Department of Mathematics

Chair of Mathematical Modeling of Biological Systems

Technische Universität München

Master's Thesis in Bioinformatics

Single-cell analysis of cancer drug response using computer vision and learning algorithms on time-lapse microtrench data

Pandu Raharja

Department of Mathematics

Chair of Mathematical Modeling of Biological Systems

Technische Universität München

Master's Thesis in Bioinformatics

Single-cell analysis of cancer drug response using computer vision and learning algorithms on time-lapse microtrench data

Wirkungsanalyse von Krebsmedikamenten in Einzeller Auflösung durch die Anwendung von Computer-Vision- und Machine-Learning-Algorithmen auf Microtrench-Videoaufnahme

Author: Pandu Raharja

Supervisor: Prof. Dr. Fabian Theis, Dr. Carsten Marr

Advisor: Prof. Dr. Fabian Theis

Prof. Dr. Dmitrij Frishman

Submitted: 15.10.2017

Abstract

Quantitative measurement of cancer drug response is esential to objectively gauge the efficacy of cancer drugs. So far, there has been no method to track and quantitatively measure single-cell response of of cancer drug treatment. A novel pipeline is presented in this thesis. First, a quasi-high-throughput method to track cells and quantitatively analyze single-cell response to drugs. We investigate the response of model cancer cell lineagues, MOLM and Jurkat, to known anti-cancer drugs Vincristine and Doxorubicine. While the method enabled relatively easy and quasi-high-throughput analysis of cancer treatment *in vitro*, our pipeline could also be adapted in varios contexts involving single-cell analysis with reasonable amount of modifications necessary.

Contents

Chapter 1

Introduction

Cancer is among the deadliest diseases ever known to human being. It is a leading cause of death in 2009, second only to cardiovascular diseases [1]. The numbers are discontenting, especially in the developed world. In the United States alone, half of men and a third of women are expected to develop some kind of cancer. According to US government, in 2016 alone, an estimated 1,685,210 people will be diagnosed with cancer, while 595,690 more will be die from it [2].

Worldwide, the International Agency for Research on Cancer's GLOBOCAN series report that, in 2014, [3] .

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

```
I.
- readable to people without background in the fields
- non technical at all
II.
- what have the researches done
-- biologics
-- technicals
III.
thesis overview
4~8 pages
```

1.1 The structure of the thesis

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet,

tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Chapter 2

Data and Methods

This chapter considers two aspects of the project: the experiment setting and the data analysis pipeline. In the first half of the chapter, the cell environment is brought forth. The highlight of this experiment, the microfluidics setup for cell containment, is elucidated in this part. Some biomedical and biochemistry aspects of the experiment are also mentioned. This includes the drugs and the auxiliary chemicals of interest used in the experiment and the cell lines probed for the experiment. Later on, the hypothesis underlying the experiment is presented.

The second part deals mostly with the quantitative methods and algorithms used to process data into meaningful observations. The part is opened with definitions used in the methods section. Afterwards, each method developed/used in the pipeline is brought forward. For each method, the rationale explaining the reason of using the method is also accompanied in the subsection.

2.1 Experimental Setting and Data

2.1.1 Microfluidics environment

Microfluidic methods for analysis and manipulation of biological cells have been done in various way and form. The miniscule spatial setting means that the family of methods is compatible for single-cell resolution analysis. For example, optical-based microfluidic methods have been used for example to sort cells with very high accuracy [4, 5, 6].

For the duration of the experiment, the cells are contained inside the so-called microtrench assay. There are several mentions of microtrench in the past from both our partner lab and several other groups(TODO cite, cite, cite). In our case, the design is based on previous work by the biophysics group of LMU München, which are also the experimentalists in this project.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur

auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2.2 Definitions

Mathematical definitions used in the thesis

2.3 Methods

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2.3.1 Laplacian of Gaussian (LoG) Cell Recognition

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2.3.2 Image Encoding

Consider whether image encoding shoud contain Lenna's picture instead.

2.3.3 Shift Correction

Consider following picture:

There are numerous encodings that could be used to internally represent this picture. Many such encodings derived from the so-called Red-Green-Blue (RGB) encodings. RGB encoding represents the pixel as a combination of red, green and blue color. This encoding is able to represent various spectra of human visible color and useful enough for most use cases (citation). To give representation on how the encoding works, the RGB encoding of some part of Figure **??** is shown in Figure 2.2. For an image of size $m \times n$ pixels, the

Figure 2.1: Lenna

RGB encoding is thus a 3-dimensional matrix of dimension $m \times n \times 3$. For time-lapsed images accordingly, the RGB encoding of a video of length T is a 5-dimensional matrix of shape $T \times m \times n \times 3$. TODO: put graphical explanation of data here.

Now, consider a case in which images are shifted in a time-lapsed movie. TODO: explain mechanism. No rotation of camera is assumed, hence there are only two degree of freedoms (vertical and horizontal). Thus, a shift can be defined as a vector movement $\vec{v}$ of all points $x_{i,j} \in M_{t_i}$ in the time-lapse from time t_i to t_{i+1}. Given two degrees of freedom and discreteness of the problem due to pixel representation, the task is reduced to finding difference in x- and y-axis (δ_x and δ_y), so that the difference of transformed pixels at t_i and t_{i+1} are minimized, i.e.:

$$argmin_{\delta_x,\delta_y}\{d(M_{t_i}, M_{t_{i+1}}^{\delta_x,\delta_y} + (\delta_x, \delta_y)^T)\}$$

Where $M_{t_{i+1}}^{\delta_x,\delta_y}$ is the entries of matrix $M_{t_{i+1}}$ after applying the shift $\vec{v} := (\delta_x, \delta_y)^T$, i.e.

$$M_{t_{i+1},x,y}^{\delta_x,\delta_y} = M_{t_{i+1},x-\delta_x,y-\delta_y}$$

For the distance function d, the in all channels absolute difference function is used, which is defined as:

$$d(M_i, M_j) = \sum_{c \in \{R,B,G\}} \sum_x \sum_y |M_{i,c,x,y} - M_{j,c,x,y}|$$

Since some pixels are lost from the field of view during a shift, only a subset of subsequent pictures is used to determine the shift, preferably those around the center point. This will allow the largest search space possible, since the distance to all four margins of the picture is maximized at the center point. The search for the optimal (δ_x, δ_y) pair is implemented as a grid search along the x- and y-axis. An example of the search grid is shown in Figure 2.3.

Since the time-lapsed data consists mainly of grayscale image, the RGB encoding could be the directly transformed to grayscale encoding. Using the transformed method also speeds up the calculation process since the distance function only computes the difference of grayscale channel's values:

$$d(M, N) = \sum_x \sum_y |M^{gray}_{c,x,y} - N^{gray}_{x,y}|$$

Due to lost pixels around the margin of before and after pictures, only the overlapping part of both slides are included after the correction. Thus, for an inferred shift of (δ_x, δ_y), the new dimension of the pictures is then $(m - \delta_x) \times (n - \delta_y)$. This change would then propagation to the other time-lapse images to maintain consistency of the images.

Ideally, the shift correction should be done for each position to reduce the track dropout rate caused by image shifts. This is however computationally very expensive and, as seen in Figure 2.4, not really necessary since the biggest shift indeed only happens right before and after the treatment, as it was expected during the experiment setting. As seen in Chapter XX (TODO: quote), the tracking allows certain amount of tolerance. In this regard, the other frame shifts are way within the tolerance of our tracking algorithm. As shown in Figure YY (TODO: add droput rate), the dropouts caused by frame shifts in the other time points are basically noisy dropout caused by random noise in time-lapse movie being tracked as cells (TODO: add reference about this phenomena).

The algorithm for shift inference could be seen in Appendix A.

2.3.4 Cell death signals

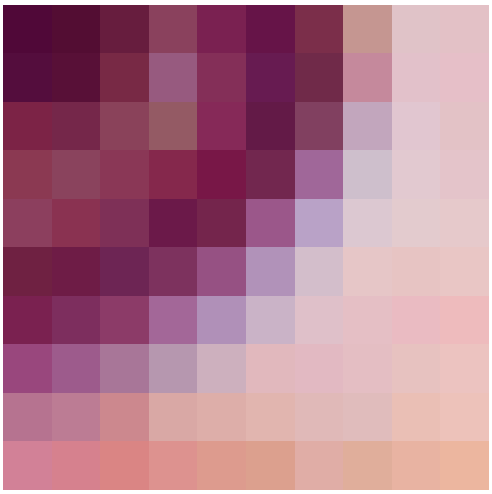
Measuring cell death is a crucial part of the experiment, as the reliable determination of it is the basis of most analysis in this thesis. There are several way to measure cell deaths with varying complexity and accuracy. Each method contains certain assumptions of cell death.

For example, determining cell death by cell movement assumes death of a cell if no movement beyond random flux is observed in certain amount of time. This obviously has certain drawbacks, such as when the observation is done in non-static environment. Moreover, defining the limit of the random flux, above which a given cell is assumed to actively move, is not a trivial task. Some kind of gold standard for a given cell line and environment has to be established manualy, which is very time consuming. This fact is again made even more complicated by the fact that many cells show different movement pattern upon introduction of treatment (TODO citation). It is well known that some cells tend to move faster or slower under stress (TODO citation), the situation many cancerous cells in our experiment will experience upon addition of cancer drugs treatment (TODO citation).

The second method is using cell size. During apopotosis, the cells would shrink. Given It is known that cell size

In this experiment, two cell signals are used as indicator of cell death.

(a) 1a



(b) 1b

```
[[255 255 255 255 255 255 255 255 255 255]
 [  0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0]]
[[255  88 120 151 132 103 112 197 226 230]
 [  0  16  41  90  47  27  42 137 193 191]
 [  0  55  69 127  88  81  73 156 202 200]]
[[255 117 138 148 134  99 129 194 225 227]
 [  0  39  66  90  41  26  64 166 198 194]
 [  0  74  90 100  88  71  96 189 208 198]]
[[255 138 138 133 120 114 160 206 226 228]
 [  0  67  55  40  23  39 103 191 201 196]
 [  0  93  86  76  71  79 153 204 208 202]]
[[255 138 126 107 116 155 185 220 227 230]
 [  0  50  48  25  37  87 162 200 203 201]
 [  0  81  87  73  76 138 199 209 206 203]]
[[255 110 109 125 150 177 211 230 231 233]
 [  0  28  37  50  81 146 190 198 196 198]
 [  0  70  84  94 131 185 203 199 195 197]]
[[255 125 140 163 176 202 223 229 234 238]
 [  0  46  59 103 144 179 192 191 187 187]
 [  0  94 104 152 184 199 201 197 194 189]]
[[255 157 168 182 205 225 226 228 231 236]
 [  0  91 118 151 176 184 185 190 194 195]
 [  0 140 152 175 190 189 194 195 192 192]]
[[255 188 204 217 221 225 224 224 234 237]
 [  0 124 136 168 174 181 185 188 191 194]
 [  0 148 142 165 169 175 184 188 181 186]]
[[255 214 218 221 221 220 224 224 232 236]
 [  0 129 133 146 155 160 173 174 179 182]
 [  0 142 132 143 142 142 166 155 162 159]]
[[255 255 255 255 255 255 255 255 255 255]
 [  0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0]]
```

(c) 1b

Figure 2.2: Figure 2.2a shows the content of red marked region in Figure **??**. Figure2.2b shows the zoomed part around Lena's right eye and matrix represented in Figure 2.2c shows the RGB representation of the eye.

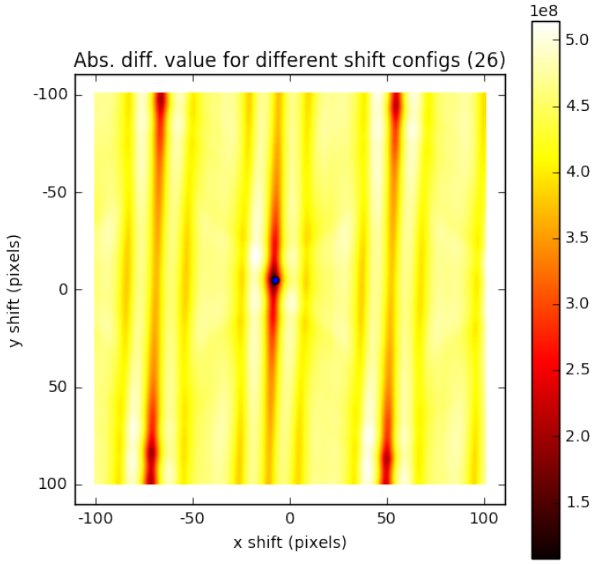


Figure 2.3: Search grid shift for Position 26. The search was conducted for shift between the last time point before and the first time point after the drugs treatment. The minimum is marked with thick black dot, which is returned after every grid-search call as inferred shift. In the position, the shift was inferred to be 8 pixels upwards and 5 pixels leftwards. Notice the repeating pattern of relatively favorable configurations after approximately 50 horizontal and 100 vertical pixels caused by lattice nature of the slits.

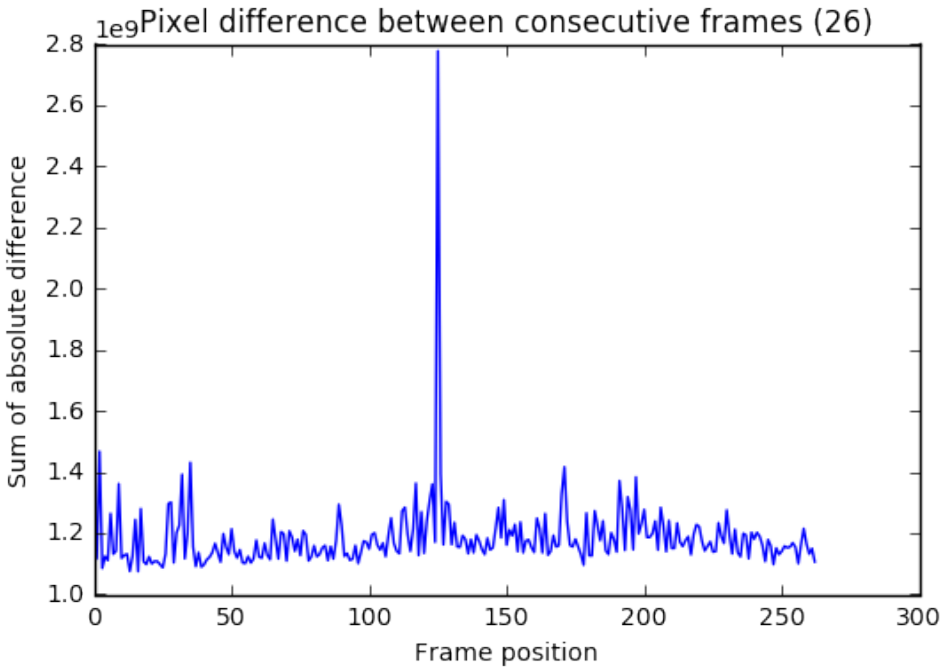


Figure 2.4: Pixel difference between consecutive frames at Position 26. In most cases, the pixel difference between the frames is mainly caused by moving cells. The diference during the treatment, on the other hand, is caused by physical shift of the frame. While moving cells mostly caused minimum noise-like pixel difference, the physical shift of field of view distorts the physical alignment and evokes immense pixel difference.

Chapter 3

Results

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

I. Pipeline

II. Quantitative Analysis

Chapter 4

Summary and Outlook

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

```
I. Summary (and discussion?)
Connect Results with Background

II. Outlook
Improvemnt capability
```

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Bibliography

[1] Akulapalli Sudhakar. History of cancer, ancient and modern treatment methods. *Journal of cancer science & therapy*, 1(2):1, 2009.

[2] The Centers for Disease Control National Cancer Institute and the North American Association of Central Cancer Registries Prevention, American Cancer Society. Cancer Statistics. `https://www.cancer.gov/about-cancer/understanding/statistics`, 2017. [Online; accessed 07-November-2017].

[3] Jacques Ferlay, Isabelle Soerjomataram, Rajesh Dikshit, Sultan Eser, Colin Mathers, Marise Rebelo, Donald Maxwell Parkin, David Forman, and Freddie Bray. Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. *International journal of cancer*, 136(5), 2015.

[4] MP MacDonald, GC Spalding, and Kishan Dholakia. Microfluidic sorting in an optical lattice. *Nature*, 426(6965):421–424, 2003.

[5] Mark M Wang, Eugene Tu, Daniel E Raymond, Joon Mo Yang, Haichuan Zhang, Norbert Hagen, Bob Dees, Elinore M Mercer, Anita H Forster, Ilona Kariv, et al. Microfluidic sorting of mammalian cells by optical force switching. *Nature biotechnology*, 23(1):83–87, 2005.

[6] Jean-Christophe Baret, Oliver J Miller, Valerie Taly, Michaël Ryckelynck, Abdeslam El-Harrak, Lucas Frenz, Christian Rick, Michael L Samuels, J Brian Hutchison, Jeremy J Agresti, et al. Fluorescence-activated droplet sorting (fads): efficient microfluidic cell sorting based on enzymatic activity. *Lab on a Chip*, 9(13):1850–1858, 2009.

Appendices

Appendix A

Algorithms

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Shift Inference

```
def infer_shift(last_slide, first_slide, search_space=(200, 200)):

  if (search_space[0] % 2 != 0) or (search_space[1] % 2 != 0):
    print("Search spaces have to be even!")
    return None
  else:

    ## calculate absolute difference for various shifts
    x1 = search_space[0]
    x2 = search_space[0]
    y1 = search_space[1]
    y2 = search_space[1]
    mid = f1.shape[0] // 2, f1.shape[1] // 2

    ## results storage
    absdiffs = np.zeros((search_space[0] + 1, search_space[1] + 1))

    ## last slide before treatment
    f1sub = f1[(mid[0] - x1):(mid[0] + x2), (mid[1] - y1):(mid[1] + y2)]

    ## search space
```

```
xdiff1 = -int(search_space[0] / 2)
xdiff2 = int(search_space[0] / 2) + 1
ydiff1 = -int(search_space[1] / 2)
ydiff2 = int(search_space[1] / 2) + 1

for xdiff in range(xdiff1, xdiff2):
  for ydiff in range(ydiff1, ydiff2):

    ## calculate absolute difference for shift
    f2sub = f2[(mid[0] - x1 + xdiff):(mid[0] + x2 + xdiff),
               (mid[1] - y1 + ydiff):(mid[1] + y2 + ydiff)]
    x = xdiff + int(search_space[0] / 2)
    y = ydiff + int(search_space[1] / 2)
    absdiff_xy = np.sum(cv2.absdiff(f1sub, f2sub).ravel())
    absdiffs[x][y] = absdiff_xy


## calculate shift based on calibration data
x = np.argmin(absdiffs) // absdiffs.shape[0]
y = np.argmin(absdiffs) % absdiffs.shape[0]

"""
True shift is the opposite of coordinate encoded
in absdiff

Let X2 the second picture and X1 the first picture.
If the sub-picture of first slide of X2 centered
at (c1 + s1, c2 + s2) fits the most with the sub-picture
of the last slide of X1 centered at (c1, c2)
then the pictures shift by (-s1, -s2) upon treatment
"""
diff = -(x - search_space[0] / 2), -(y - search_space[1] / 2)

return diff
```