the shift vector for the input image, I select the motion vector that owns the largest frequency in the background part to avoid some error computation with intensity difference.

**Acceleration**

As demonstrated before, image pyramid is an effective method for acceleration, it equals to enlarge the block size in when using BMA algorithm, however, the level should be controlled to an extent for a relative reliable results, which is shown in section 4.4.1.

### Summary

In this part I employed two methods for image registration, the first one is done by an XOR operator and is simpler, but after experiments, it performed less robust, the second has a segmentation of background and foreground, which is an inherent character for time – lapse photography, the results turn out to be pleased.

## Interpolation

### Motion Estimation for Interpolation

An exciting point from motion estimation method is the segmentation results, which could be extend to interpolation, but this should under the condition that images have been aligned, otherwise, it could gain computation for registration despite of a successful interpolation.

**Motion vector selection**

As input to the interpolation algorithm are aligned images, only the motion blocks from foreground (vector norms larger than threshold) is considered, similarly, I count frequencies for all vectors, the largest is regarded as the motion information of the object.

**Creation of intermediate frames**

Now that we have computed how much the objects moves from one frame to another, the intermediate position could be estimated, firstly, the motions are considered not large, so that the problem could be simplified as linear, if one image is need to be interpolated, the new position should at half way from the previous frame to the next one, then at these positions, pixels could be copied from either the previous or next frame.

Accounting for the background part, which are considered as static, I set the mean values of the neighboring two frames since they are calibrated and will not cause offsets.

In addition, there could be two methods for creating more images, first, repeat previous process, and the second could been implemented when new position calculation, if one more images is needed, for example, two positions should be computed by a step of one third from reference frame to current frame rather than half of that. However, this is out of project scope so I did not apply experiments

**Occlusion handling**

In image interpolation problems, there will be empty pixels in the newly created frames, which is displayed in Figure 8, as part of the edge is not being filled. Particularly, when previous method is applied, there could be holes at some edges because the input images are regarded as a 2 – D array, when processing the blocks near the edges, there will be out – of – bound problems, which have been purposely disregarded when comparing blocks, it is reliable because after that the motion vectors with largest frequency is forecast as the actual motion, therefore, empty pixels in the new frame is filled the same way as copying background.



Figure 8 Empty pixels after interpolation

**Algorithm overview**

As described before, Algorithm 6 gives the process from selected motion vectors. It is currently designed for creating one image between images, if more novel frames are required, the process could be repeated.

**Algorithm 6 Novel frame creation (Considered for one image).**

1. Assume the foreground contains motion blocks with motion vectors, .
2. Select the vectors that owns largest frequency and estimated as motion vector, demoted by .
3. New position coordinates , where are the pixels positions in the current image, or , where denote the pixels positions in the reference image.
4. Copy pixels from the corresponding coordinates in current or reference frame.
5. For values of pixel in the background part of the new frame, , where denotes the pixel values of a pixel, note that the value here is in RGB colour channels.
6. Fill empty pixels with value calculation similar to 5).

### Summary

Motion estimation offers a linear and illustrative method for image interpolation, and if the correct positions of moving objects are able to achieve, plausible new frame could be obtained by copying pixels from source frames, in addition, it is indispensable to fill blank pixels caused by this method.

## Image Sequence Consideration

Previous methods still work for two images, it should be extend to the entire image sequence to create totally automatic program, only in this way, users are able to select image sequence and wait for their results. In this section, both registration and interpolation methods are developed for image sequence, but for the threshold and XOR algorithm, it is difficult to determine whether two images are aligned and hence less possible to find a correct reference point, as a consequence, it has been categorized as a unsuccessful attempt in this project.

### Find Reference Point for Registration

For two images to be aligned, either could be selected as the reference image, but in an image sequence, it is completely different, a correct reference point could massively save computation, for example, consider an image sequence consisting of 100 images that have been captured via a tripod, if there were slightly movement after capturing the 5th image, the fastest registration should apply calibration to the first five images rather than the last 95, so any one from image 6 to 100 could be selected as a reference, if unfortunately image 1 to 5 is selected, the registration algorithm should be run for 95 times, thus it is significant to consider the speed issue by detecting a best reference image.

**Assumptions**

The assumption for reference point detection is formed from time – lapse practice, usually a photographer will be careful with the tripod to avoid movement, but it is almost impossible to re – install the tripod to exactly the same position after movement, this could be true when the process covers a long time, three days, for example, the photographer could mark the first position but it is still difficult to ensure perfectly aligned images, and that is the reason why I do registration. Under such consideration, it could be assumed that the input images are actually divided as groups, each group are perfectly aligned because of usage of tripod (it could also be possible that only one image is in the group). To write the assumption into mathematical form, for all images in image sequence , they are divided into groups, which could be denoted by a set , their corresponding image quantities is written as .

Group2

RP

Group1

Group3

Figure 9 An illustration of reference point detection

Start

**Main process**

Figure 9 gives an example of three image groups compromising an image sequence, the algorithm starts from the first image in the first group, selecting the first image as reference point, then time, check one image afterwards, if the image pair is aligned, an accumulator indicating the quantity of images that are aligned with the reference will increase by one, otherwise, record the quantity, move the reference point the first unaligned image, then repeat this process until all the images have been checked. At this stage, the images are grouped with each group has an aligned image numbers, the reference point will be the first image of the group containing the largest quantity.

**Condition of alignment**

At this point, I found the thresholding – XOR algorithm less reliable for time – lapse images because the motion could cause large errors, the smallest difference in entire image may not indicate the right offsets, thus this algorithm is abandoned and regarded as a failure in this project, in contrast, motion estimation could generate a correct result, when comparing only in the background part, if most blocks return vectors (0, 0), this background could be estimated as still, representing the alignment.

**Algorithm description**

The following box demonstrates the entire algorithm.

**Algorithm 7 Reference point detection**

1. For Groups in image sequence in total images, aligned image quantities for each group are written as (to be detected).Set the first image as the reference point, set up an image quantity accumulator .
2. Check one image each time after the reference point, if image satisfies the aligned condition, , then check the next image.
3. If two images are not aligned at image , store to , then move the reference point to , , repeat from 2) until the last one.
4. Find the largest , set the final reference point to the first image of .

### Determine the Interpolation Points

Similarly to registration, not all the gap between two frames requires interpolation, only these are sufficiently large will apply the algorithm. From the a motion vector , the distance of motion could be calculated as . For an image sequence, I first detect all the distances between frames, then select the points that are within a specified range (set to 0.8 in test, could be set by user in practice) of the largest distance. Algorithm 8 in the next table describes this process.

**Algorithm 8 Interpolation point detection**

1. For image sequence with images , start from the first pair (index = 1).
2. Compute the motion information and get the motion distance , then index = index + 1, repeat 2) until the last pair.
3. Find the largest , set as ..
4. Restart from the first pair, if , do interpolation between this pair.

### Summary

When extend the methods to the entire image sequence, for registration, it is significant to consider the speed issue, a best reference must be selected for the sequence since it could save enormous amount of computation, with regard to image interpolation, the magnitude for creating new frame could be selected by user, smaller range will limit the condition of interpolation and thus accelerating.

## Methods Summary

In section 3, all the methods and algorithms that are employed in this project have been demonstrated in detail with figures, firstly, registration and interpolation algorithms are applied for only two images for testing the effectiveness and robustness, among there is a failure which employs thresholding and then compares images with an XOR operator. Then, comparison is implemented before a background and foreground segmentation which is done by motion estimation, it is exciting when motion estimation is positive with both image operations. Finally, the methods are extended to image sequence and could be able to run automatically.

# Implementation and results

The 4th section illustrates the results I have achieved with some implementation methods, in section 4.1 some data storing structures and objects are introduced, followed by some failure examples with analysis, then, motion estimation method results are shown in the next two sections respectively.

## Introduction

The entire project is implemented by C++ using Open Source Computer Vision (OpenCV), program environment is a 32bit Windows 7 Running on a virtual machine, with Microsoft Visual Studio 2010. Min configuration of this virtual system is listed in the following table.

Table 1 Test machine configuration

|  |  |
| --- | --- |
| **Component** | **Models** |
| CPU | Intel(R) Core(TM) i7 – 2620M @ 2.70GHz, 4 processors |
| Memory | 3GB @ 1333MHz |
| Hard Disk | 80GB |
| Graphic Card | Intel HD 3000 |

**Cameras used**

* Panasonic LUMIX DMC – LX5GK
* Nikon DSLR D80

The implementation of this project should first consider the image sequence problem, thus in this section, firstly vectors are introduced for storing image names, queue are utilized to store motion vectors, motion distances, which are described in section 4.1.2.

### Image Names

Due to the input to my program is an image sequence, and OpenCV can easily read images by names, I create string - typed vectors to store image names, the only parameter that left to the user is the address of image sequence directory.

After the directory is specified, my program can search all files inside and record all their names (appears with directory) into a string vector instance called ‘srcFileNames’, simultaneously, a symbol string ‘aligned\_’ is attached to each image name and they form another string vector ‘dstFileNames’, which is used for writing aligned output images. All input images have an output including their offsets is (0, 0), in such case, images are aligned to the reference so that they are only copied, which mainly boils down to the convenience for showing images in the program.

For showing the results of alignment, I write codes to overlay images with transparency reduced, in this way aligned images would be similar to the effect of showing a single image.

For image interpolation, there are slightly differences for output file names, when image pairs are computed for new frames, string prefix ‘inter\_’ will be attached to the name of reference image indicating that it is a newly created frame, for example, file name ’inter\_0001.jpg’ is the frame interpolated between images ‘0001.jpg’ and ‘0002.jpg’.

### Data Structure

Queue are mostly used to store motion vectors, to accumulate distance frequencies for data optimization, as shown in Figure 10, three structures are designed.

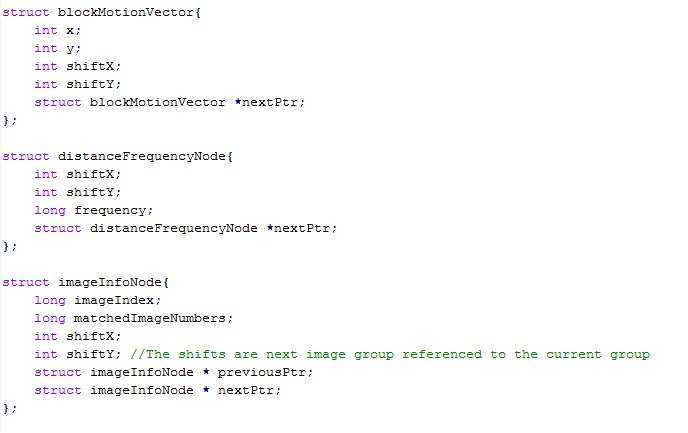


Figure 10 Structures definition

* **struct blockMotionVector**

For storing motion vectors when comparing two images, is an index of motion blocks, denotes the motion vector.

* **strct disranceFrequencyNode**

Used for counting frequencies of motion vectors.

* **struct imageInfoNode**

Only used for registration, this structure corresponds to the first images of each group, represents the image position with quantities of matched images stores in , the nodes form a bi – direction queue since the reference point need to reference to both nodes in the contexts.

Due to the First – In – First – Out (FIFO) character of queue, the motion information could be accessed strictly as the sequence has been stored, this is extremely significant for registration in image sequence, when file names are in sequence stored in vectors and could be accessed by an index, in that way, the elements in the vectors could be projected to nodes from , so that each image could easily get an shift vector when applying shifting operation.

## Unsuccessful Results

In this section, some results that are not pleased but significant to the project is illustrated, including failure examples of thresholding – XOR algorithm and some error in image interpolation.

### XOR Registration

Registration by comparing the differences of entire images could obtain positive results, but less reliable when motion is presented in images, a expressive example of this algorithm is shown in Figure 11, images are captured by the LUMIX in Edinburgh, with hand – held exposures, it can be seen that buildings are decently aligned in spite of the clouds in the sky, in the right side of the image, the bell tower is failed because of the massive distortion of the lens, which is set on a focal length of 24mm equivalent to a full – frame system.



Figure 11 A result aligned by the thresholding – XOR algorithm

Up: Overlay effect of source images Down: Aligned result

In contrast, the failure in Figure 12 shows almost no differences before and after registration, the model car is actually moving and should be neglected, but the background still kept displacements, I experimented with several other examples but unfortunately obtain similar results. It can be seen that the XOR algorithm is less robust, in further work, it could still be difficult to find a reasonable standard to determine whether two images are aligned and then apply shifting to images, as a consequence, this part of work has been classified as unsuccessful category.



Figure 12 Failure example of the XOR computation

In addition, (G. Ward, 2003) proposed a method for handling noises with an exclusion image, which is comprised by 0’s and 255’s, pixels are set to 0 if grey values are within a certain distance of the threshold, otherwise, they are set to 255, in this way, the noises near threshold could be shown in white points in the exclusion image, Figure 13 illustrates the noises in the red square in the actual pixels, and Figure 14 is an example of exclusion image of the 2 – value image in Figure 13, this exclusion image is computed by an AND operator with the result from XOR calculation, but the result turned out to be no differences.



Figure 13 Noises near threshold

### Summary

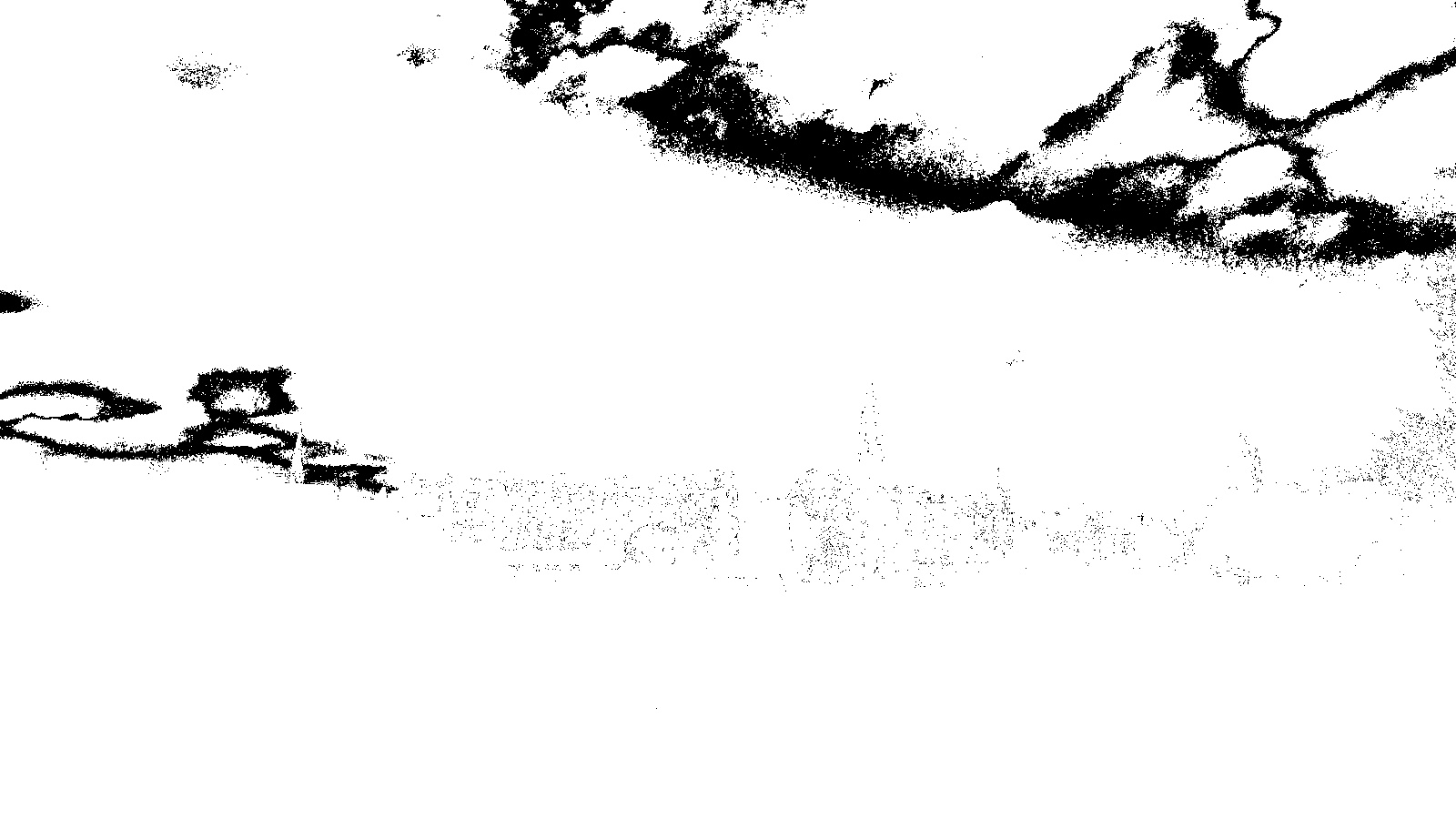


Figure 14 Exclusion image example

In conclusion, although the XOR algorithm could achieve some effect, when compare to motion estimation based method, computation in the entire image shows obvious disadvantages, and considering the difficulty for sequence extension, this algorithm is waived in the later part of the project.

## Motion Estimation Registration



Figure 14 Success alignment for the failure example of the XOR computation

Up: Unaligned results

Down: Aligned images

In this section a number of images that are aligned by motion estimation algorithm, including experiments with only two images and an entire image sequence.

### Results

In figure 14, the result that is not achieved by the thresholding - XOR method is shown, it is noticeable that the background are reasonably aligned, but the source images still representing the motion, and hence the overlapped image has ghosts of the model car. Another representative example is demonstrated in Figure 15, where the trees and mountains have been obviously calibrated, comparing to the model car image, segmentation of images of Figure 15 could gain difficulty because the moving parts are actually the cloud, while background, where should be apply alignment are the mountains and trees, which requires the motion blocks to be smaller (large blocks could fail, because background takes smaller proportion of the image size), and Figure 15 is in similar condition to the previous Edinburgh images.

The Figures have expressed the effectiveness of motion estimation method, when applied to image sequence, it could still detect the correct image reference point and apply registration, Figure 16 illustrates a result when experiment with an image sequence, which is purposely formed as Figure 9, Group 1, 2, 3 contains 5, 6, and 14 images respectively, and image index, relatively, starts from 0 to 24, the camera displacement is intentionally set and the program achieve an expected result with the reference image selected as image 11, thus only 11 images (image 0 to 10) requires registration rather than 14 if simply set the first image as reference. As experimented, each registration operation could cost up to roughly 1 second, therefore, it could be forecast that the computation would be massively reduced when the image sequence contains larger number of images.



Figure 17 Failure of motion estimation method



Figure 15 Another example aligned by motion estimation

Up: Unaligned results

Down: Aligned images

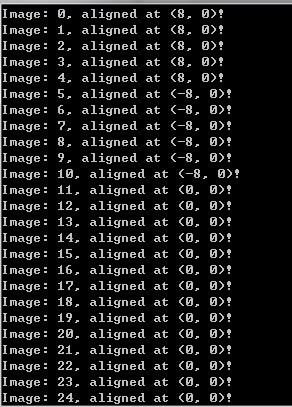


Figure 16 Image sequence registration information output

An unsuccessful example is shown in Figure 17, in there two frames, camera movement becomes abnormally large, which is just very close to the car model, but they move towards different directions, so that the segmentation is failed and thus has negative effects on image shifting.



Figure 18 Comparison before and after motion vector selection

Left: No selections Right: Selection by frequency

Two intermediate images are interpolated frames, the first and third row are source input images

### Summary

The motion estimation method, due to a background and foreground segmentation, could perform better than single XOR operation in the image, it could work on images with apparent background and moving objects, as well as wide angle scenes, but only for some efficiency differences, one situation that fails the algorithm is that the background and objects moves in different directions but in very similar distance.

## Interpolation

The interpolation results are shown in this section with comparisons between different parameter configurations, the unpleased results in the early stages are also expressed as a significant process to obtain plausible new frames.

### Results

**Motion block selection**

In the early stage, after motion vectors are recorded, the new frame is computed by the raw data, which could result in incorrect position of some blocks, as shown in Figure 18 (left), a block containing the trunk of the car model located in the background and cause an obvious error, but the main body of the car is just in the correct position, it can be seen that the most motion blocks obtain the reasonable motion information apart from a few exceptions, which result in errors, from this example I tried to select the motion vectors from frequency, use the motion vectors with largest frequency to represent the entire object, by implementing that, incorrect motion vectors are re – assigned with values from most motion vectors, thus avoid improper motion blocks. A comparison of images after motion vector selection is displayed in the right part of Figure 18.



Figure 19 Different block size comparison

Block size: 8 \* 8

Search range: 40 \* 40

Pyramid level of top image: 4

Pyramid level of bottom image: 3

Pyramid change equals the effects of resizing the block and search range because images are linearly shrunk to a quarter size of one lever higher

**Motion block size**

It is difficult to specify a motion block size, experiments show that large motion blocks could reduce the computation amount and hence increase speed, but some part of the objects could be under – segmentation, to put another way, the foreground could be unexpectedly classified as background. However, smaller blocks make contributions to accuracy, since images are sufficiently divided and motion blocks are moving to the correct position. Figure 19 demonstrates an interpolation at the same point with different block sizes, in this image, block size are actually not changed but the image pyramid is modified as an alternative, both operations are equivalent because the computation varies linearly. In the upper image, the size is larger, so one block containing the part of left back wheel is classified as background and hence kept still. In contrast, when reduce one level in pyramid, the motion blocks equally shrunk to a quarter of previous area, the car is sufficiently divided as foreground. Furthermore, if the pyramid level continues goes up, smaller motion blocks could result in extremely negative effect, firstly the computation workload could shoot up due to such square – magnitude growth and secondly, over – segmentation would be presented, since the difference of block is calculated from grey values, slightly difference of intensity could cause the problem. Figure 20 is an unsuccessful interpolation result obtained in a long time beyond tolerance, hence, a reasonable pyramid configuration becomes significant. With a number of experiments, I found that a reasonable configuration could be 8 \* 8 block size with 40 \* 40 search range, and a pyramid level of 3 and 4, under such conditions, a relative higher probability of obtaining plausible images could be achieved. Comparisons are shown in Figure 21 and Figure 22, same sequence are input to the program, when pyramid level is set to 4, four images are created and three are plausible, when the pyramid level reduced by 1, five images are created and the result is slightly improved.



Figure 20 Different block size comparison



Figure 21 Results from image sequence interpolation

Pyramid level: 4

Except for the last frame, other images could be regarded as plausible, the model car is in the correct position while segmentation is reasonable, interpolation for the 4th image is a bit more difficult, motion blocks are relative large for this context, it is improved in the case of Figure 22.

**Interpolation point**

For an image sequence, the interpolation is applied wherever the motion distance between two frames are within a specified range of the largest one, the range is set to 0.8 in experiments, and various number of images will be created, Figure 22 is an example that five images are created, compared to the results shown in Figure 21, the only difference at the beginning of the program is the level of image pyramid.



Figure 22 Another result

Pyramid level: 3

One more image is created at a new point (first image), the last image is improved, while other three are similar

### Summary

In this program for interpolation, after images sequence is input, process is completed automatically and acceptable results have been achieved, there are several significant elements that could affect the result, first, optimization of motion vectors could avoid individual incorrect motion block movements Second, adjusting motion block size is a trade – off between speed and accuracy, large block sizes gives fast algorithm with less accuracy, the result could be under segmented, while vice versa. Finally, the condition of interpolation could be left to the user, using the largest motion as a measurement, could create images between large motions, which contributes to a detailed story telling when images are projected to films due to an equivalent effect of reducing playing back rate.



Figure 23 Another visual plausible result

## Results Summary

The failure of thresholding – XOR algorithm in this project leads to a decision making for smaller region segmentation, it can be seen that motion estimation could work for both registration and interpolation, a reasonable segmentation is the key factor to get successful results. In this project, the method achieve some decent images but still find some unsolved problems, for example, when background and foreground move to similar distance in two directions, it is fairly difficult for segmentation, and difficulty for interpolation is to find a proper block size.

Some more images that have been tested are shown in the following Figures.



Figure 24 Unpleased results of excessive small block size

Small block size could give more accuracy, however images could be over segmented, some blocks in the background are categorized as foreground.

# Discussion

In this section the results observed are being discussed, with comparison when relevant, it will also give a conclusion to each method whether it is effective for this project, finally, aims and objectives of the project will be reviewed and compared to what actually have been down.

## The Failed Method

### Reason Analysis

There are two methods employed for image registration part, the algorithm from (G. Ward, 2003) was firstly preferred boils down to previous HDR image processing experience, this method applies segmentation by a median threshold, which is the grey value from the intermediate position of the pixel after sorting by grey values, this could eliminate the effect from various intensities in the image. In HDR image construction, the intensity is completely different for three source images (usually use three images to represent details from specular highlight, darkness and intermediate grey parts in the scene), while for the sequence of time – lapse, the intensity is unpredictable, it could change massively when recording a sunrise process, for example, and sometimes it only fluctuates slightly. Hence, theoretically, after such segmentation, the algorithm could be applied for aligning images, however, the result turned out to be frustrated despite of some successful examples, the reason for that is that motions are usually presented in both input images (no motions in HDR source images), and to obtain offsets in x and y directions, two images are computed by an XOR operator, which is reluctant to ensure that the shifting vector is returned from a result containing least non – zero values, and furthermore, when applied into the image sequence, it is difficult to judge whether an image pair is aligned or not.

### Further Thinking

After recognizing the causes of failure, it is necessary to find ways to avoid the influence from motion, in time – lapse practices, motions should actually be neglected when applying registration, thus it should only work with the background parts.

## Motion Estimation

### Comparison with Thresholding – XOR method

Motion estimation provides an algorithm to recognize the background and foreground, which, inherently circumvent the problem from the thresholding – XOR method, background is recognized by motion vectors from motion blocks. The motion information of the background could be used for registration, while that of foreground indicates the object movements.

Each motion Block owns a motion vector predicting the position relationship between two images, among them some vectors are incorrectly estimated since the slightly changes in intensity, these could be neglect because most vectors are in the same motion, which is detected from the motion vectors that are in largest frequency, then the motion vector could represent the position relationship which is returned for shifting images. This is reliable because the background parts in the image are actually moving together when camera is slightly moved, the image is subdivided to motion blocks, comparing with the thresholding – XOR method, it has two main advantages, first, difference computations are operated only in backgrounds, second, since subdivisions are actually input into comparison, the genuine motion of the background is indicated by most motion blocks, as a consequence, the vector with largest proportion is selected, by doing that, larger sample volume is examined so that the probability of correct decision making is enlarged.

### Motion Block Size and Image Pyramid Levels

Motion block size and image pyramid levels are two equivalent factors that could control speed and accuracy in the motion estimation method, excessively small motion blocks (or very low pyramid level) will lost both because of a substantial computation quantity in conjunction with over segmentation, on the opposite, very large motion blocks (or high pyramid level) could result in fast speed but under segmentation, thus it is significant to determine a equilibrium point, from the experiments carried out with HD (1920 \* 1080) level, an acceptable point could be 8 \* 8 block size with 40 \* 40 () search range, image pyramid level is set at 3 (actual input image resolution 240 \* 135).

### Applicability

Motion estimation based method could be effective if an image is successfully segmented, throughout the experiments, for image registration, 50 image pairs are tested and 40 are reasonably aligned, while for interpolation, there are 40 successful results out of 50 input frames, to sum up, roughly 80 per cent image could be processed with this method, including those with clear difference between background and foreground and some wide angle scenes.

There are limitations for this method, firstly, in our experiment, images with similar motions in background and foreground are difficult to be segmented and therefore less likely to obtain plausible registration and interpolation. Secondly, since the method is designed for 2 – D motion field, problems in 3 – D motion field are not included.

## Applied to Image Sequence

The most interesting and exciting point of this project is when the methods are extended to the entire image sequence, for image registration, a key point is to determine a reference image, because it could optimize the processing speed when the reference is the one to which majority images in the sequence are matched, otherwise, if the first one is selected or applying arbitrary selection, there might be massive computation cost. In addition, by implementing the reference detection, the entire image sequence could be automatically registered, which could contribute to time – lapse photographer since it is common to deal with thousands of images.

For the interpolation part, as interpolation is applied whenever a relative large motion is detected, more details are displayed in the film and hence the time become non – linear, it becomes more expressive when the photographer misses some motions when capturing images, similar to registration, in this project the interpolation is also automatic for the purposes of a more illustrative story expression.

In conclusion, an automatic registration method for image sequence has been implemented in this project and hence, it is convenient to process source images from time – lapse photography, it saves the time to find the images that requires alignment. Furthermore, time could be automatically adjusted when large motions are presented between two frames, this could offer more details to the audience.

## Evaluation

The first objective of this to implement registration between two images, and two algorithms have been experimented with, although the thresholding – XOR method is less robust, it made contributions to beginning exploring with the motion estimation method, which could extract the background (segmentation) and shift images as per the motion vectors, in addition, motion estimation could be applied to aligned images and compute with foreground to predict the motion of moving objects between frames, with occlusion handling, plausible new frames could be computed and this obtains the interpolation objective of this project.

Another significant objective is to extend the methods to image sequence and automatically implement registration and interpolation, firstly, vector container in C++ is employed for recording all file names in an specified directory, in conjunction with creating new file names. Then vector objects are working with queue structure to detect reference image for registration, and to determine interpolation point between two frames. In short, this project has developed image processing techniques to fit a large number of images without any manual operation.

For acceleration, image pyramid is an effective and efficient to shrink images to smaller size with a quarter computation amount lessened when reducing a level, specifically, in my program it works with motion block size and search range which are together pre – defined, usually it could work at the found equilibrium combination and for software development consideration it could be selected by the user in case of some exceptions.

With regard to robustness, it is pleased to achieve an approximate 80 per cent successful probability in both registration and interpolation methods, testing with 50 example image pairs respectively.

In conclusion, the key contribution in this project is the implementation of an automatic processing method for image sequence, which offers convenience in time – lapse photography.

## Further Work

It has been introduced that

# References

1. Iain E.Richardson,The H.264 Advanced Video Compression Standard John Wiley & Sons ltd, 978-0-470-51692-8, 2010.
2. I.E.G. Richardson, H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia, John Wiley & Sons, Ltd, 0-470-84837-5, 2003.
3. Aseem A, Maneesh A, Michael C, et al. Photographing Long Scenes with Multi-Viewpoint Panoramas. ACM Transactions on Graphics, 25(3), pp. 853-861, 2006.
4. Loewke K, Camarillo D, Piyawattanametha W, et al. Realtime Image Mosaicing with a Hand-held Dual-axes Confocal Microscope. Progress in Biomedical Optics and Imaging. Proc of SPIE, Endoscopic microscopy III, Proceeding of SPIE, 6851: 68510F1-9, 2008.
5. Heinze N, Esswein M, Krüger W, et al. Automatic Image Exploitation System for Small UAVs. Proc of SPIE on Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications, 6946: 69460G1-10, 2008.
6. Q. Zhang, Z. Zhang, D. Zeng, A fast, Automatic and Robust Image Registration Algorithm, International Conference on Virtual Reality and Visualization, 2011.
7. F. L. Seixas, L. S. Ochi, A. Conci, D. C. M. Saade, Image Registration Using Genetic Algorithms, Proceedings of the 10th annual conference on Genetic and evolutionary computation, 2008.
8. G. Ward. Fast, Robust Image Registration for Compositing High Dynamic Range Photographs From Hand-held Exposures. Journal of Graphics Tools: JGT, 2003.
9. D. Mahajan, F. Huang, W. Matusik, R. Ramamoorthi, P. Belhumeur Moving gradients: a path-based method for plausible image interpolation, ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2009 TOG Homepage, Volume 28 Issue 3, 2009.
10. T. Stich, C. Linz, C. Wallraven, D. Cunningham, M. Magnor, Perception-motivated Interpolation of Image Sequences, ACM Transactions on Applied Perception: TAP, 2011.
11. C. L. ZITNICK, N. JOJIC, S. B. KANG, Consistent segmentation for optical flow estimation, Proc. IEEE Int. Conf. Computer Vision, pp. 1308–1315, 2005.
12. V. KOLMOGOROV, R. ZABIH, Computing visual correspondence with occlusions using graph cuts, Proc, IEEE Int. Conf. Computer Vision, pp. 508–515, 2001.
13. J. SUN, Y. LI, S. KANG, H. SHUM, Symmetric stereo matching for occlusion handling, Proc. IEEE Conf. Computer Vision and Pattern Recognition 2, pp. 399–406, 2005.
14. A. W. FITZGIBBON, Y. WEXLER, A. ZISSERMAN, Image-based rendering using image-based priors, Proc. IEEE Int. Conf. Computer Vision, pp. 1176–1183, 2003.
15. H. WANG, R. RASKAR, N. AHUJA, Seamless video editing, Int. Conf. on Pattern Recognition, vol. 3, pp. 858–861, 2004.
16. J. JIA, C. TANG, Image stitching using structure deformation, IEEE Trans. Pattern Analysis and Machine Intelligence 30, 4, pp. 617–631, 2008.
17. M. Sonka, V. Hlavac, R. Boyle, Image processing, analysis, and machine vision, second edition, Brooks/Cole Publishing Company, 1999.

# Appendix

1. This equation is from the OpenCV function that are used for color space transformation. [↑](#footnote-ref-1)