DETERMINE PRODUCTIVITY MEASURES USING BUSINESS PROCESS IMAGES

BY
MR. J.M.C. VAN STADEN
29013799

Supervisor: Ms. E. van Wyk

A project submitted in partial fulfilment of the requirements for the degree of

BACHELORS IN INDUSTRIAL ENGINEERING

at the

FACULTY OF ENGINEERING, BUILT ENVIRONMENT

AND INFORMATION TECHNOLOGY

UNIVERSITY OF PRETORIA

November 2013

EXECUTIVE SUMMARY

Time studies have long been one of the only means of obtaining data for use of analysing a business process. However useful this tool has been in the past, using only time draining, primitive stop-watch based time studies have seemed to run its course and it has surely become necessary to make use of available technology to further analysis of business productivity measures.

Image analysis has been used in the past to not only predict weather patterns and position of stars and planets, but also with success in the fields of bio-informatics to analyse genes. If it is possible to analyse small entities like genes, as well as planet bodies millions of kilometres away, why not try and use some of these same techniques in business?

This document provides a basic outline of how an image can be described, as well as examples of techniques that can be used to divert an image to either mathematical or graphical metrics. A basic timeline and project plan will also accompany the document to give the reader a sense of when to expect which deliverables.

Accompanying this project plan will be explanations of how image analysis is used in other disciplines. The aim of this project is to see whether it is possible and plausible to measure various productivity levels using image analysis and afterwards using the obtained data as input to successfully calculate performance indicators important in the industry of managing productivity such as throughput and inventory levels.

The research into the application of using business process images to determine productivity measures will be proven to have a valid place in the industry as a practical and beneficial methodology, using available technology to improve not only the way that business is done, but allowing for radical change in the small to medium business sector, giving more business owners the advantage of using methods that are not only less tedious, but can be shared with the industry to benefit all.

As a whole, this dissertation provides the platform from which not only further research can be done to illustrate the necessity of an alternative technique to determine productivity measures, but gives the reader insight into software that may be applicable in developing the above-mentioned technique of determining productivity measures by using business process images.

ACKNOWLEDGEMENTS

Many times when a project comes together, I tend to settle down and congratulate myself on my accomplishments by having a nice cold and gold beverage. In the case of this dissertation, however, I have never thought of it as an individual effort. I may have done much of the research, written the document and spent hours editing, but none of this would have been at all possible if it was not for certain contributions made along the way.

My beautiful wife, Linda, thank you for your constant support, not only with this final chapter of university, but also throughout the last five years while I was yet again a student, trying to catch up with how wonderful you have always been. I love you completely, and know that I dedicate every page and minute I spent on this document to you.

My wonderful family. Harry and Linda, for all the encouragement and times shared, picking me up when I was down, and for never giving up on me when I already did. Thank you so much for the prayers and support. Riaan, my Boet, thank you for the calls from Sweden, the love you have always given me, even when I least deserved it. I love you all dearly.

I have also been blessed with amazing in-laws. Thank you so much Ferdie and Antoinette, for making sure I always knew that I was fit to be an engineer. Your love for me means the world. A special word of thanks to Ma Antoinette for every call she made, every cent she sponsored and every prayer that was said to help get me through university.

I would also like to thank Estelle van Wyk, my supervisor for this dissertation. Thank you for all the support, the advice and friendship throughout the year. Your guidance was enormously beneficial, and is much appreciated.

Finally, my friends. I will not state names for the sake of anonymity; I know many of you wished at times that I would call someone else to share my 56 herbs and spices... As an aspiring engineer though, I think X would be applicable, and feel free to enter your name to solve for X. Thank you X. Thank you for the last five years and all the support you have given even though you did not even know it. Thank you for always being available to share both the good and the bad times, and may there be many more opportunities for all of us to keep on doing so in the future.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

1.1. INTRODUCTION

Productivity has always been defined as the average efficiency of a process that has certain inputs and outputs that are related to aforementioned process. Above mentioned efficiency, in turn, is linked directly to the time that elapses as that process progresses. Therefor it has always been imperative that proper time studies be made that can produce data for calculations of how productive a system can be. Figure 1 is a basic list of the metrics that is used to interpret productivity.

Figure 1: Table of Productivity Metrics.

Utilization
Throughput
Queue Time (hr)
Cycle Time (hr)
Cumulative Cycle Time (hr)
WIP in Queue (jobs)
WIP (jobs)
Cumulative WIP (jobs)

These are the metrics that are normally calculated with the data that is derived from time studies. Normally time studies have been done with physical people that measure all the needed inputs either by hand (stopwatches) or by using more technology advanced methods like RFID or sensors. It seems however, that there is no middle ground that allows for ease of use, while not breaking the bank. In many businesses, be it consumer driven or external customer driven, there seems to be a niche wherein a method can be derived to fill this “gap”. This method relies on the probability of successfully analysing images to attain input data as showcased in Figure 2.

Figure 2: Table of Input Variables.

Arrival Rate (parts/hour)
Arrival Coefficient of Variation (CV)
Natural Process Time (hr)
Natural Process Squared Coefficient of Variation (SCV)
Number of Machines
Mean Time To Failures (hr)
Mean Time To Repair (hr)
Batch Size
Setup Time (hr)
Setup Time SCV
Yield

Undoubtedly the question arises of how this may even be possible. The answer lies in the fact that the way that humans perceive images is more of an artistic impression, whereas certain computer software can transform these images into mathematical graphs that in most cases may be as unique as a fingerprint. Figure 3 is what image analysts refer to as a "cartoon", which in general is a two dimensional illustrated image. This image is stored as a 200 × 200 matrix of integers {1, 2, 3, 4} and was used for the quantitative assessment of the classification of Synthetic Aperture Radar (SAR) data, (Mejail, et al., 2003). Frames were used to distinguish the images from the background. The formal definition of an image can be found in APPENDIX A: DEFINITIONS. In general, however, "images whose properties are perfectly known are referred to as "phantoms". They are useful for assessing the properties of filters, segmentation, and classification algorithms, since the results can be contrasted to the ideal output", (Frery & Perciano, 2013).

Figure 3: Original "Cartoon".

Figure 4 and Figure 5 are examples of how images can be represented mathematically as graphs.

Figure 4: (a) Beta Densities, (b) Phantom under the Beta Model Four Beta Densities and Observed Image.

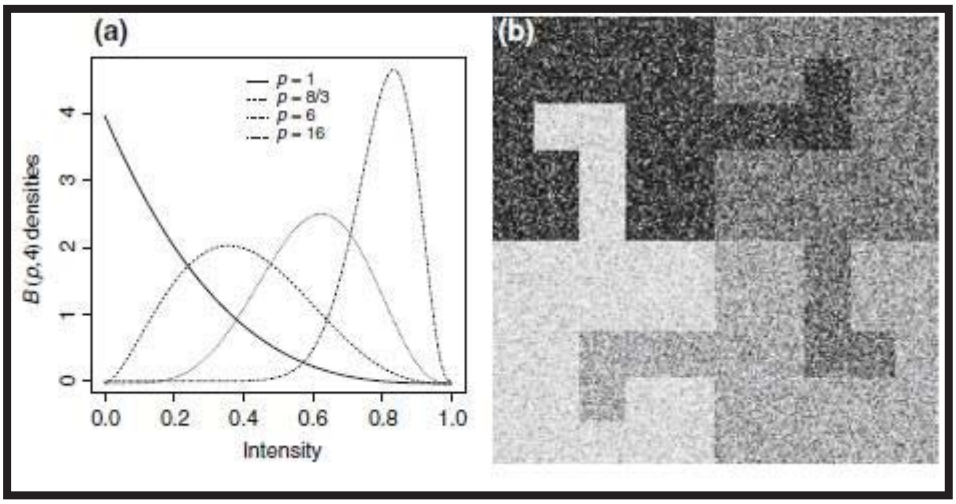
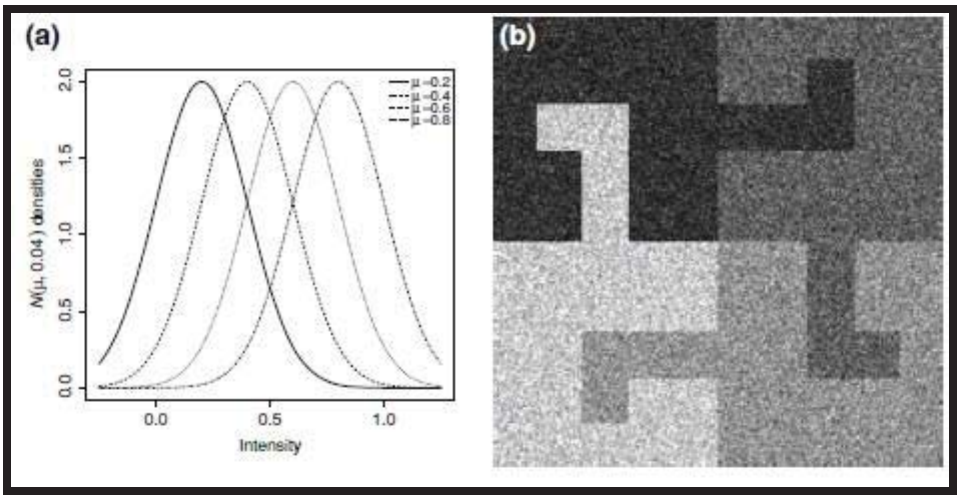


Figure 5: (a) Gaussian Densities, (b) Phantom under the Gaussian Model Four Gaussian Densities and Observed Image under the Gaussian Model.



These mathematical and graphical representations will be the foundation on which analysis of the images will be built on to eventually compare images and attain the input data necessary to calculate the metrics of productivity.

1.2. PROBLEM STATEMENT

The problem that can be formulated out of the above mentioned introduction is that there is currently various old, albeit tried and tested, and out dated methods of doing time studies, as well as very modern, expensive technology that can be implemented, and herein lies the problem, there is currently no middle ground that can replace older methods, without going to the other extreme of modern technology. A transition is needed to use as an interim, where data can still be processed and analysed, while in the meantime saving for that special modern technology.

1.3. PROJECT AIM

The aim of this project is to devise a method of processing and analysing images, preferably still images taken with photographs, to provide input metrics for use in calculating productivity measures. It is a method not as yet used in the industry, and may be implemented much easier and with fewer expenses than other modern methods such as radio technology and sensors.

1.4. PROJECT SCOPE AND DELIVERABLES

The scope for this project is currently limited to research because of the limited resources that have been available and the fact that the idea for analysing business processes by using images is currently not in use at all. The scope may expand to building a model once the proof of the theory of using above mentioned method can be produced, as well as expanding into implementation in a real world environment.

The initial deliverables will consist of documentation relating to the use of various methods of image analysis, as well as literature relating to software programming using the open source software “R”. As a solution a statistical method will be proposed that will best suit future development of the software, and grant assistance in identifying the possibilities that remain in implementation.

As the suggested solution is as yet fairly unused in the area of business, the only plausible way of proving its applicability is to run it through various analysis methods to also prove that it is strategically sound. The deliverables will thus include results of a Porter Analysis, a PESTLE analysis as well as some more qualitative and quantitative techniques to analyse in which area use will be most applicable. Various modelling techniques will also be considered in order to suggest the path to take in terms of development.

1.5. PRELIMINARY LITERATURE REVIEW

Following extensive time researching miscellaneous sources, and gathering data from individuals who work in the field of Industrial Engineering, it has been found that there is sadly not much of use in the literature currently available for determining productivity measures via image processing and analysis. There are however, several other fields where these studies have not only been used to great effect, but have led to breakthroughs in technology and use of modern techniques. The following two abstracts are examples of some of the preliminary research that was done.

1.5.1. Astrophysics

"Optical astronomers normally collect digital images by means of charge-coupled device detectors, which are blurred by atmospheric motion and distorted by physical noise in the detection process. They examine Bayesian procedures to clean such images using explicit models from spatial statistics for the underlying structure, and compare these methods with those based on maximum entropy." (Molina & Ripley, 1993)

1.5.2. Bio-Informatics

"Biological imaging is now a quantitative technique for probing cellular structure and dynamics and is increasingly used for cell-based screens. However, the bioinformatics tools required for hypothesis-driven analysis of digital images are still immature. We are developing the Open Microscopy Environment (OME) as an informatics solution for the storage and analysis of optical microscope image data. OME aims to automate image analysis, modelling, and mining of large sets of images and specifies a flexible data model, a relational database, and an XML-encoded file standard that is usable by potentially any software tool. With this design, OME provides a first step toward biological image informatics." (Swedlow, et al., 2003)

In the above abstracts it can clearly be seen that whether in the universe or within micro-cells, image processing and analysis is an amazing method to be used in various disciplines.

1.6. RESEARCH DESIGN

1.6.1. Proposed Solution

The proposed solution is currently to make use of an open source software package known as "R", as well as an applicable statistical model, and theoretically develop a solution to analyse still images taken by camera. Within these images one intends to find "markers" that will be useful in tracking either people within shops to see how they work and how productive they are, as well as trace inventory levels as a specified time progresses.

1.6.2. Outcome

The possible outcome of this project will be a methodology to develop a software model that an image, or set of images can be given as input to derive the input markers as stated in the introduction, and then to use these inputs to perform calculations to analyse and increase productivity. In conjunction with this methodology an in depth analysis on behalf of the potential benefits and possible use of this solution will be the final deliverable.

1.7. RESEARCH METHODOLOGY

1.7.1. Literature Review

As stated in the PRELIMINARY LITERATURE REVIEW, there are not any studies done in the field of using image analysis to perform time studies. There is however multi-disciplinary approaches on using image processing and analysis that provide insight into the effectiveness and opportunities of using these techniques. The brunt of the literature that will be used will remain in the fields of astrophysics and bio-informatics, as well as literature dedicated to explaining the software and methods that will be used to analyse images.

1.7.2. Data Gathering

The data itself will be gathered by analysing images that will be taken in different scenarios and that utilize different variables. The input data that is needed for the calculations are mostly time oriented and therefor will be of either customers arriving at different times in a shop, the time they spend and the movement associated with the time spent in shops, and whether it is possible to track these customers. Another prospect is to analyse the probability of tracking stock levels during the day with the same images to eventually succeed in forecasting that may be real-time.

1.7.3. Model Formulation

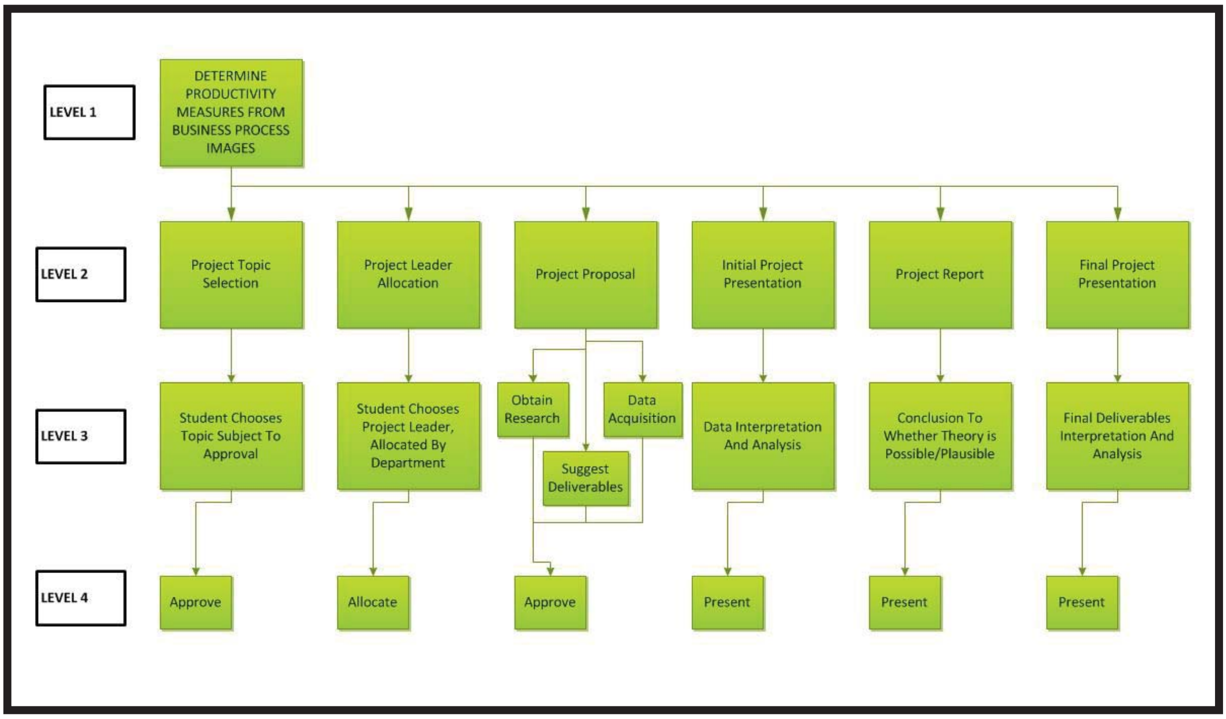
The initial attempts at modelling will be in two dimensions, as this will give an indication of the possibility of using other forms of images. After the initial testing phase is completed, a three dimensional model will be used in the form of a diorama, to give an indication if the theory can be applied to real world.

1.7.4. Model Solution

The solutions will be rendered using open source software named “R”. This software will be used to analyse the still images taken and to revert the physical data from the images to either mathematical of graphical data to be used in further analysis of the process and to determine specific productivity measures.

1.8. PROJECT PLAN

Figure 6: Work Breakdown Structure.



1.8.1. Activities and Tasks

As can be seen in Figure 6, there are several sub-projects within the main project, each with associated tasks and work activities. As this document is the deliverable for sub-project Project Proposal, only the sub-projects following it will be described.

- Initial Project Presentation (Runs in tandem with Preliminary Project Report)
    - Research software (“R”) methods and techniques to be used.
    - Build model and analyse alternatives.
    - Finalise scope and deliverables.
    - Prepare and submit Preliminary Project Report.
    - Prepare and present Oral Examination Presentation.
- Project Report (Runs in tandem with First Draft of Report)
    - Implement solutions.
    - Prove theory.
    - Finalise model to be used.
    - Decide on data to be presented.
- Final Project Presentation
    - Finalise all data to be presented.
    - Interpret final data.
    - Prepare and present Final Project Presentation.

1.8.2. Resources

Execution of this project relies on the following resources:

- Internet access
    - Research
    - Correspondence
- Fuel
    - Meetings
    - Future implementation
- Library

As with all final year projects, the resource responsible for the brunt of the work will be the engineering student, and is responsible to meet all specifications and present all deliverables. The student is also responsible to keep all relevant parties in the communication loop. As this is a final year research project, it seems unnecessary to spend copious amounts of time to prepare a budget that will not be adhered to. The constraints in this project are therefor only time and availability of data and research.

The deliverables, activities and resources within this document are purely estimates, and will only be finalised at a later stage. This Work Breakdown Structure links with APPENDIX B: GANTT CHART, and therefor has no dates in the corresponding blocks.

1.9. CHAPTER SUMMARY

In conclusion it can be seen that although the research that is required seems overwhelming, and the facts are not readily available, methods of obtaining the required data is available, albeit in disciplines that are different from engineering. By working together as multi-disciplined groups it seems that any output can be obtained, and the theory of whether it is possible to determine productivity measures from business process images, may be proven successfully.

CHAPTER 2

2.1. INTRODUCTION

Before designing and implementing a major project, it is utterly important to consolidate the data that has been gathered with the thorough research that was done. Because of this dissertation being a research project in nature, the data that will be presented will be mostly literature. The brunt of this chapter will consist of an extension of the PRELIMINARY LITERATURE REVIEW as summarised in CHAPTER 1 as well as the literature that was researched in the time that has elapsed since the Project Proposal. The goal of this chapter is to set up the research to prepare a conceptual design that will be presented in CHAPTER 3

At the initial stages of the literature review the only available data was that of application of image analysis within the areas of astrophysics and bio-informatics. Since then more information has surfaced on the application of image analysis within business environments. For this part of the dissertation, more time will be spent on explaining the use of the software "R", which will be used to process the images, as well as the application of this software within the fields previously discussed in the PRELIMINARY LITERATURE REVIEW. This will give the reader a much better idea of the capabilities of the software package "R". Many ideologies pertaining to the physical aspects of the imaging regarding the topic of the dissertation will be discussed to give way to the processes and methodology that will be applicable within the initial conceptual design of image analysis of business processes to determine productivity measures.

2.2. LITERATURE REVIEW

2.2.1. "R"

To define the software package "R", the following piece was taken from the introductory document about the software (Venables, et al., 2013):

"'R' is an integrated suite of software facilities for data manipulation, calculation and graphical display. Among other things it has:

- an effective data handling and storage facility,
- a suite of operators for calculations on arrays, in particular matrices,
- a large, coherent, integrated collection of intermediate tools for data analysis,
- graphical facilities for data analysis and display either directly at the computer or on hardcopy, and
- a well-developed, simple and effective programming language (called 'S') which includes conditionals, loops, user defined recursive functions and input and output facilities. (Indeed most of the system supplied functions are themselves written in the 'S' language.)

The term "environment" is intended to characterize it as a fully planned and coherent system, rather than an incremental accretion of very specific and inflexible tools, as is frequently the case with other data analysis software.

'R' is very much a vehicle for newly developing methods of interactive data analysis. It has developed rapidly, and has been extended by a large collection of packages. However, most programs written in 'R' are essentially ephemeral, written for a single piece of data analysis."

'R' as a language and environment was mainly developed for statistical calculating and graphics. Because of the "open sourced" nature of the package however, the environment has expanded as many more applications within the environment was developed by users outside of the initial software development team and has since become popular for use of image processing and analysis.

It was pointed out in the PRELIMINARY LITERATURE REVIEW that image analysis could be used in the fields of Astrophysics and Bio-Informatics. The following literature will shed some light on how "R" is utilised within these fields.

2.2.2. "R" and Astrophysics

Within the field of astrophysics "R" is mainly used as a statistical tool, with many more applications in actual data processing to produce graphs to analyse these data sets than actual image processing, although the application is more than possible.

The following are examples of the statistical analysis that is indeed possible with "R" within an astrophysical context. Each is accompanied by graphs that can be plotted by using each technique, but no explanations will accompany these as it is unfortunately not within context of this dissertation, and is merely present to illustrate the graphical capabilities of "R" (Feigelson & Babu, 2012):

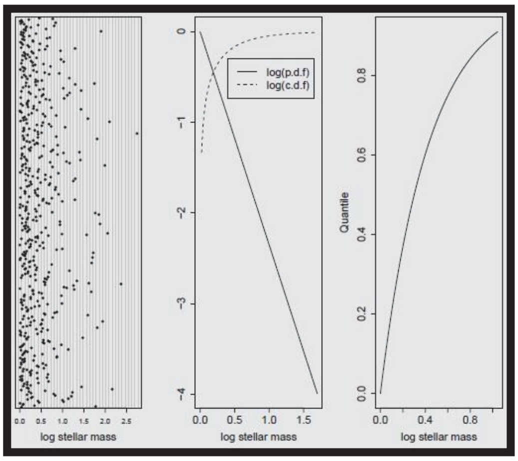
Statistical Inference

The application of "R" for statistical inference is unfortunately not discussed in much detail as it is used in many of the other applications, for example in the processing of the data used for Figure 7.

Probability Distribution Functions

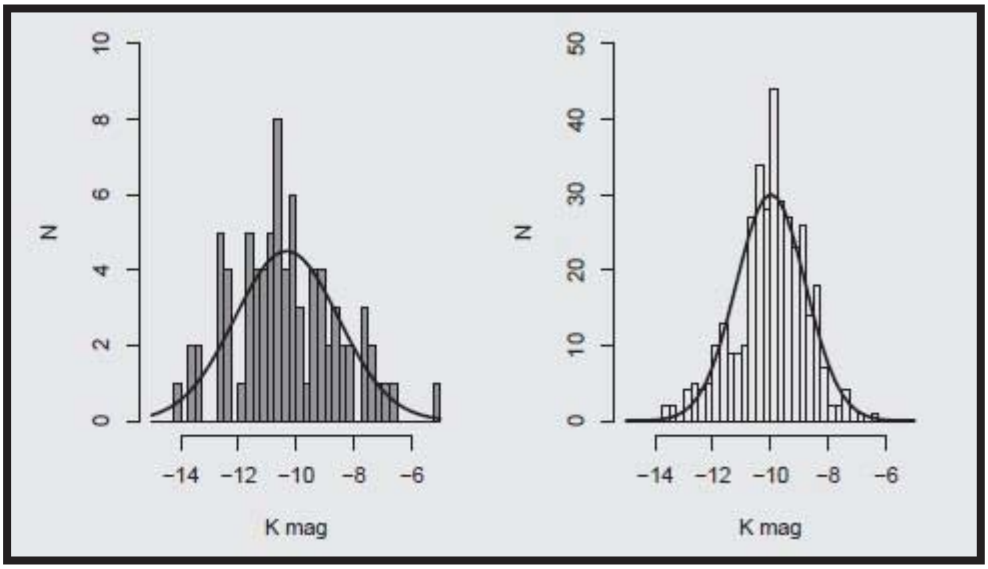
Comparing Pareto distribution estimators

Figure 7: A simulation of the initial mass function of 500 stars assuming a power-law (Pareto) distribution with α = 1.35. Panels show: (left) individual values; (center) logged probability density function and cumulative density function and (right) quantile function.



Fitting distributions to data

Figure 8: Distribution of K-band absolute magnitudes of globular clusters in the Milky Way Galaxy and M 31 (Andromeda) Galaxy.



Non-Parametric Statistics

Figure 9: Simple plots for a small univariate sample of asteroid density measurements. Top: A dot chart. Bottom: Plot with heteroscedastic measurement errors.

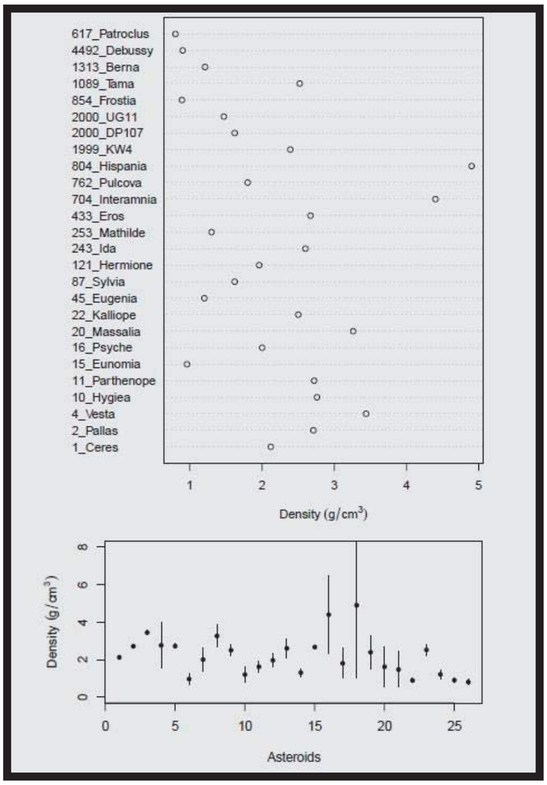


Figure 10: Boxplots of asteroid densities (left) and their measurement errors (right).

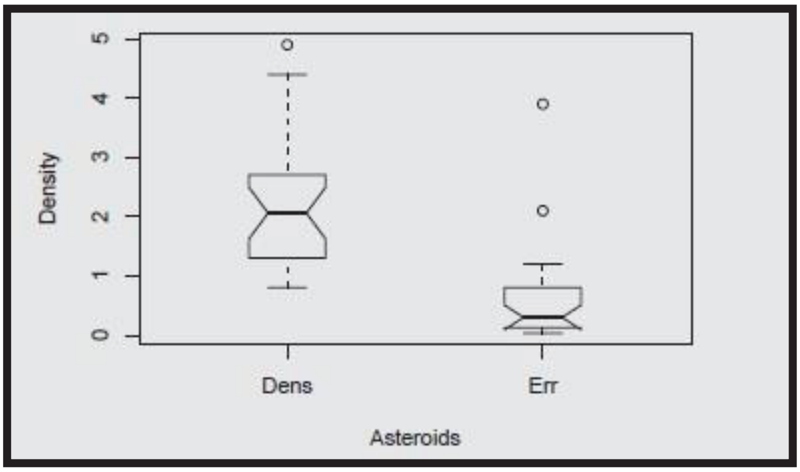


Figure 11: Empirical distribution functions for the K-band magnitudes of globular clusters in the Milky Way Galaxy and the Andromeda Galaxy (M 31).

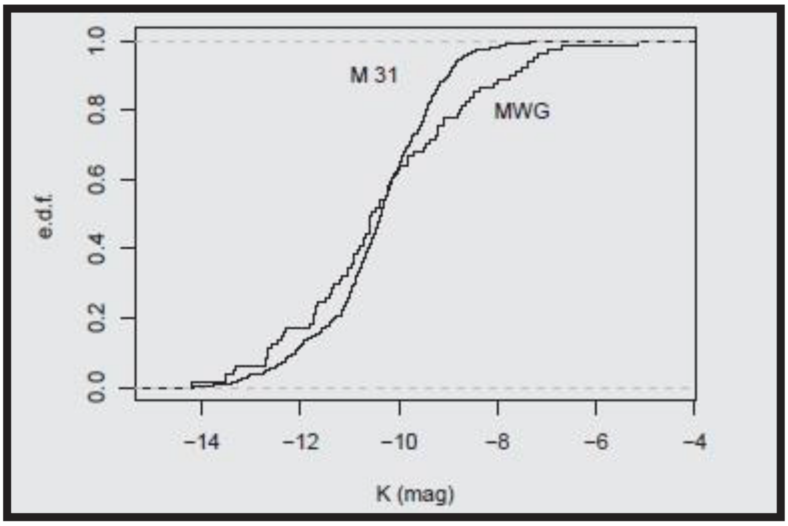


Figure 12: Quantile-quantile plots for comparing univariate empirical distributions. Left: Comparison of Milky Way and M 31 globular cluster distributions. Middle and Right: Normal Q-Q plot comparing the Milky Way and M 31 globular cluster distributions to a Gaussian.

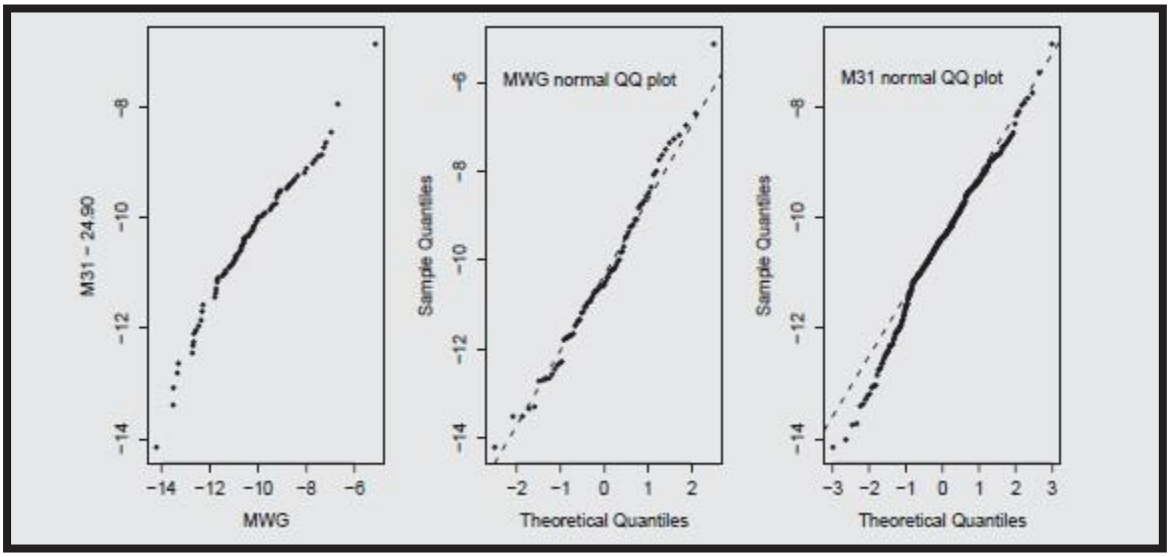
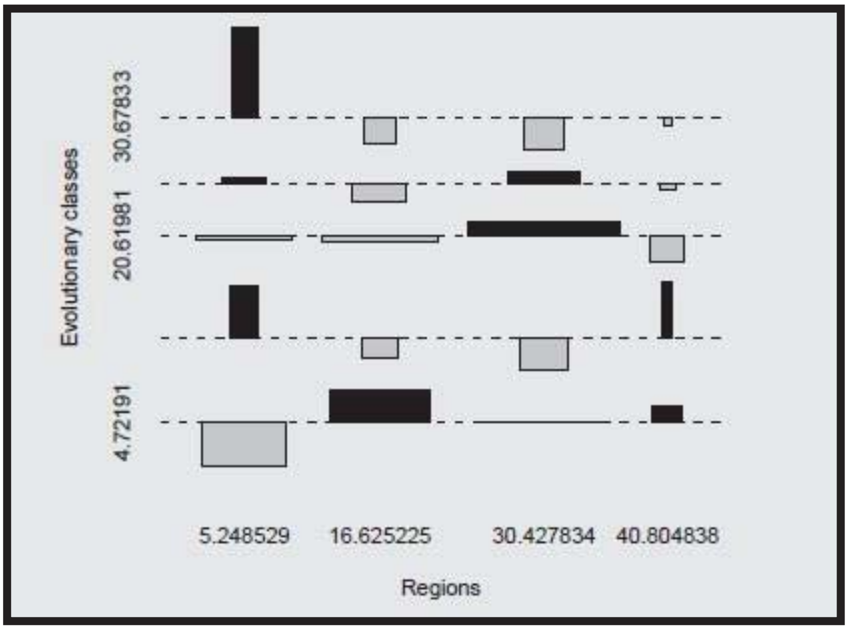


Figure 13: Association plot for an r ×c contingency table showing the population of young stars in five evolutionary classes and four star-forming regions.



These are only a few applications of statistical graphing that can be done by "R". Some of the other applications can be in the following data processing techniques, but only a few more will be accompanied by figures so as to not miss the point of this dissertation.

- **Data Smoothing: Density Estimation**
- **Regression**
- **Multivariate Analysis**
- **Clustering, Classification and Data Mining**
- **Time Series Analysis**
- **Spatial Point Processes**

To illustrate the power of "R" however, it must be seen that "R" is capable of plots that distinguishes between colours as well such as in Figure 14 and Figure 15as well as the 3D capabilities as illustrated in Figure 16. When looking at these images, it is easy to see that image processing within "R" is not only a possibility, but can be used to illustrate the needed data within business processes as well.

Figure 14: Bivariate distribution of r − i colour index and redshift of SDSS quasars. Gary-scale and contours are calculated using averaged shifted histograms. Red and green curves are calculated using LOESS local regression.

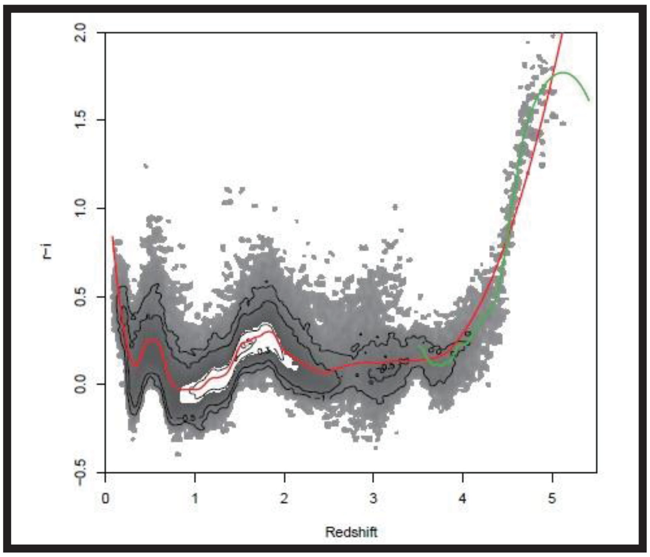


Figure 15: The low-density region of the Shapley Super cluster galaxy redshift survey plotted with CRAN's spatstat package. The colour image shows a kernel density estimator of the galaxy distribution, and the circles show the individual galaxies with symbol size scaled to their velocities.

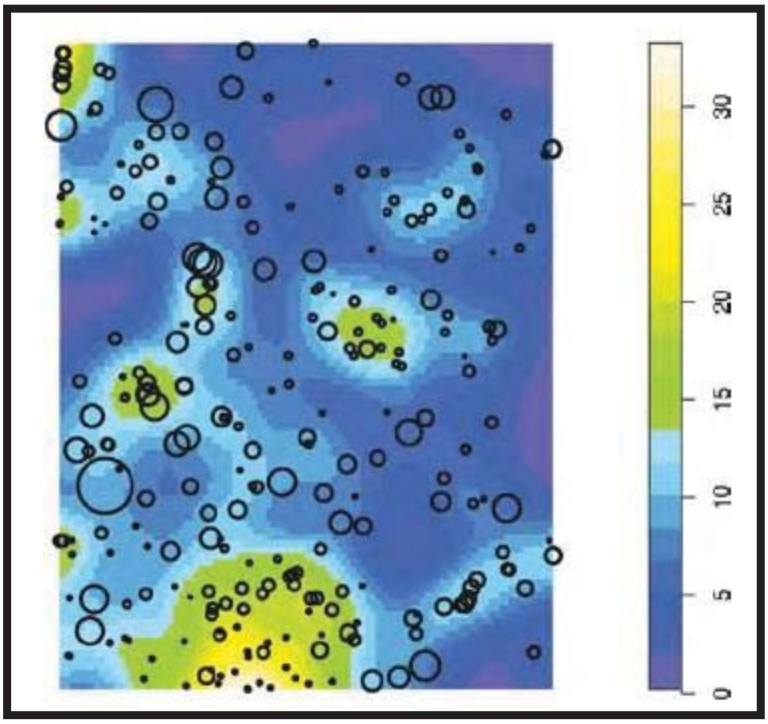
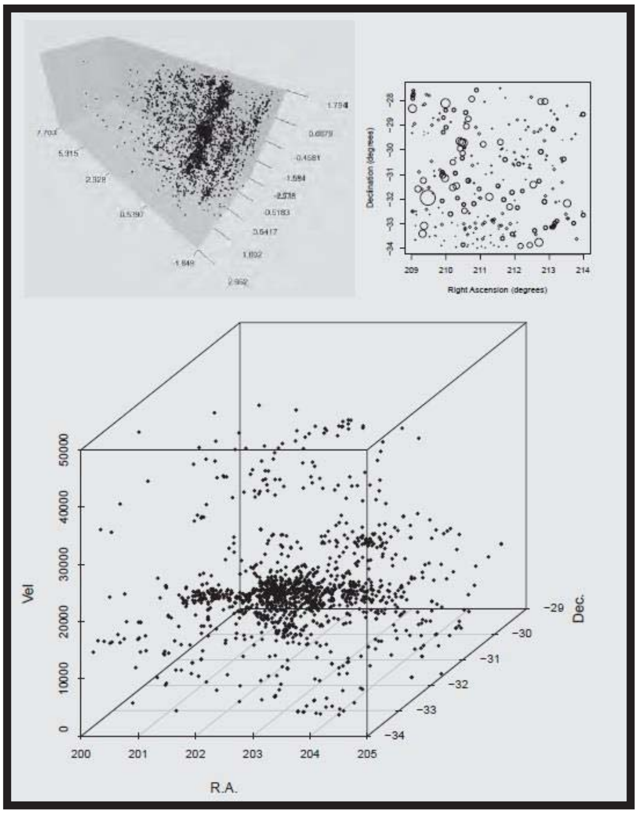


Figure 16: Three-dimensional views of the Shapley Super cluster galaxy redshift survey: Top left: Full survey produced with R's *rgl* package; *top right*: low-density region produced with R's *plot* function with symbol sizes scaled to galaxy distance; *bottom*: high-density region produced with CRAN's *scatterplot3d* package.



2.2.3. "R" and Bio-Informatics

The following abstract is an explanation of an actual processing package used within the "R" environment to analyse images in the field of bio-informatics (Pau, et al., 2010):

"EBImage provides general purpose functionality for reading, writing, processing and analysis of images. Furthermore, in the context of microscopy-based cellular assays, EBImage offers tools to segment cells and extract quantitative cellular descriptors.

This allows the automation of such tasks using the R programming language and use of existing tools in the R environment for signal processing, statistical modelling, machine learning and data visualization."

To illustrate the power of this package, Figure 17 to Figure 21 will show the steps that are taken to analyse these images from microscope images to cellular phenotypes.

Figure 17: Fluorescent microscopy images from three channels of the same population of HeLa cells perturbed by siRluc.

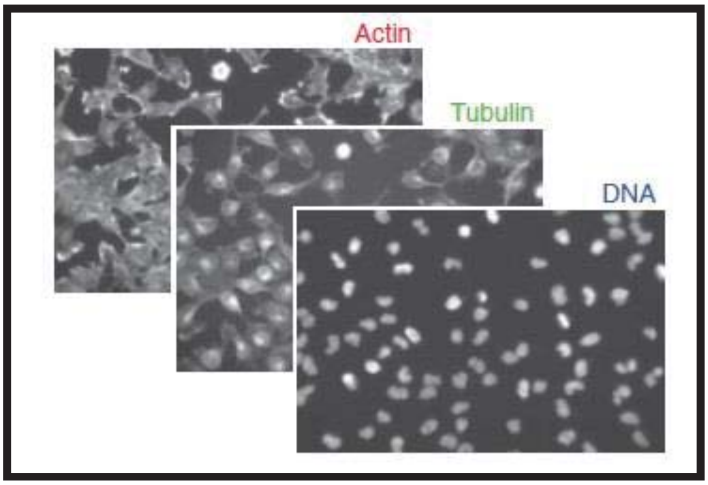


Figure 18: A false colour image combining the actin (red), the tubulin (green) and the DNA (blue) channels.

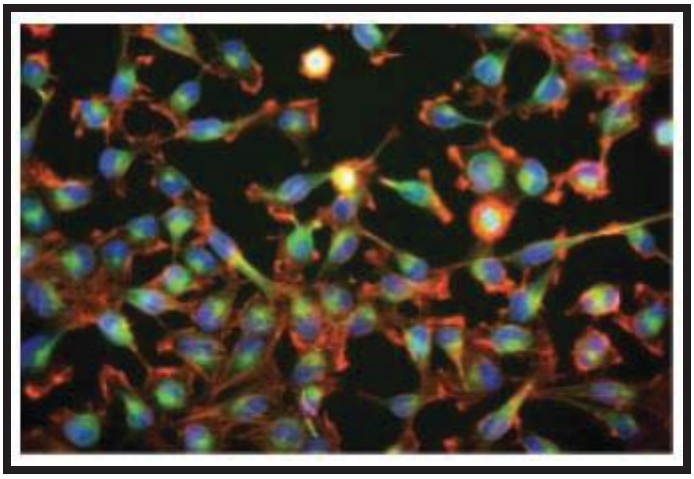


Figure 19: Nuclei boundaries (yellow) were segmented with adaptive thresholding followed by connected set labelling.

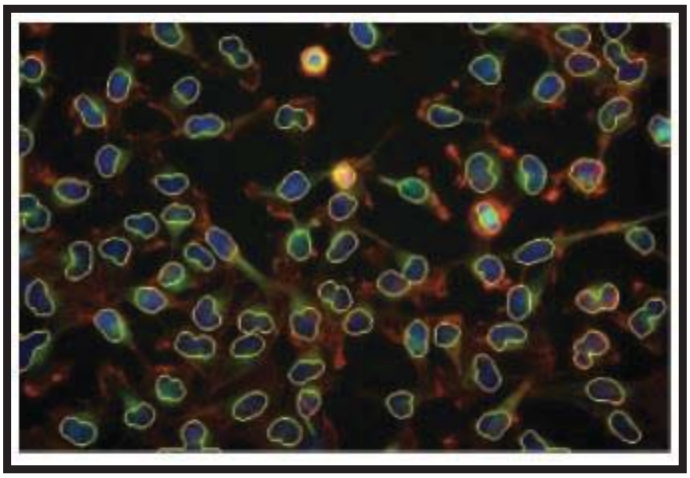


Figure 20: Cell membranes (magenta) were determined by Voronoi segmentation.

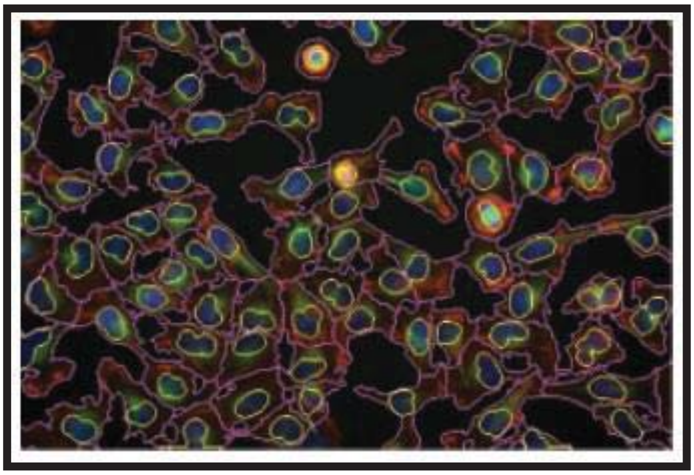
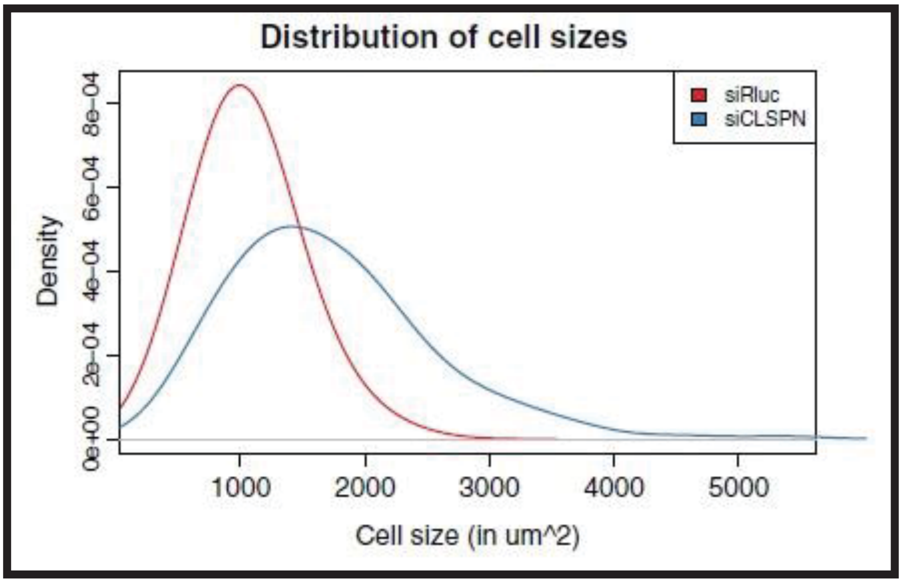


Figure 21: Distribution of the cell sizes compared to a population of HeLa cells perturbed by siCLSPN. Cells treated with siCLSPN were significantly enlarged compared to those perturbed with siRluc (Wilcoxon rank sum test, $P<10^{-15}$).



As can be seen yet again the capabilities within the environment of "R" seems extensive enough to give the idea of relating these methodologies used in astrophysics and bio-informatics to processing business images sufficient plausibility.

2.2.4. Image Analytics in a Business Environment

After seeing that it seems plausible to process images within the context of business processes, it seems that the dissertation is now at the point of trying to determine which pathway needs to be taken and which methods need to be used to successfully determine productivity measures from business process images.

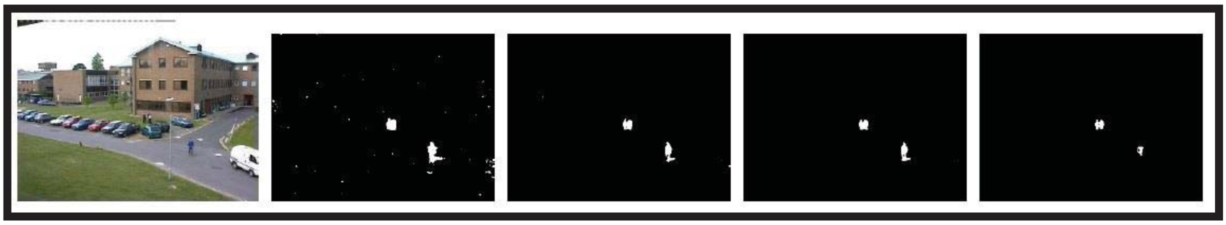
Before getting into the possible application of the software within this specified environment it is important to first consider which types of images will ultimately be used for analysis.

In general most businesses, factories, and plants are fitted with CCTV (Closed Circuit Television) which is primarily used as for detection of shoplifting and have proved to have a variety of uses to justify investment, such as deterrent, records for insurance claims, public safety, stock tracking and employee fraud (Senior, et al., 2007). It seems though that this is still fairly labour intensive if used in many of these manners. The possibility of image analysis comes to mind, and when in combination with image analysis software, will give users a method of running these still images through a single package to determine facial recognition and stock taking so as not only providing the above stated uses, but other uses such as possibly determining productivity measures for a company on an on-going basis, which will not only provide better seasonal forecasting, but gives the possibility of providing a middle ground solution for smaller businesses to have data resulting in growth , but not breaking the bank.

The question however is where to start and which images will be sufficient for analysis. The first type of image that needs to be considered is to track customers within an environment to provide details such as time spent in a queue, time spent being serviced as well as total time within a system. The following methods may be applicable:

Parzen Density Estimation (Tanaka, et al., 2007)
This method is a simplified version tracking a person, and can be used to see where a person is in an environment at a specific time. It is however flawed, but seems a good starting point.

Figure 22: Tracking using Parzen Density Estimation.

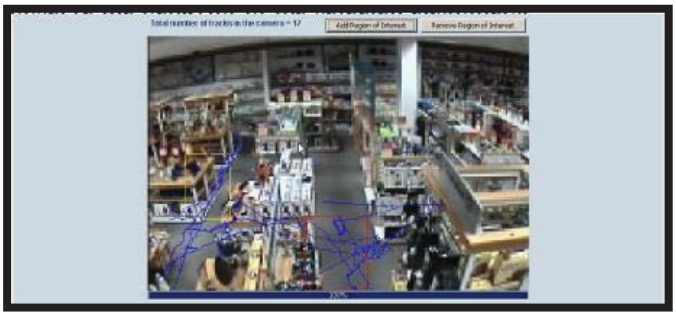


In this experiment, supposing that Red, Green, and Blue components of pixel values are independent of one another, it is estimated as a one-dimensional Probability Density Function of each component. Then, a pixel is judged as a foreground pixel when at least the probability of one component, Red, Green or Blue, is below a given threshold. In this manner only the people in the forefront are supposed to be left on the image and all else is left in black. Depending on the kernel width h and the pixels of the image, the smoothness of the images is approximated. Even though this obviously determines position, it leaves no way of identifying individuals, which makes it less desirable.

Another experiment shown in Figure 23 and Figure 24 illustrates how image analysis is used to track the traffic flow of people within a shop (Senior, et al., 2007).

Figure 23: Colour density plot of customer activity.

Figure 24: Customer trajectories.



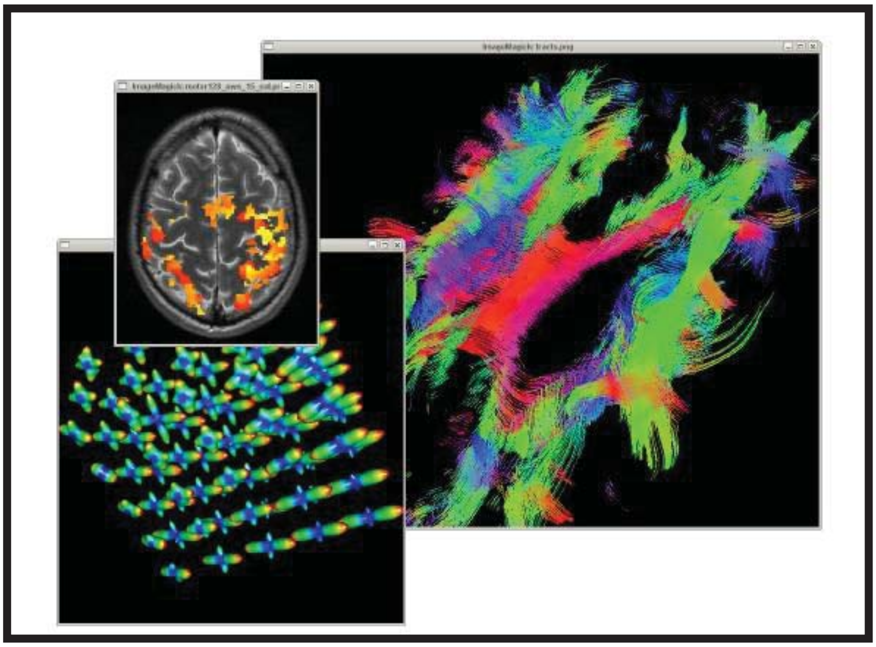
These two examples are only a fraction of the possibilities that can be explored in the field of image analysis, but offer the examples of the two most important methods that need to be explored, object recognition, and object tracking.

2.2.5. Possible Application of "R" within a Retail Environment

Out of all of the above mentioned reading material, it seems that the possible application of the software "R" and the methods to use lie predominantly in the linking of both the statistical processing capabilities of the software with its image analysis capabilities. To provide insight within this, Figure 25 and Figure 26 illustrates how image analysis and statistical inference in neuroimaging is combined within "R". This yet again illustrates that within this software package, image analysis within other fields of work is not only probable, but definitely possible as long as the correct statistical data and imaging package is used in accordance to the programs capabilities.

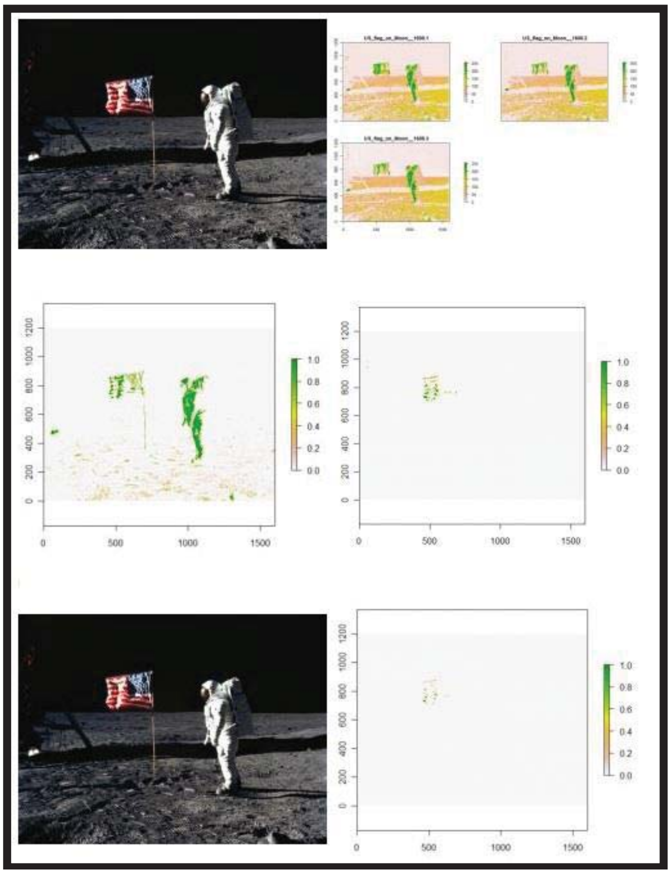
Figure 25: Snapshot of AnalyzeFMRI at work.

Figure 26: Results from the FMRI and DTI package.



Another possibility that was explored by a "R" user is to search for colours that may be predominant to track people in an environment, as illustrated in Figure 27 that proves it is indeed possible to eliminate unnecessary noise to discover the desired (Charpentier, 2012), even though it may be hard in other scenarios.

Figure 27: Flag Search.



2.3. CHAPTER SUMMARY

R provides an excellent environment for all levels of image analysis various data, from basic image processing to advanced statistical techniques via the current list of contributed packages in the needed environment, as well as integration with many other platforms. It is possible to analyse data sets and images individually, as well as having the option of automated bulk analysis of data sets and imaging data. The user is free to create additional data structures or analysis routines using the programming environment in “R” - making it easily customizable. It may be run in either interactive or batch-processing modes in order to scale with the application, and may be combined with other computing environments to allow for even greater flexibility (Tabelow, et al., 2008).

CHAPTER 3

3.1. INTRODUCTION

The technical analysis and literature behind a dissertation is only as good as its physical implementation possibilities, and therefore it is necessary to not only provide the research behind the desired model, but a plausible design to determine whether said model is not only possible as a study, but as a physical model. Is it possible to determine productivity measures from business process images? In theory, yes, but for all its practicality it seems to not as yet have been implemented in any businesses.

3.2. CONCEPTUAL DESIGN

A conceptual design usually gives precedence to the hypothesis of the functioning on the subject of this dissertation. As it seems that image analysis is poorly implemented in the field of business except for analysing security and to sustain security measures, the options that are available need to be explored to determine whether it is indeed plausible to take the research that was done and actually implement. As a first step it must be determined which approach of conceptual design will be followed, i.e. a phase based design or an activity based design. As this dissertation is mostly research based, it seems more viable to use a phase based approach to determine the conceptual design. A simple example of the phase based design that can be seen in Figure 28, (Wodehouse & Ion, 2010) will be used to lay the foundation of the conceptual design that will be used for this dissertation.

Figure 28: Phase-based concept design process.

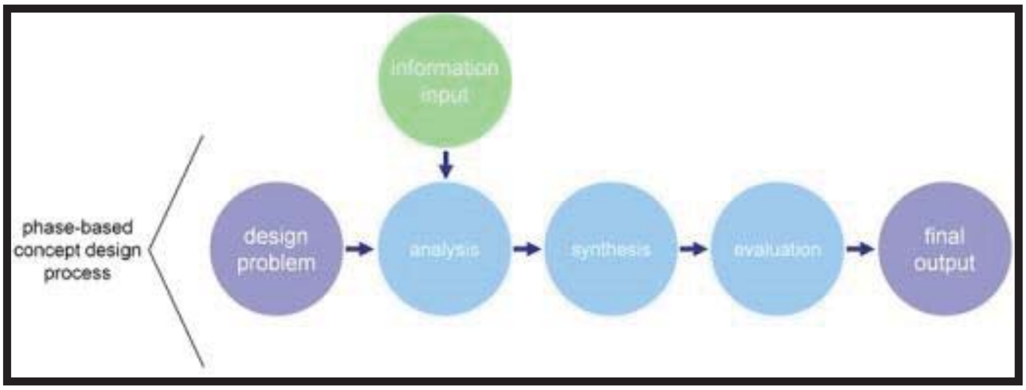
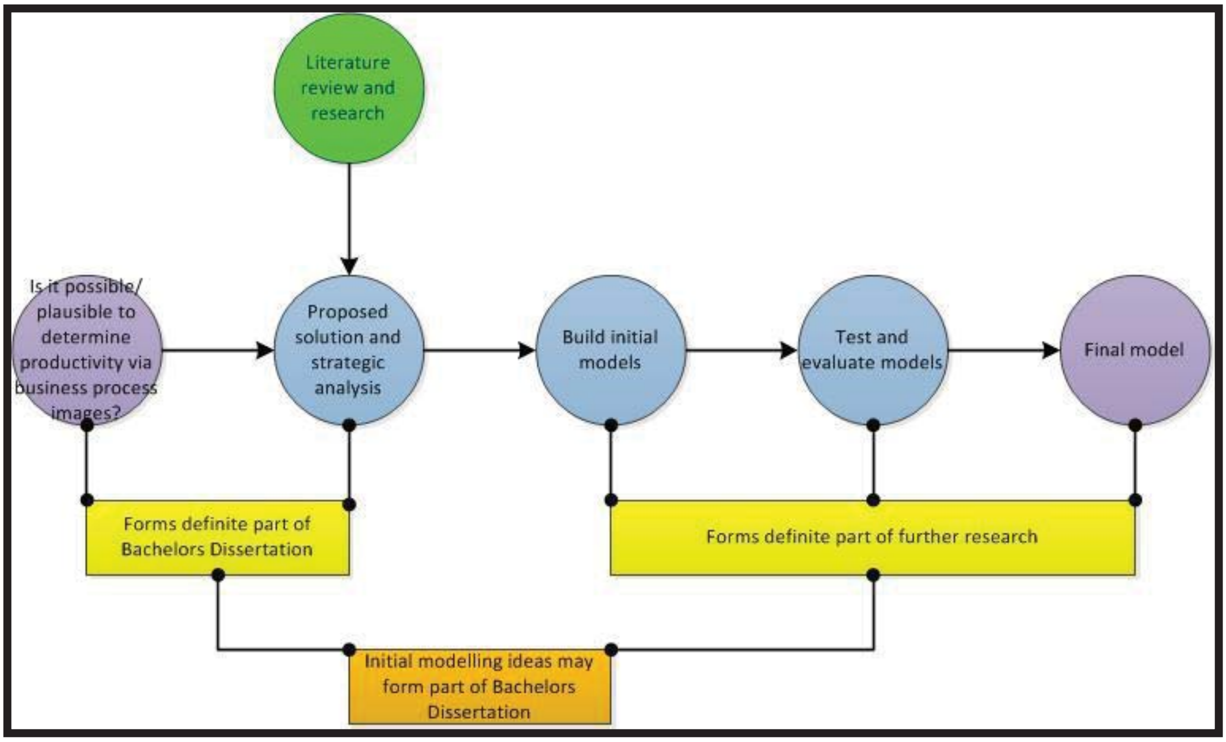


Figure 29 simply illustrates the phases this project will follow and the possibilities that will be available after completion of this dissertation. The scope for this dissertation was determined to remain in the analysis phase, but part of the initial synthesis will be done by using literature pertaining to statistical analysis and image analysis packages within the "R" software environment.

Figure 29: Simple Concept Design



3.2.1. Software Model in "R"

The initial part of the design consists of building a basic model in the software package "R". To build this model it is necessary to first determine which one of the discussed statistical models will conform best to the type of image analysis that is to be used. This model will be first built as basic as possible to firstly incorporate tracking designated forms from still images. If that can be done seamlessly for basic forms, the next step will be to identify forms that are not uniform, to find whether it is traceable at different positions in time. When the software model is refined to trace non-uniform object from still images, the required outputs are achieved and the code is established to be uncluttered and thoroughly usable in all tests, the next phase will start. Applicable examples of the code that may be used in the developing of the software model will be shown in APPENDIX D: "R" .

3.2.2. Physical Model

Decision of whether to progress to further implementation of the above mentioned software model that is to be built will be determined by the outcome of the analysis of the physical model.

The idea is to build a scale model of a basic room. Within this room there will be randomly placed models of people as well as models of stock. These models will be simple and of colouring that mostly consists of prime colours red, green and blue, and with the walling being predominantly white. Still images will then be taken from different angles as well as of the models in different positions to establish whether it is possible to determine position at a certain time to also find the effectiveness and accuracy of different methods of tracking. Photos will also be taken at different resolutions to determine at which resolutions accuracy will become problematic.

As this will be a step for step analysis, it is pertinent that every angle is researched before moving on to the next step. It is also utterly important to make sure that the outputs remain constant and that all desirable outputs are found from the images that are used as input.

The following step will then be to enhance the model in such a way as to determine how accurate the model will be within different conditions such as where the walls are coloured or has patterns and whether this can be removed to use the model as planned.

3.2.3. Case Study

If, and only if it is deemed possible that, after the analysis of the physical model, the output data is not only statistically correct and usable for input into calculating productivity measures but usable within many different conditions, will the software model be implemented into an environment that is not primarily a test environment.

A case study is set up to be performed after successful testing done in the previous stages, at a fuel garage in Pretoria, where digital copies of images taken of the environment will be used to analyse, as well as point of sale information to determine whether if the two coincide. The main objective of this case study is to study the movement of cars to determine throughput and volumes.

The imagery will also be used to determine service time at the fuel pumps, estimation of queue times at peak times at the pumps, and of how the customers move within the shops.

3.3. CHAPTER SUMMARY

To prepare for implementation further studies need to be made and all probable angles need to be tested, as this phase concludes the easy and inexpensive back out option. Within the confines of this phase it is still easy to change scope or better define the outcomes, but as this phase progresses, it will become harder to make changes. It is therefore pivotal that all is done that can be to assure the positive outcome of the research within this dissertation. As further research has commenced, it was noted that the intended models that is proposed in this section will expand the scope of this project in the general direction of Master's Degree dissertation, where the implementation of this research model can be better built and funded, and where due credit may be achieved in the future.

CHAPTER 4

4.1. INTRODUCTION

During all the previous chapters the fact that there is a need for newer and cheaper methods of measuring productivity was highlighted, as well as the fact that there is not a method that can be used by businesses in the small to medium market, and by discussing these shortcomings, an alternate solution, that being the determination of productivity measures from video analysis by using software called “R”, was proposed. The application of this software within other fields of study was also discussed, as well as the success of these applications. Within the following chapter, the development behind the software application within the business industry will be explained, as well as another alternative that was developed in the past year that is actually in a beta phase, and by using this alternative as an example, will prove that the application of this method of determining productivity measures is not only relevant, but much more practical within a business environment as was initially expected. The research into “R” and the development behind the software that will be used as a guide for this dissertation will not only be evaluated against the “off the shelf” alternative as usual, but will form the foundation of the proof needed to substantiate the claims that this proposed method of determining productivity measures is indeed viable as well.

4.2. PRISM SKYLABS

The following extract was taken from the website of Prism Skylabs (Prism Skylabs, 2013) and is a short explanation of what the company is about:

“Prism Skylabs is the global leader in visually understanding and optimizing offline commerce. Its unique cloud service transforms any video camera into a visual merchandising, auditing, and business intelligence tool that can be accessed from any device.

With Prism Skylabs, businesses can measure offline conversion, understand long-term trends, and dig in to powerful analytics — from dwell to footpaths to product lift — using the cameras they already have. Prism Skylabs condenses customer interaction and movement into stunning imagery and reports that provide instant understanding of any moment or period of time.”

This very interesting company is the brainchild of Silicon Valley veterans Stephen Russell (digital video guru, founding CEO and current chairman of 3VR Security) and Ron Palmeri (the guy behind Grand Central/Google Voice, Scout Labs, and OpenDNS), and have through the process of trying to understand retail business better, come up with the ingenious idea which is Prism Skylabs. After the company was in stealth development mode for several months, it was launched at an annual competition where start-up companies get to showcase their products and ideas, and thereby get sponsors and support from the wider business community. This competition is known as Techcrunch Disrupt, and at the 2011 event Prism Skylabs walked away with the second place. This has since earned the company in excess of

$10 million (approximately R100 million) in funding, showing the relevance of their software within the retail environment.

The product in itself boasts a myriad of features, being versatile enough to work in several areas of business, such as marketing, merchandising, store operations and loss prevention, while the application within industry is just as wide, covering retail, consumer packaged goods and restaurants just to name a few. At this moment in time it is possible to use the product on a 30 day free trial, after which further use with exceptions will be on a month-to-month basis at a price of $99 per month (approximately R1000). With this version one can get access to 10 cameras, 1 user and 1 advanced analytic. Any optional extras will obviously need to be purchased at extra expense to the user. The product works on the same principle as many of the new-age action cameras, in which it takes background "noise" to enhance the image quality and stabilise the image to receive great high definition images. It can be used on several mobile devices, as is showcased in Figure 30, and is easily downloaded from their website. Integration with cameras is done automatically, as the application scans all available networks associated with the owner of the stores even if the stores are not on the same premises.

Figure 30: Log in available from any mobile devices.

As can be seen in Figure 31, the owner then has access to the cameras from these mobile devices, and can enjoy the versatility of real time views of the stores, as well as the flexibility of jumping between stores as can be seen in Figure 32 and Figure 33.

Figure 31: Enhanced, real time views of the business area.

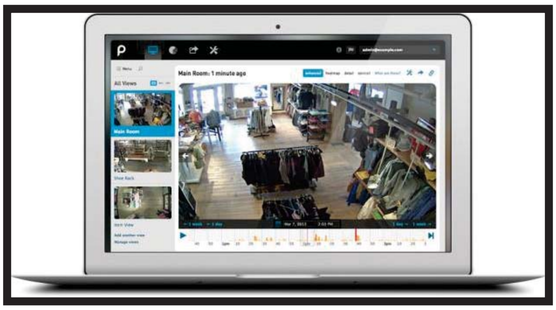


Figure 32: Jumping between stores, Store 1.

Figure 33: Jumping between stores, Store 2.

The ability to have access to these stores at any moment in time is what really gives owners the freedom to explore the rest of the features that are associated with this product. The features that are highlighted on their website when taking the free tour is seen from Figure 34 to Figure 42, and will be briefly discussed as the images are rather self-explanatory.

Figure 34 shows how easily the user can zoom into any of the images to inspect shelves, products and displays, and gives the user the freedom to conveniently manage when to fill depleted stock and in cases where a retail shop may not have a stock management system installed, gives the user access the ability to notice when more stock is needed.

Figure 34: Ability to zoom into shelves, products or displays.

As consumers in most cases have the right to privacy, this software is able to simply "grey out" customers, as can be seen in Figure 35, but still retain the same functionality of the software, be it measuring customer dwell, Figure 36, or measuring product lift, Figure 37. In both cases the customers need not worry that their privacy be violated, and the user, in addition to the software, can assure customers another value added service, in terms of protecting their privacy.

Figure 35: Ability to protect customer privacy.

Figure 36: Measuring customer dwell.

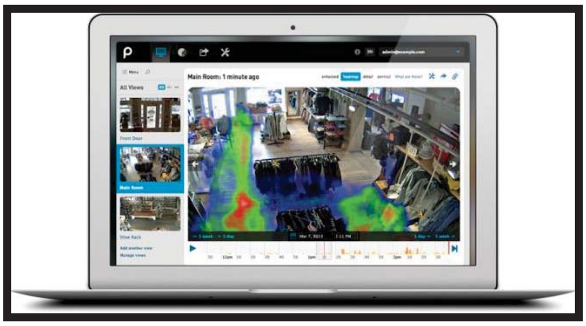


Figure 37: Measuring product lift.

By measuring the customer dwell and the product lift, the customer movement, Figure 38, and crowd sizes, Figure 39, within a store can be evaluated and help the owner of the store better design the layout of his facility to accommodate better flow and to introduce customers to more products.

Figure 38: Evaluate customer movement.

Figure 39: Evaluate crowd size.

At any point in time the owner of the store has access to calculated crowd size trends, Figure 40, and can easily let the software measure and calculate trend with on demand reports generated by the software, as can be seen in Figure 41.

Figure 40: Calculate crowd size trends.

Figure 41: Measure and calculate trends with on demand reports.

Another useful feature that the software offers is the capability to receive email and text notifications as events occur that the owner of the facility can set up and customise. When a software package seems to do so much, and many of the user's useful data runs the risk of getting lost, it is mandatory to have a paperless storage method, and the software remedies this obstacle by giving access to cloud storage of which the size can be extended as needed.

Figure 42: Ability to store data and images safely on cloud.

This product from Prism Skylabs is thus the basis on which one can argue that conventional methods of measuring productivity may well be managed from these mobile devices on software that is easy to use, with data that is easily available at any time, which not only gives business owners the flexibility to adapt their businesses on account of this data, but also the freedom to enhance their productivity while not paying an arm and a leg for it.

4.3. PROPOSED SOLUTION

As time studies become ever more tedious, and prices for the use of technology rises, the solution needs to take into account the following factors:

- Readily available
- Flexibility
- Ease of use
- Customisable
- Applicable to the South African business environment

As can be seen by the product that is available from Prism Skylabs, all of the above factors can easily be packaged in a beautiful piece of programming architecture that integrates easily with social media and ensures privacy of customers. The only item it does not seem to satisfy is its applicability in South African business environment, and is another reason why a solution is tailor made for the local market is necessary. This will also make out part of the main focus for the development of the solution for this dissertation.

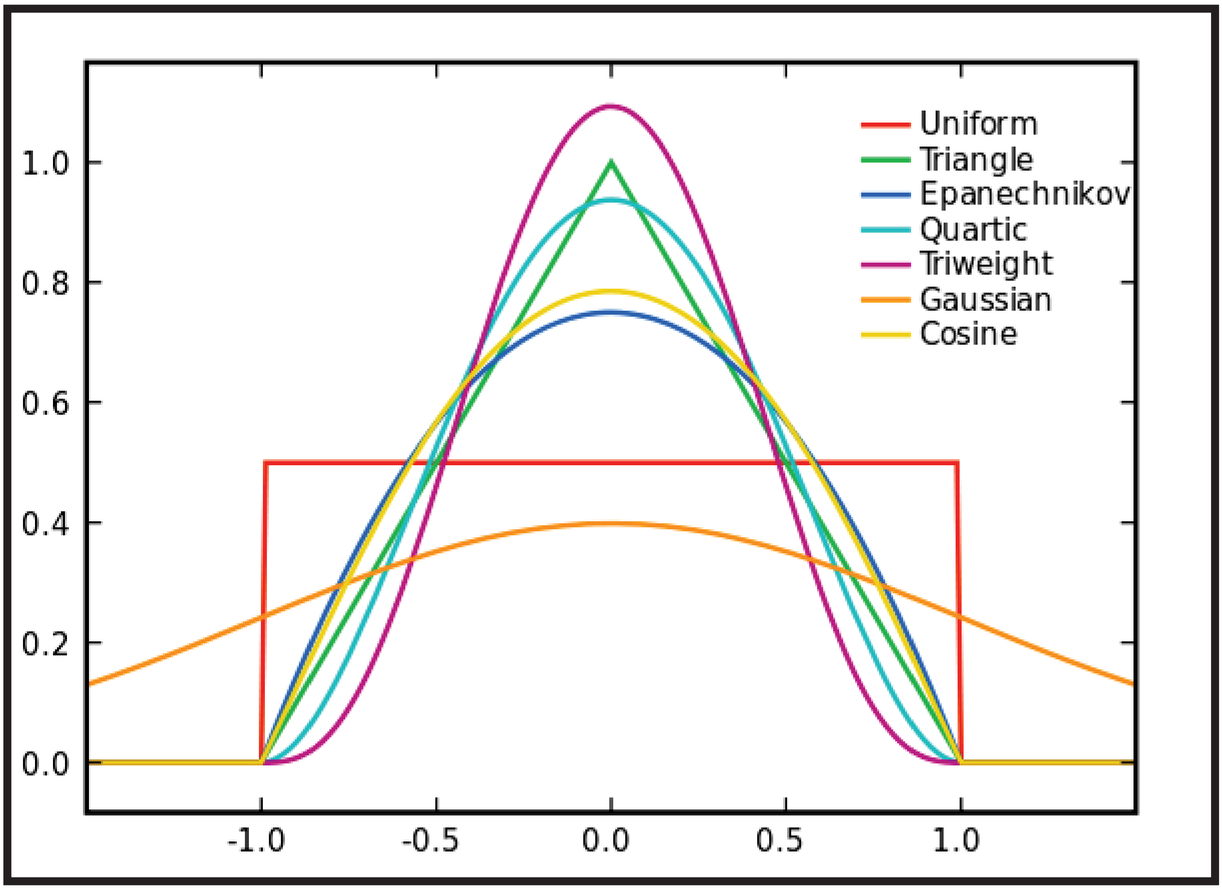
To perform the image analysis, a proper statistical method needs to be proposed as well to further leverage off to implement the technique in a business environment. The solution that was developed by Prism Skylabs seems to rely heavily on a form of pattern recognition that has the ability to distinguish between the people in a store and the products within the store. In a dissertation about Parzen-Window Density Estimation (Awata, 2007), it was noted that this statistical model is best known for its use in the areas of pattern recognition, classification, image registration, tracking, image segmentation, and image restoration. It therefore seems imperative that it needs to build the foundation of the proposed solution. To refresh the memory the basic formula for Parzen-Window Density Estimation is shown in Figure 43 (Tanaka, et al., 2007).

Figure 43: Parzen-Window Density Estimation Equation.

$$P(X) = \frac{1}{N} \sum_{i=1}^{N} K\left(\frac{X - X_i}{h}\right).$$

This equation needs to be modified however depending on the kernel function K that needs to be used. A kernel in statistics is a weighting function commonly used in density estimations. The kernel width must always be specified when running nonparametric density estimation. Figure 44 displays the different functions in graph form, (Amberg & Vetter, 2012).

Figure 44: Kernel density functions.



The functions that relate to most of the afore-mentioned graphs can be seen in Figure 45 (Zucchini, 2003), and also interesting to note is the efficiency that relates to the use of each of the functions.

Figure 45: Kernel function equations

Kernel	$K(t)$	*Efficiency (exact and to 4 d.p.)*
Epanechnikov	$\frac{3}{4}(1-\frac{1}{5}t^2)/\sqrt{5}$ for $\lvert t\rvert < \sqrt{5}$ 0 otherwise	1
Biweight	$\frac{15}{16}(1-t^2)^2$ for $\lvert t\rvert < 1$ 0 otherwise	$\left(\frac{3087}{3125}\right)^{1/2} \approx 0.9939$
Triangular	$1-\lvert t\rvert$ for $\lvert t\rvert < 1$, 0 otherwise	$\left(\frac{243}{250}\right)^{1/2} \approx 0.9859$
Gaussian	$\frac{1}{\sqrt{2\pi}}e^{-(1/2)t^2}$	$\left(\frac{36\pi}{125}\right)^{1/2} \approx 0.9512$
Rectangular	$\frac{1}{2}$ for $\lvert t\rvert < 1$, 0 otherwise	$\left(\frac{108}{125}\right)^{1/2} \approx 0.9295$

It is however important to note that the graphs and equations are only given as background to the statistical model, and will not necessarily be utilized in the case of this dissertation, as the software package "R" has the necessary functions built into the code to be able to easily switch between different kernel widths. The following code is an example of the density source code that is used in "R" (R Foundation, 2000):

```
1    density(x, ...)
2    ## Default S3 method:
3    density(x, bw = "nrd0", adjust = 1,
4            kernel = c("gaussian", "epanechnikov", "rectangular",
5                       "triangular", "biweight",
6                       "cosine", "optcosine"),
7            weights = NULL, window = kernel, width,
8            give.Rkern = FALSE,
9            n = 512, from, to, cut = 3, na.rm = FALSE, ...)
```

In the above source code it can be seen how easily the user can jump between the different kernel functions as needed by just substituting with the applicable function in line 4, and "R" will automatically call the function. This will only be part of the image analysis code, and will most probably run within one of the image analysis packages in the "R" environment. The most common packages that can be used are found as part of CRAN, which is the main resource package downloadable within the "R" environment. The literature review shed some light on the more common packages such as "AnalyzeFMRI" (Lafaye de Micheaux &

Marchini, 2013), “biOps” (Bordese & Alini, 2013), and “EBImage” (Pau, et al., 2013). The example that will be showcased in the following section is a practical example of how the package “EBImage” was used to identify that there is indeed data hidden in all images, and this will serve to indicate that it is indeed possible to extrapolate data from images, and that when used in conjunction with the correct statistical model, it is not only possible to determine productivity measures from image analysis, but may lend a hand to improve business as usual within any institution.

Figure 46 was read into “R” using the following code (Campbell, 2012):

```
1    > require(EBImage)
2    > require(spatstat)
3    > Cat <- readImage("cat.jpg")
```

Figure 46: Toy cat in "R" environment.

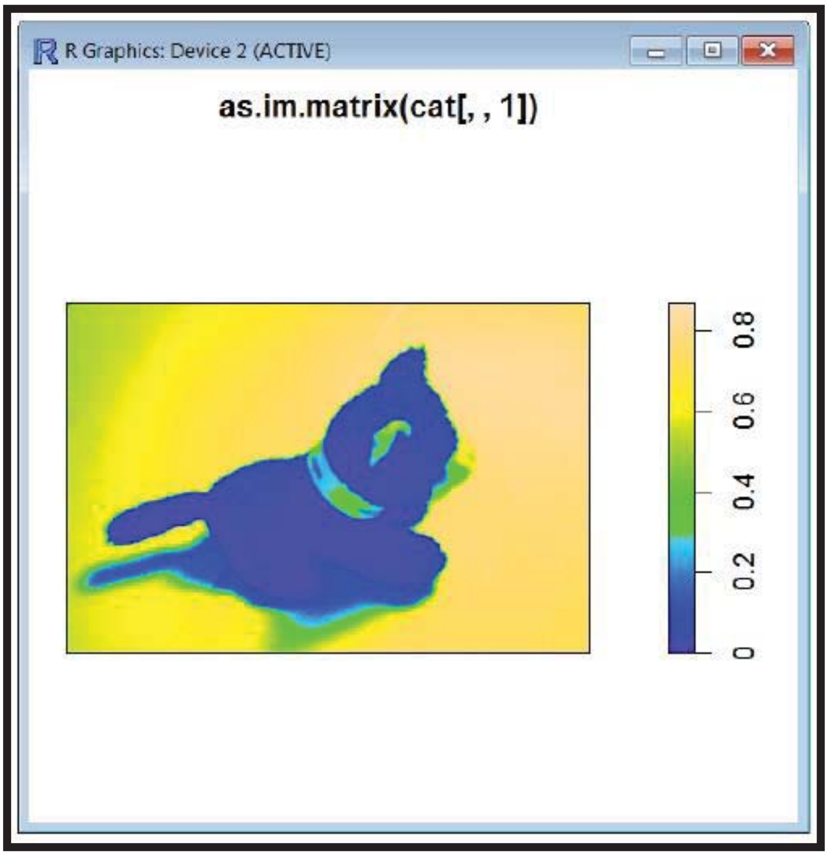
In line 1 it can be seen that the “R” environment calls up the package “EBImage” to do the processing of the image and scans the image into the “R” environment. To analyse the image, it first needs to be rendered to a compatible format, in this case, an RGB-image. The following code was used (Campbell, 2012):

```
1    > plot(rgbim(
2    + as.im.matrix(Cat[, , 1]),
3    + as.im.matrix(Cat[, , 2]),
4    + as.im.matrix(Cat[, , 3]),
5    + maxColorValue = 1),
6    + valuesAreColours = TRUE,
```

```
+ main = NULL)
> plot(as.im.matrix(Cat[, , 1]))
> Cat[101:103, 121:123, 2]
```

Figure 47 illustrates the image that resulted of the above mentioned code and the data that follows Figure 47 is the output that was calculated from the image.

Figure 47: Toy cat - RGB image.



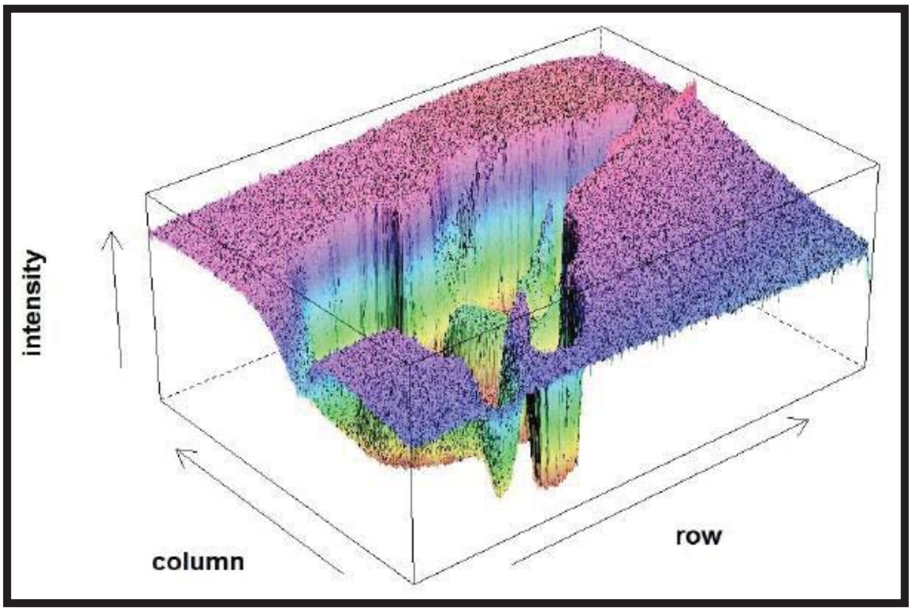

```
Image
    colormode: Color
    storage.mode: double
    dim: 3 3
    nb.total.frames: 1
    nb.render.frames: 1

imageData(object)[1:3,1:3]:
          [,1]      [,2]      [,3]
[1,] 0.5411765 0.5411765 0.5411765
[2,] 0.5450980 0.5411765 0.5372549
[3,] 0.5490196 0.5450980 0.5372549
```

The final step that illustrates how much data really is hidden within an image is to plot it as a graph, and the following code was used (Campbell, 2012):

```
> require(lattice)
> intensity <- unclass(Cat[, , 1])
> wireframe(intensity, shade = TRUE)
```

The result from this code is the 3-dimensional graph shown in Figure 48. From this it can be theorised that when Parzen –Window Density estimation is used, this data plot will consist of the densities of an object that is moving, and as the position of the object varies from image to image, the data will change as well, and can be easily compared and then used to plot movement of the object, as well as easily identify the same object in different time frames. Ideally the product of this analysis is a still image where different objects can be traced throughout the image, and productivity measures can be extrapolated from these final images.

Figure 48: Data plot from toy cat image.



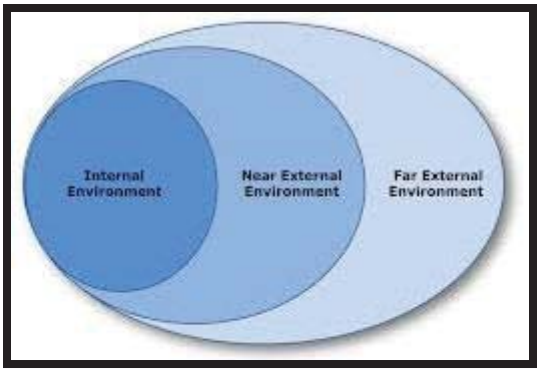
The logical approach illustrated in the above mentioned section is a theoretically developed solution that, with further research and development, can be developed into a working software model that successfully illustrates that this method can be much more beneficial as opposed to traditional methods of measuring metrics.

4.4. STRATEGIC ANALYSIS

Even though looking at a product such as the one showcased by Prism Skylabs and giving them credit for developing a product that can be so easily utilised fully in the country of origin, United States of America, it is very important to properly scrutinise the product against the primary solution suggested in this dissertation. South Africa is not the U.S.A., and while it may seem that this solution is the all mighty solution abroad, South African businesses may have other needs and infrastructure, which must allow for a solution that may be developed on the same ideas, but tailored to use in the South African environment.

Proposing a statistical model and considering software for the use of the programming may not prove sufficient and therefore it is very important to perform a proper strategic analysis based on the applicability of either product within this environment, and to fill a gap where there is a need, rather than only proposing a solution because there is one. Before developing a product it can thus be seen clearly that an eternal focus on strategy is important. For this dissertation two methods will be used to highlight which external factors to focus on when developing a strategy. These external environments are better illustrated in Figure 49, and are explained in-depth in APPENDIX E: ANALYSIS TECHNIQUES EXPLAINED (Killeen, 2012).

Figure 49: Three environments of business.

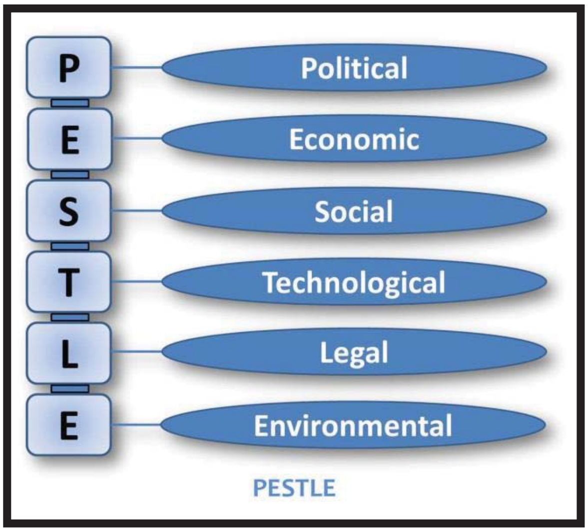


Firstly, a PESTLE Analysis will be done to analyse the macro environment which the product will function in, or as stated by Killeen, The Far External Environment. Secondly Porter's Five Forces will analyse the market that the product will function in, which is also known as the Near External Market.

4.4.1. PESTLE Analysis

Using PESTLE Analysis (Van Assen, et al., 2009), it is easier to determine which factors will affect the product in the present as well as in the near future, and will be discussed by component as illustrated in Figure 50. An in-depth explanation of PESTLE Analysis can be read in APPENDIX E: ANALYSIS TECHNIQUES EXPLAINED. In this environment the factors that are encountered are at most times factors that cannot be controlled, and many times businesses need to respond to developments in this environment. This is why planning is crucial. If the planning is done and one has some insight into possible setbacks, one can better react to these setbacks.

Figure 50: Components of a PESTLE Analysis.



Political

The political implications of this product and the business that may be a follow up on the product must be taken into consideration. Funding will be a necessity for initial as well as further development of the proposed solution, and to involve the reigning political party may be a move in a positive direction, as this can be seen as developing South African business, and the proposed solution can be leveraged even more if there is a possibility of having a B-BBEE partner. Government may then see the need to offer grants to further the research as well as development. It is however suggested that it may be prudent to wait until the current term of office is served and the general elections have passed, as this may give a measure of more stability within the country.

Economic

As the idea behind the software is to create a solution that can be distributed freely and used by all, the only economic implication of importance is funding. Even though the objective of the solution is to advance small to medium businesses and give that sector the ability to grow economically, which in turn will help the country economically as well, there will be a need for research and development funding. It can be seen in the piece about Prism Skylabs that product development was boosted by winning a major competition, giving them their initial funding. Saying that, the current state of South Africa's economy is not looking well, but many claim that it may be adjusting to an improvement in the following year (2014) and that the GDP will grow by an estimated 2% in 2013, making it a good time to invest in a venture sure to culminate the growth of the small to medium business sector.

Social

As the social needs of fellow South Africans evolve to a culture where being succesful is a venture, not a right, more entrepeneurs will surface, with ideas to not only to eleviate the individual, but to advance the country. Within this culture it is iminent that a product that is able to further the cause of these individuals is not only developed, but done in such a manner that it can be used to also help these individuals with ideas to further the cause of small to medium businesses and startup ventures.

Technological

In terms of technology South Africa has a bit of a setback. Whereas the Prism Skylabs product works on the principle that the industry they cater for all have access to established networks and internet, the South African industry cannot guarantee either. It is therefore pivotal to develop a solution that can be used by businesses that do have access to technology, as well as give businesses which do not have access to this technology a solution that is just as versatile and flexible in its use as normal. To develop for this situation, various infrastructure types and configurations need to be taken into account, ensuring availability of a solution that can cater to any technological need.

Legal

The legal implications associated with the initial development of this solution falls within the South African legal system, and need to be followed as prescribed by law. As legal matters usually arise as products are developed, a thorough understanding of law is necessary to not complicate development, and need to be dealt with as issues arise.

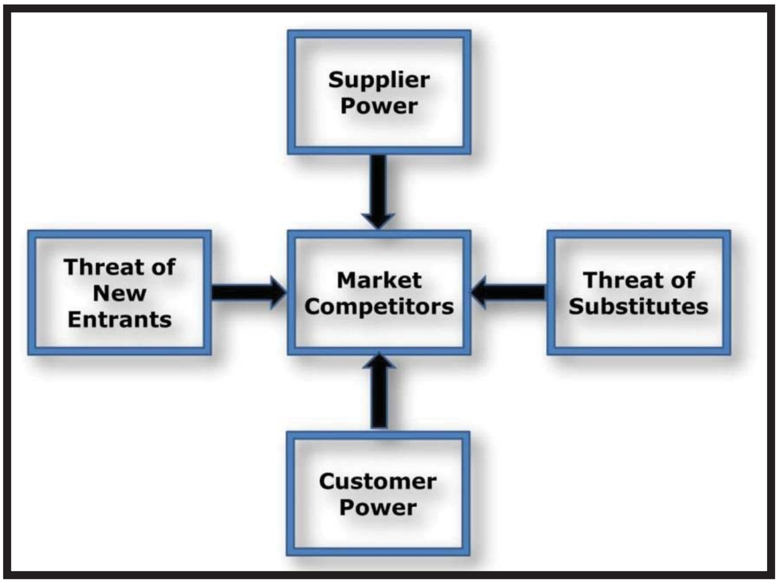
Environmental

As with many technological innovation in the present, the proposed solution will probable developed as a “Green” project, ensuring that no harm comes to the environment as the product is used. The use of a “cloud” storage system allows for reduction in use of paper, and if possible, the wider community that will make use of the product will be made aware of how the environment can be affected if proper “Green” technology is utilised.

4.4.2. Porter’s Five Forces

Using Porter’s Five Forces (Porter, 2008) as shown in Figure 51 (Killeen, 2012), an analysis of the industry where the product will function in can be done. To better understand the principle of Porter’s Five Forces, the in-depth explanation can be read in APPENDIX E: ANALYSIS TECHNIQUES EXPLAINED, but in simple terms, these are factors that can be influenced by the business owner, and most of these pressures can be handled to exert a positive outcome. For the use of this analysis, each segment will be scored out of 10 with regards to the level of pressure that may be experienced within each segment, with 1 being LOW amount of pressure, 5 being MEDIUM amount of pressure, and 10 being HIGH amount of pressure.

Figure 51: Porter's Five Forces That Shape Industry Competition.



Market Competition

3 – LOW TO MEDIUM Pressure

This is a market that currently is not exploited, as many small to medium businesses do not have the capital to afford that measures be put in place to measure productivity. The market has its fair share of big corporations that deliver other products, but the pressure here will only increase once the solution is successfully implemented and other competitors surface. It is an open-sourced solution which may mitigate the threat of competition by rather working together to develop a solution that can be used by a wide business sector.

Customer Power

6 – MEDIUM Pressure

The fact that the solution is yet to be implemented, makes for the pressure being relative to whether the product is successful or not. If the product is not successful the customers may easily revert back to either not having a solution or to more archaic solutions. If, however, the product delivers on its promises and abilities, the amount of customers that can be pulled into using the proposed method will be staggering, and may lead to more pressure to deliver yet more solutions to be incorporated with the product.

Supplier Power

8 – MEDIUM TO HIGH Pressure

As there are currently only older methods available for determination of productivity measures, there is a lot of pressure to deliver on the promise to provide a flexible and easy to use solution to measure productivity and provide seamless integration with aspects such as storage and social media. In a market where the solution initially will be developed to be distributed for much less than normal rates of these productivity measuring solutions, it can also be hard to keep the pricing under control. The fact that most of the code will be readily available for use can boost the development of new features, but may also provide a foothold for competitors to surface, and thus needs to be handled accordingly.

Threat of New Entrants

2 – LOW Pressure

Initially the threat level of new entrants will be low, but there may be a chance that the pressure may escalate as the development coding gets distributed. If the distribution is handled correctly though, and proper legal measures are in place that ensures that the solution can only be distributed to enhance flexibility and further development of the solution, it may be hard for a competitor to enter the market, as any solutions that have the same base of operations will be easily traceable. The fact that the solution consists of specialised code and a high level of skill in computer programming may also play a part in entrance into the market being less desirable,.

Threat of Substitutes

8 – MEDIUM TO HIGH Pressure

Within the context of the proposed solution, substitutes do not necessarily imply that there is a product that can produce the same results and works in the same manner, but can rather refer to the methodologies that is currently used in most businesses to determine productivity measures. As a solution it is easily be theorised that it is practical and relevant to implement a solution based on determining productivity measures from business process images, but within South Africa it is at times hard to convince people of change. If engineers have used time studies and newer methods for extensive periods in time, why develop something new. South Africa still has a culture of “If it isn’t broken, don’t fix it”. This keeps the country from fully realising its potential for growth, but can be sidestepped. As this solution increases the productivity in small to medium sized businesses, bigger companies and corporations may also adopt the technology for use in their own sectors.

4.5. CONCLUSION

Spending time watching television has taught individuals that only huge corporations and government agencies have funds to purchase image analysis technology and have individuals capable of using it. With this dissertation it is not only shown that image analysis can be done by any individual, but that the software that can be developed is easy to use, it can be freely available and the code can be distributed.

It is then important that this software is used as optimally as can be, as it has many applications within many fields of industry.

The possible use of this software as a tool to determine productivity measures from business images is an application that may not only provide insight into the use of “R” as a serious contender in statistical data analysis and image analysis, but may provide a great application in the business environment that has not been previously utilised.

When looking at the amount of development and funding that needs to go into establishing a world class product that can radically change the small to medium business sector, it may seem that spending that time might be better utilised elsewhere, but taking into account the state of Africa and other 3rd World Countries, it can possibly be a change that can put these countries on par with the world powers. Even if this product may take time to fully develop, it is always important to try and advance business, as not utilising technology may lead to businesses declining in their growth.

The strategic analysis done with Porter's Five Forces illustrates that there is some foreseeable risk included in the initial development of the proposed solution, and it is important to do further in-depth analysis to identify the most crucial risks, and it is in the mitigation of these crucial risks that successful results will be seen, and the benefits will soon outweigh the risks as the solution expands into a multitude of uses in various environments.

The most positive part of the proposed solution being developed in "R" is the fact that it is developed on an open sourced basis, giving all with sufficient skill the benefit of a product that is not only easy to use, but flexible in its use. It is important to note that the ideology of combining the package "EBImage" with statistical model such as Parzen-Window Density Estimation is just one of the alternatives to build a superior model, and still needs to be benchmarked in terms of efficiency and effectiveness. The "R" community is one that grows daily, and there will always be someone that will be available to readily provide even more help and insight.

The most important factor to consider is always the customer, the user of the product, or even the person that will be able to utilise the product when released, or who can benefit from the product, even though it may not currently be known to that person. The development of this solution therefore is of the utmost importance in playing a role in helping develop not only Africa, but any other country that may benefit from the use of the solution.

This dissertation in general, does not only provide proof that there are indeed alternatives to traditional methods of determining productivity measures, but also illustrates that the proposed method of determining productivity measures from business process images has the potential to be used as a stand-alone product, as well as a product integrated into current methods used in the business arena. The purpose of this dissertation was never to disprove traditional methods, but to provide research to prove that the possibility does exist for progress, as well as provide insight into an alternative method for determining productivity measures for the sake of future research and development if needed.

BIBLIOGRAPHY

Amberg, B. & Vetter, T., 2012. *GraphTrack: Fast and Globally Optimal Tracking in Videos.* Basel, University of Basel.

Awata, S. P., 2007. *Parzen-Window Density Estimation,* Utah: University of Utah.

Bordese, M. & Alini, W., 2013. *Package 'biOps'.* [Online]
Available at: http://cran.r-project.org/web/packages/biOps/biOps.pdf
[Accessed 20 July 2013].

Campbell, C., 2012. *BirminhamR: Good Looking Models.* [Online]
Available at: http://files.meetup.com/2906882/Good%20Looking%20Models.pdf
[Accessed 20 July 2013].

Charpentier, A., 2012. *Freakonometrics.* [Online]
Available at: http://freakonometrics.blog.free.fr/index.php?post/2012/04/18/foundwaldo
[Accessed 10 May 2013].

De Espindola, G. M. & Pantaleão, E., 2009. *Parzen windows and nonparametric density estimation applied in high resolution imagery classification.* Natal, Instituto Nacional de Pesquisas Espaciais – INPE.

Feigelson, E. D. & Babu, G. J., 2012. *Modern Statistical Methods for Astronomy: With "R" Applications.* 1st ed. New York: Cambridge University Press.

Frery, A. C. & Perciano, T., 2013. Introduction to Image Processing Using R. *SpringerBriefs in Computer Science,* Issue DOI: 10.1007/978-1-4471-4950-7_1.

Killeen, J., 2012. *Understanding your Business Environment,* Essex: Results Consortium Ltd..

Lafaye de Micheaux, P. & Marchini, J. L., 2013. *Package 'AnalyzeFMRI'.* [Online]
Available at: http://cran.r-project.org/web/packages/AnalyzeFMRI/AnalyzeFMRI.pdf
[Accessed 20 July 2013].

Mejail, M. E., Jacobo-Berlles, J., Frery, A. C. & Bustos, O. H., 2003. Classification of SAR images using a general and tractable multiplicative model. *International journal of remote sensing,* 24(18), p. 3565–3582.

Molina, R. & Ripley, B. D., 1993. Using spatial models as priors in astronomical image analysis. *Journal of Applied Statistics,* 20(5-6), pp. 281-298.

Pau, G. et al., 2010. EBImage - A "R" package for image processing with applications to cellular phenotypes.. *Bio-Informatics Application,* 26(7), pp. 979-981.

Pau, G. et al., 2013. *Package 'EBImage'.* [Online]
Available at: http://www.bioconductor.org/packages/2.12/bioc/manuals/EBImage/man/EBImage.pdf
[Accessed 20 July 2013].

Pau, G., Sklyar, O. & Huber, W., 2013. *Introduction to EBImage, an image processing and analysis toolkit for R.* [Online]
Available at: http://www.bioconductor.org/packages/2.12/bioc/vignettes/EBImage/inst/doc/EBImage-introduction.pdf
[Accessed 20 July 2013].

Porter, M. E., 2008. The Five Competitive Forces that Shape Strategy. *Harvard Business Review*, 1 January, pp. 79-93.

Prism Skylabs, 2013. *Prism Skylabs.* [Online]
Available at: https://www.prismskylabs.com
[Accessed 28 May 2013].

R Foundation, 2000. *R: Kernel Density Estimation.* [Online]
Available at: http://stat.ethz.ch/R-manual/R-patched/library/stats/html/density.html
[Accessed 26 September 2013].

Senior, A. W. et al., 2007. *Video analytics for retail.* London, IEEE.

Swedlow, J. R., Goldberg, I., Brauner, E. & Sorger, P. K., 2003. Informatics and Quantitative Analysis in Biological Imaging. *Science 300,* Issue 5616, pp. 100-102.

Tabelow, K. et al., 2008. *Image Analysis and Statistical Inference in Neuroimaging with R*, Vienna: R Development Core Team.

Tanaka, T., Shimada, A., Arita, D. & Taniguchi, R., 2007. *A Fast Algorithm for Adaptive Background Model Construction.* London, IEEE, pp. 528-533.

Van Assen, M., Van den Berg, G. & Pietersma, P., 2009. *Key Management Models: The 60+ Models Every Manager Needs to Know.* 2nd ed. London: Pearson Education.

Various Authors, 2005. *Proceedings 2nd Joint IEEE International Workshop on VS-PETS.* Beijing, IEEE.

Venables, W. M., Smith, D. M. & R Core Team, 2013. *An Introduction to R.* [Online]
Available at: http://www.r-project.org/
[Accessed 15 April 2013].

Wodehouse, A. J. & Ion, W. J., 2010. Information Use in Conceptual Design: Existing Taxonomies and New Approaches. *International Journal of Design*, 4(3), pp. 53-63.

Zucchini, W., 2003. *APPLIED SMOOTHING TECHNIQUES - Part 1: Kernel Density Estimation.* [Online]
Available at: http://isc.temple.edu/economics/Econ616/Kernel/ast_part1.pdf
[Accessed 26 09 2013].

APPENDICES

Image

A formal description of the objects which will be referred to as "images" in order to be able to deal with them in an unambiguous way, is given in the following excerpt, (Frery & Perciano, 2013):

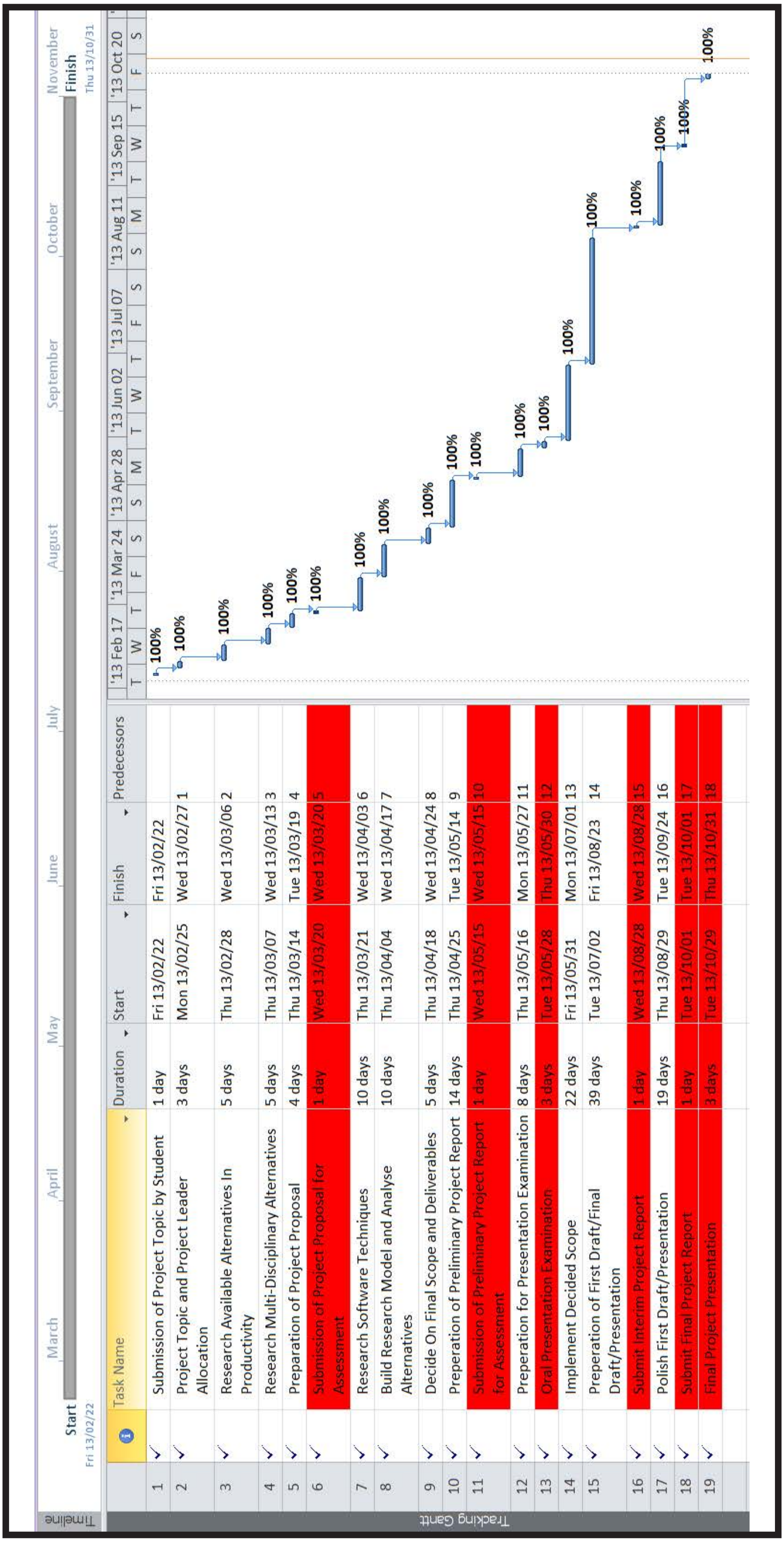
"Images are defined on a finite regular grid S of size $m \times n$. Denote this support $S = \{1, ..., m\} \times \{1, ..., n\} \subset N2$. Elements of S are called *sites*, *coordinates* or *positions*, and can be denoted by $(i, j) \in S$, where $1 \leq i \leq m$ is the row, and $1 \leq j \leq n$ is the column. If the coordinates are not necessary, a lexicographic order will be used to index an arbitrary element of S by $s \in \{1, ..., mn\}$. Using the convention employed by R for storing data in a matrix, $s = (n-1) j+i$, as illustrated in the following code where the entries of S, that has $m = 3$ lines and $n = 4$ columns, are the values of the lexicographic order.

An image is then a function of the form f: $S \rightarrow$ K, i.e., $f \in S$K, where K the set of possible values in each position. A *pixel* is a pair $(s, f(s))$. In many situations K has the form of a Cartesian product, i.e., K = $K\ p$, where p is called the "number of bands" and K is the set of elementary available values.

Images are subjected to storage constraints in practical applications, and, therefore, the set K is limited to computer words. Examples are the sets which can be described by one bit {0, 1} and eight bits {0, 1, . . . , 255}, either signed or unsigned integers, floating point and double precision values, and complex numbers in single and double precision.

An image is a mathematical object whose visualization has not been yet defined. A visualization function is an application v: SK $\rightarrow (M, C)$, where M is the physical area of the monitor and C is the set of colours available in that monitor for the software which is being employed (the *palette*). Given an image f, $v(f)$ is a set of spots in the monitor corresponding to each coordinate $s \in S$, where colours are drawn according to the values $f(s)$. It is, therefore clear that a binary image can be visualized in black and white, in red and blue, or in any pair of colours, even the same colour. In this last case, the information conveyed by the image is lost in the visualization process, but can be retrieved by switching to a different set of colours."

APPENDIX B: GANTT CHAR)2/11/2013)



	Task Name	Duration	Start	Finish	Predecessors
1	Submission of Project Topic by Student	1 day	Fri 13/02/22	Fri 13/02/22	
2	Project Topic and Project Leader Allocation	3 days	Mon 13/02/25	Wed 13/02/27	1
3	Research Available Alternatives In Productivity	5 days	Thu 13/02/28	Wed 13/03/06	2
4	Research Multi-Disciplinary Alternatives	5 days	Thu 13/03/07	Wed 13/03/13	3
5	Preparation of Project Proposal	4 days	Thu 13/03/14	Tue 13/03/19	4
6	Submission of Project Proposal for Assessment	1 day	Wed 13/03/20	Wed 13/03/20	5
7	Research Software Techniques	10 days	Thu 13/03/21	Wed 13/04/03	6
8	Build Research Model and Analyse Alternatives	10 days	Thu 13/04/04	Wed 13/04/17	7
9	Decide On Final Scope and Deliverables	5 days	Thu 13/04/18	Wed 13/04/24	8
10	Preparation of Preliminary Project Report	14 days	Thu 13/04/25	Tue 13/05/14	9
11	Submission of Preliminary Project Report for Assessment	1 day	Wed 13/05/15	Wed 13/05/15	10
12	Preparation for Presentation Examination	8 days	Thu 13/05/16	Mon 13/05/27	11
13	Oral Presentation Examination	3 days	Tue 13/05/28	Thu 13/05/30	12
14	Implement Decided Scope	22 days	Fri 13/05/31	Mon 13/07/01	13
15	Preperation of First Draft/Final Draft/Presentation	39 days	Tue 13/07/02	Fri 13/08/23	14
16	Submit Interim Project Report	1 day	Wed 13/08/28	Wed 13/08/28	15
17	Polish First Draft/Presentation	19 days	Thu 13/08/29	Tue 13/09/24	16
18	Submit Final Project Report	1 day	Tue 13/10/01	Tue 13/10/01	17
19	Final Project Presentation	3 days	Tue 13/10/29	Thu 13/10/31	18

APPENDIX C: INDUSTRY S FORM

Department of Industrial & Systems Engineering
Final Year Projects
Identification and Responsibility of Project Sponsors

All Final Year Projects are published by the University of Pretoria on *UPSpace* and thus freely available on the Internet. These publications portray the quality of education at the University and have the potential of exposing sensitive company information. It is important that both students and company representatives or sponsors are aware of such implications.

Key responsibilities of Project Sponsors:

A project sponsor is the key contact person within the company. This person should thus be able to provide the best guidance to the student on the project. The sponsor is also very likely to gain from the success of the project. The project sponsor has the following important responsibilities:

1. Confirm his/her role as project sponsor, duly authorised by the company. Multiple sponsors can be appointed, but this is not advised. The duly completed form will considered as acceptance of sponsor role.
2. Review and approve the Project Proposal, ensuring that it clearly defines the problem to be investigated by the student and that the project aim, scope, deliverables and approach is acceptable from the company's perspective.
3. Review the Final Project Report (delivered during the second semester), ensuring that information is accurate and that the solution addresses the problems and/or design requirements of the defined project.
4. Acknowledges the intended publication of the Project Report on UP Space.
5. Ensures that any sensitive, confidential information or intellectual property of the company is not disclosed in the Final Project Report.

Project Sponsor Details:

Company:	Sathu
Project Description:	
Student Name:	
Student number:	
Student Signature:	
Sponsor Name:	A van Rensburg
Designation:	Director
E-mail:	info@sathu.co.za
Tel No:	012 997 3791
Cell No:	083 2005693
Fax No:	
Sponsor Signature:	

The following documentation relates to the statistical model and package that was used in the proposed solution.

Kernel Density Estimation (R Foundation, 2000)

Description

The (S3) generic function density computes kernel density estimates. Its default method does so with the given kernel and bandwidth for univariate observations.

Usage

```
density(x, ...)
## Default S3 method:
density(x, bw = "nrd0", adjust = 1,
        kernel = c("gaussian", "epanechnikov", "rectangular",
                   "triangular", "biweight",
                   "cosine", "optcosine"),
        weights = NULL, window = kernel, width,
        give.Rkern = FALSE,
        n = 512, from, to, cut = 3, na.rm = FALSE, ...)
```

Arguments

x	the data from which the estimate is to be computed.
bw	the smoothing bandwidth to be used. The kernels are scaled such that this is the standard deviation of the smoothing kernel. (Note this differs from the reference books cited below, and from S-PLUS.) bw can also be a character string giving a rule to choose the bandwidth. See bw.nrd. The default, "nrd0", has remained the default for historical and compatibility reasons, rather than as a general recommendation, where e.g., "SJ" would rather fit, see also V&R (2002). The specified (or computed) value of bw is multiplied by adjust.
adjust	the bandwidth used is actually adjust*bw. This makes it easy to specify values like 'half the default' bandwidth.
kernel, window	a character string giving the smoothing kernel to be used. This must be one of "gaussian", "rectangular", "triangular", "epanechnikov", "biweight", "cosine" or "optcosine", with default "gaussian", and may be abbreviated to a unique prefix (single letter). "cosine" is smoother than "optcosine", which is the usual 'cosine' kernel in the literature and almost MSE-efficient. However, "cosine" is the version used by S.
weights	numeric vector of non-negative observation weights, hence of same length as x. The default NULL is equivalent to weights = rep(1/nx, nx) where nx is the length of (the finite entries of) x[].

width	this exists for compatibility with S; if given, and bw is not, will set bw to width if this is a character string, or to a kernel-dependent multiple of width if this is numeric.
give.Rkern	logical; if true, *no* density is estimated, and the 'canonical bandwidth' of the chosen kernel is returned instead.
n	the number of equally spaced points at which the density is to be estimated. When n > 512, it is rounded up to a power of 2 during the calculations (as fft is used) and the final result is interpolated by approx. So it almost always makes sense to specify n as a power of two.
from,to	the left and right-most points of the grid at which the density is to be estimated; the defaults are cut * bw outside of range(x).
cut	by default, the values of from and to are cut bandwidths beyond the extremes of the data. This allows the estimated density to drop to approximately zero at the extremes.
na.rm	logical; if TRUE, missing values are removed from x. If FALSE any missing values cause an error.
...	further arguments for (non-default) methods.

Details

The algorithm used in density.default disperses the mass of the empirical distribution function over a regular grid of at least 512 points and then uses the fast Fourier transform to convolve this approximation with a discretized version of the kernel and then uses linear approximation to evaluate the density at the specified points.

The statistical properties of a kernel are determined by *sig^2 (K) = int(t^2 K(t) dt)* which is always = *1* for our kernels (and hence the bandwidth bw is the standard deviation of the kernel) and *R(K) = int(K^2(t) dt)*.
MSE-equivalent bandwidths (for different kernels) are proportional to *sig(K) R(K)* which is scale invariant and for our kernels equal to *R(K)*. This value is returned when give.Rkern = TRUE. See the examples for using exact equivalent bandwidths.

Infinite values in x are assumed to correspond to a point mass at +/-Inf and the density estimate is of the sub-density on (-Inf, +Inf).

Value

If give.Rkern is true, the number *R(K)*, otherwise an object with class "density" that's underlying structure is a list containing the following components.

x	the n coordinates of the points where the density is estimated.
y	the estimated density values. These will be non-negative, but can be zero.
bw	the bandwidth used.

n	the sample size after elimination of missing values.
call	the call which produced the result.
data.name	the deparsed name of the x argument.
has.na	logical, for compatibility (always FALSE).

The print method reports summary values on the x and y components.

Introduction to EBImage,
an image processing and analysis toolkit for R

Gregoire Pau, Oleg Sklyar, Wolfgang Huber
gpau@ebi.ac.uk

April 18, 2013

Contents

1 Reading/displaying/writing images

The package EBImage is loaded by the following command.

```
> library("EBImage")
```

The function `readImage` is able to read images from files or URLs. Current supported image formats are JPEG, PNG and TIFF.

```
> f = system.file("images", "lena.png", package="EBImage")
> lena = readImage(f)
```

Images can be displayed using the function `display`. Pixel intensities should range from 0 (black) to 1 (white).

```
> display(lena)
```

Color images or images with multiple frames can also be read with `readImage`.

Figure 1: `lena`, `lenac`

```
> lenac = readImage(system.file("images", "lena-color.png", package="EBImage"))
> display(lenac)
> nuc = readImage(system.file('images', 'nuclei.tif', package='EBImage'))
> display(nuc)
```

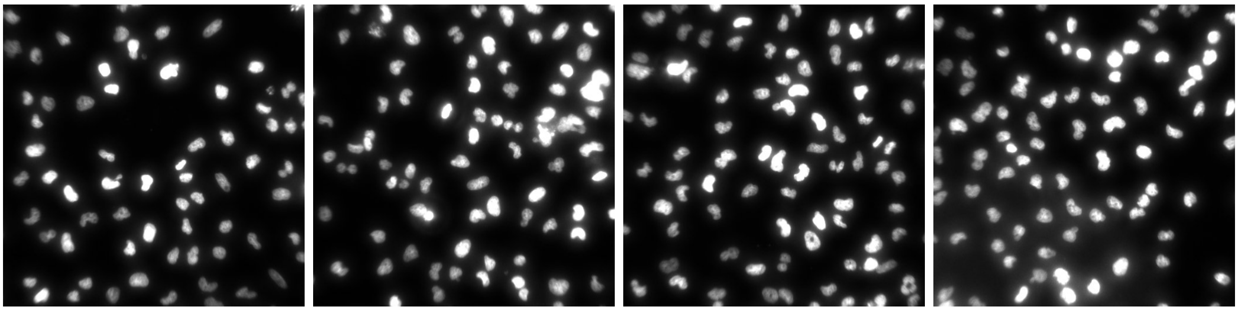


Figure 2: `nuc`

Images can be written with `writeImage`. The file format is deduced from the file name extension. This is useful to convert image formats, here from PNG format to JPEG format.

```
> writeImage(lena,  'lena.jpeg', quality=85)
> writeImage(lenac, 'lenac.jpeg', quality=85)
```

2 Image objects and matrices

The package EBImage uses the class `Image` to store and process images. Images are stored as multi-dimensional arrays containing the pixel intensities. All EBImage functions are also able to work with matrices and arrays.

```
> print(lena)

Image
  colormode: Grayscale
  storage.mode: double
  dim: 512 512
  nb.total.frames: 1
  nb.render.frames: 1

imageData(object)[1:5,1:6]:
          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[1,] 0.5372549 0.5372549 0.5372549 0.5372549 0.5372549 0.5490196
[2,] 0.5372549 0.5372549 0.5372549 0.5372549 0.5372549 0.5490196
[3,] 0.5372549 0.5372549 0.5372549 0.5372549 0.5372549 0.5137255
[4,] 0.5333333 0.5333333 0.5333333 0.5333333 0.5333333 0.5098039
[5,] 0.5411765 0.5411765 0.5411765 0.5411765 0.5411765 0.5333333
```

As matrices, images can be manipulated with all R mathematical operators. This includes `+` to control the brightness of an image, `*` to control the contrast of an image or `^` to control the gamma correction parameter.

```
> lena1 = lena+0.5
> lena2 = 3*lena
> lena3 = (0.2+lena)^3
```

Figure 3: `lena`, `lena1`, `lena2`, `lena3`

Others operators include `[` to crop images, `<` to threshold images or `t` to transpose images.

```
> lena4 = lena[299:376, 224:301]
> lena5 = lena>0.5
> lena6 = t(lena)
> print(median(lena))

[1] 0.3803922
```

Images with multiple frames are created using `combine` which merges images.

```
> lenacomb = combine(lena, lena*2, lena*3, lena*4)
> display(lenacomb)
```

Figure 4: `lena`, `lena4`, `lena5`, `lena6`

Figure 5: `lenacomb`

3 Spatial transformations

Specific spatial image transformations are done with the functions `resize`, `rotate`, `translate` and the functions `flip` and `flop` to reflect images.

```
> lena7 = rotate(lena, 30)
> lena8 = translate(lena, c(40, 70))
> lena9 = flip(lena)
```

Figure 6: `lena`, `lena7`, `lena8`, `lena9`

4 Color management

The class `Image` extends the base class `array` and uses the `colormode` slot to store how the color information of the multi-dimensional data should be handled.

As an example, the color image `lenac` is a 512x512x3 array, with a `colormode` slot equals to `Color`. The object is understood as a color image by EBImage functions.

```
> print(lenac)
```

```
Image
  colormode: Color
  storage.mode: double
  dim: 512 512 3
  nb.total.frames: 3
  nb.render.frames: 1

imageData(object)[1:5,1:6,1]:
          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[1,] 0.8862745 0.8862745 0.8862745 0.8862745 0.8862745 0.8901961
[2,] 0.8862745 0.8862745 0.8862745 0.8862745 0.8862745 0.8901961
[3,] 0.8745098 0.8745098 0.8745098 0.8745098 0.8745098 0.8901961
[4,] 0.8745098 0.8745098 0.8745098 0.8745098 0.8745098 0.8705882
[5,] 0.8862745 0.8862745 0.8862745 0.8862745 0.8862745 0.8862745
```

The function `colorMode` can access and change the value of the slot `colormode`, modifying the rendering mode of an image. In the next example, the `Color` image `lenac` with one frame is changed into a `Grayscale` image with 3 frames, corresponding to the red, green and blue channels. The function `colorMode` does not change the content of the image but changes only the way the image is rendered by EBImage.

```
> colorMode(lenac) = Grayscale
> display(lenac)
```

Figure 7: `lenac`, rendered as a `Color` image and as a `Grayscale` image with 3 frames (red channel, green channel, blue channel)

The color mode of image `lenac` is reverted back to `Color`.

```
> colorMode(lenac) = Color
```

The function `channel` performs colorspace conversion and can convert `Grayscale` images into `Color` ones both ways and can extract color channels from `Color` images. Unlike `colorMode`, `channel` changes the pixel intensity values of the image. The function `rgbImage` is able to combine 3 `Grayscale` images into a `Color` one.

```
> lenak = channel(lena, 'rgb')
> lenak[236:276, 106:146, 1] = 1
> lenak[236:276, 156:196, 2] = 1
> lenak[236:276, 206:246, 3] = 1
> lenab = rgbImage(red=lena, green=flip(lena), blue=flop(lena))
```

Figure 8: `lenak`, `lenab`

5 Image filtering

Images can be linearly filtered using `filter2`. `filter2` convolves the image with a matrix filter. Linear filtering is useful to perform low-pass filtering (to blur images, remove noise, ...) and high-pass filtering (to detect edges, sharpen images, ...). Various filter shapes can be generated using `makeBrush`.

```
> flo = makeBrush(21, shape='disc', step=FALSE)^2
> flo = flo/sum(flo)
> lenaflo = filter2(lenac, flo)
> fhi =  matrix(1, nc=3, nr=3)
> fhi[2,2] = -8
> lenafhi = filter2(lenac, fhi)
```

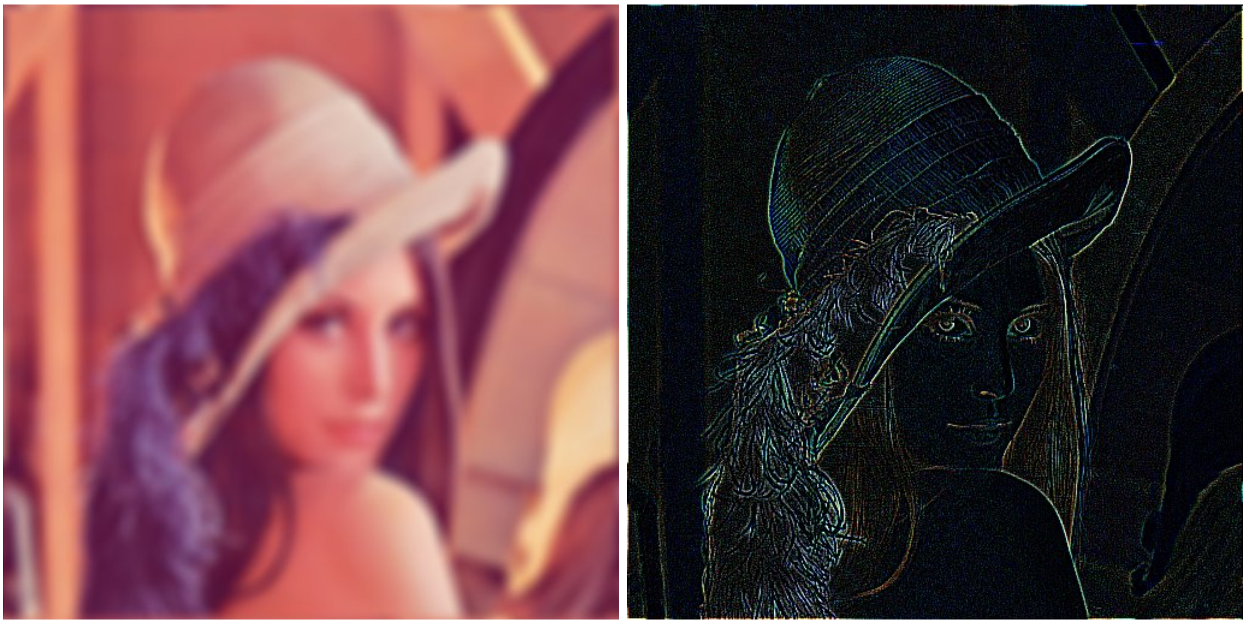


Figure 9: Low-pass filtered `lenaflo` and high-pass filtered `lenafhi`

6 Morphological operations

Binary images are images where the pixels of value 0 constitute the background and the other ones constitute the foreground. These images are subject to several non-linear mathematical operators called morphological operators, able to `erode` and `dilate` an image.

```
> ei = readImage(system.file('images', 'shapes.png', package='EBImage'))
> ei = ei[110:512,1:130]
> display(ei)
> kern = makeBrush(5, shape='diamond')
> eierode = erode(ei, kern)
> eidilat = dilate(ei, kern)
```

Figure 10: `ei` ; `eierode` ; `eidilat`

7 Segmentation

Segmentation consists in extracting objects from an image. The function `bwlabel` is a simple function able to extract every connected sets of pixels from an image and relabel these sets with a unique increasing integer. `bwlabel` can be used on binary images and is useful after thresholding.

```
> eilabel = bwlabel(ei)
> cat('Number of objects=', max(eilabel),'\n')

Number of objects= 7

> nuct = nuc[,,1]>0.2
> nuclabel = bwlabel(nuct)
> cat('Number of nuclei=', max(nuclabel),'\n')

Number of nuclei= 74
```

Figure 11: `ei`, `eilabel/max(eilabel)`

Since the images `eilabel` and `nuclabel` range from 0 to the number of object they contain (given by `max(eilabel)` and `max(nucabel)`), they have to be divided by these number before displaying, in order to fit the [0,1] range needed by `display`.

The grayscale top-bottom gradient observable in `eilabel` and `nuclabel` is due to the way `bwlabel` labels the connected sets, from top-left to bottom-right.

Adaptive thresholding consists in comparing the intensity of pixels with their neighbors, where the neighborhood is specified by a filter matrix. The function `thresh` performs a fast adaptive thresholding of an

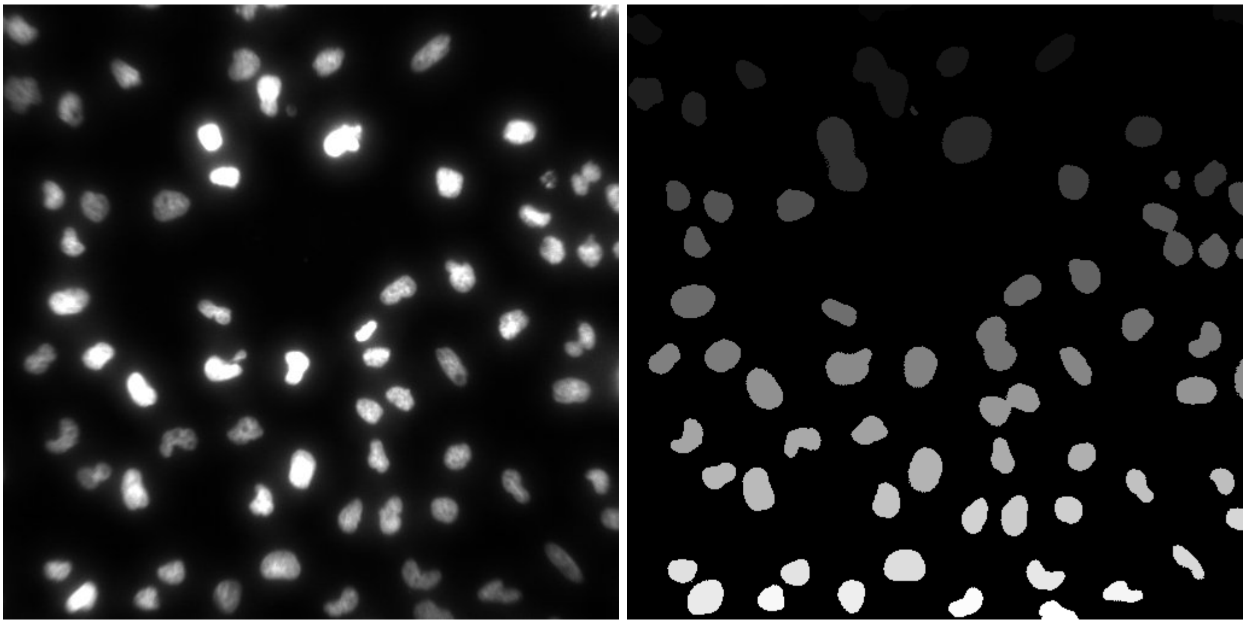


Figure 12: nuc[, ,1], nuclabel/max(nuclabel)

image with a rectangular window while the combination of filter2 and < allows a finer control. Adaptive thresholding allows a better segmentation when objects are close together.

```
> nuct2 =  thresh(nuc[,,1], w=10, h=10, offset=0.05)
> kern = makeBrush(5, shape='disc')
> nuct2 = dilate(erode(nuct2, kern), kern)
> nuclabel2 = bwlabel(nuct2)
> cat('Number of nuclei=', max(nuclabel2),'\n')

Number of nuclei= 76
```

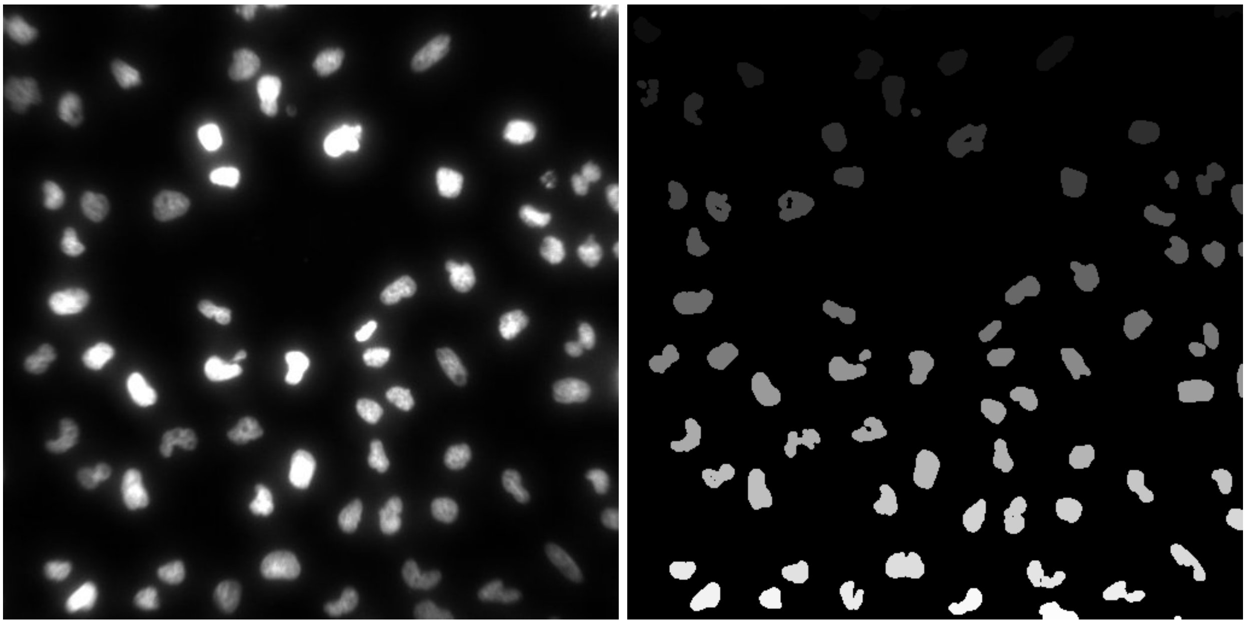


Figure 13: nuc[, ,1], nuclabel2/max(nuclabel)

8 Object manipulation

Objects, defined as sets of pixels with the same unique integer value can be outlined and painted using `paintObjects`. Some holes are present in objects of `nuclabel2` which can be filled using `fillHull`.

```
> nucgray = channel(nuc[,,1], 'rgb')
> nuch1 = paintObjects(nuclabel2, nucgray, col='#ff00ff')
> nuclabel3 = fillHull(nuclabel2)
> nuch2 = paintObjects(nuclabel3, nucgray, col='#ff00ff')
```

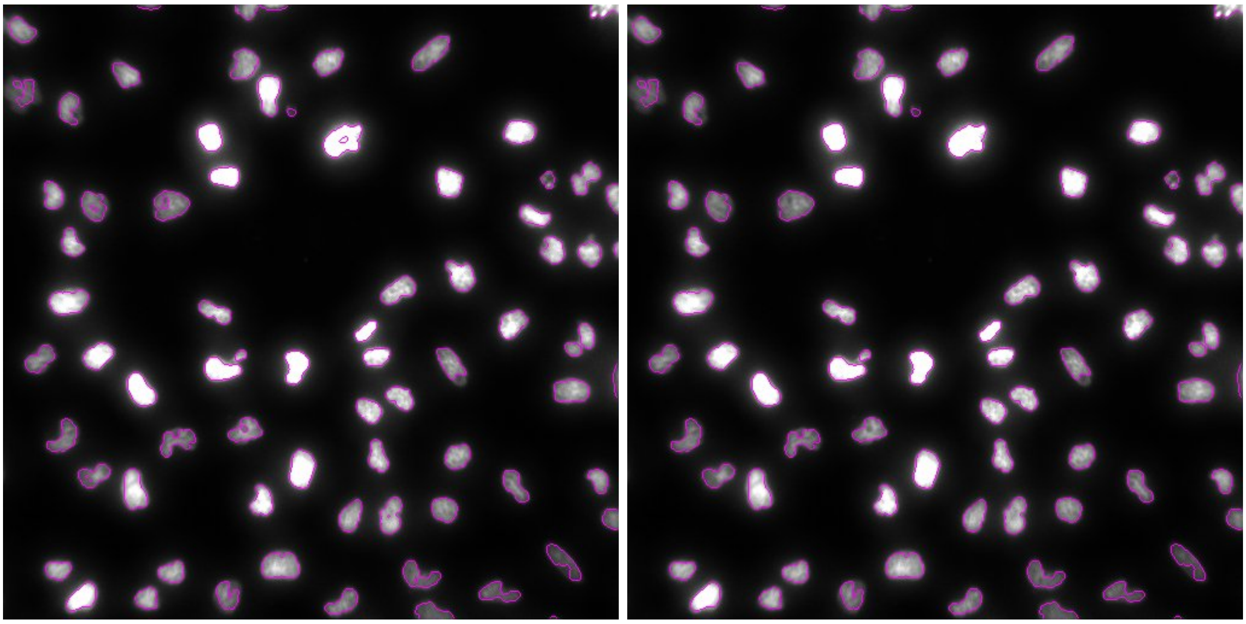


Figure 14: `nuch1`, `nuch2`

A broad variety of objects features (basic, image moments, shape, Haralick features) can be computed using `computeFeatures`. In particular, object coordinates are computed with the function `computeFeatures.moment`.

```
> xy = computeFeatures.moment(nuclabel3)[, c("m.cx", "m.cy")]
> xy[1:4,]

       m.cx      m.cy
1 121.74667  2.466667
2 210.19231  4.611888
3 497.44550  5.165877
4  15.99688 22.140187
```

9 Cell segmentation example

This is a complete example of segmentation of cells (nucleus + cell bodies) using the functions described before and the function `propagate`, able to perform Voronoi-based region segmentation.

Images of nuclei and cell bodies are first loaded:

```
> nuc = readImage(system.file('images', 'nuclei.tif', package='EBImage'))
> cel = readImage(system.file('images', 'cells.tif', package='EBImage'))
> img = rgbImage(green=1.5*cel, blue=nuc)
```

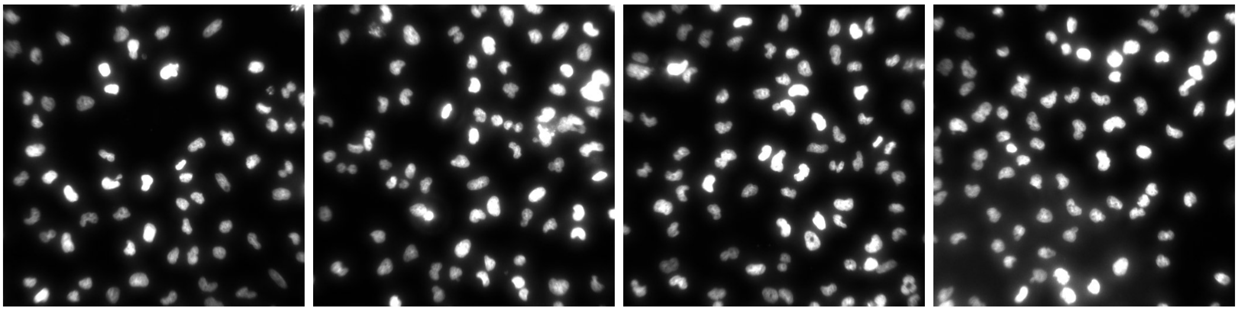


Figure 15: `nuc`

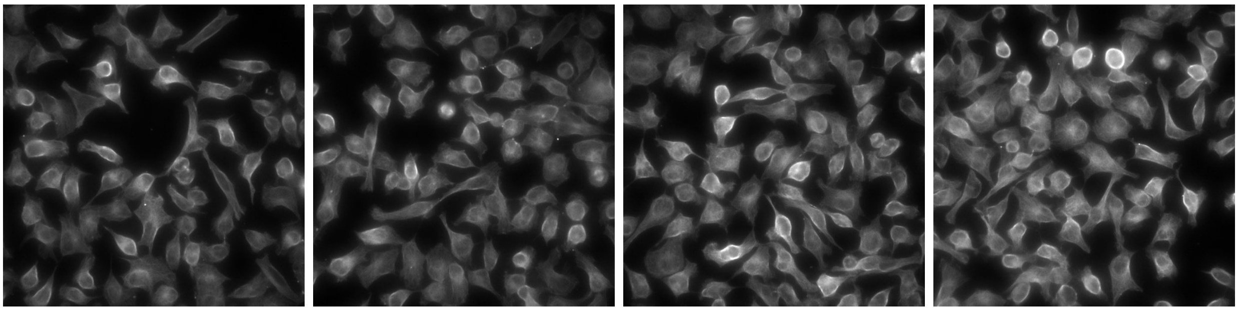


Figure 16: `cel`

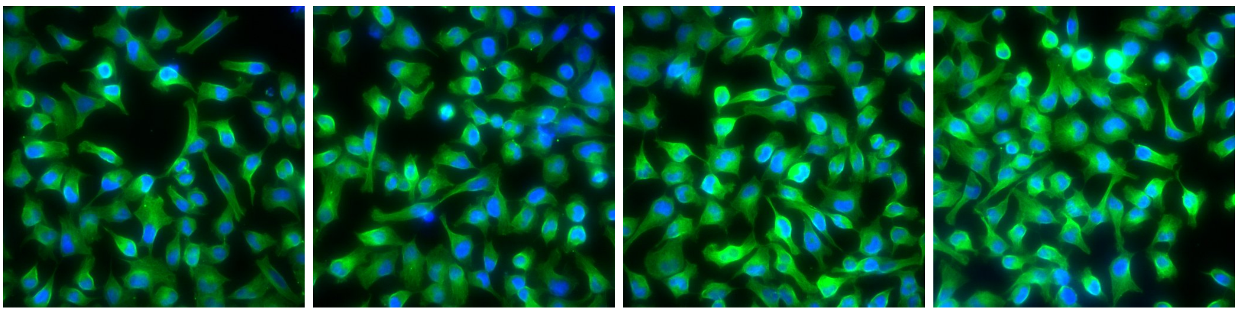


Figure 17: `img`

Nuclei are first segmented using `thresh`, `fillHull`, `bwlabel` and `opening`, which is an `erosion` followed by a `dilatation`.

```
> nmask = thresh(nuc, w=10, h=10, offset=0.05)
> nmask = opening(nmask, makeBrush(5, shape='disc'))
> nmask = fillHull(nmask)
> nmask = bwlabel(nmask)
```

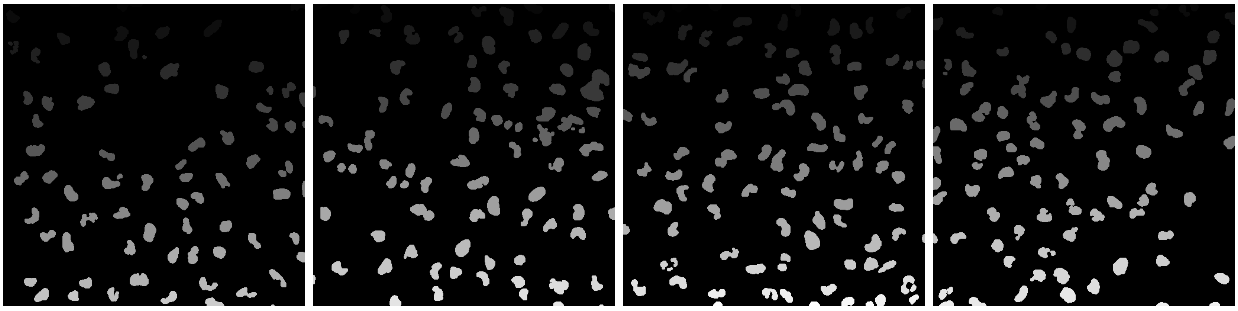


Figure 18: nmask/max(nmask)

Cell bodies are segmented using propagate.

```
> ctmask = opening(cel>0.1, makeBrush(5, shape='disc'))
> cmask = propagate(cel, seeds=nmask, mask=ctmask)
```

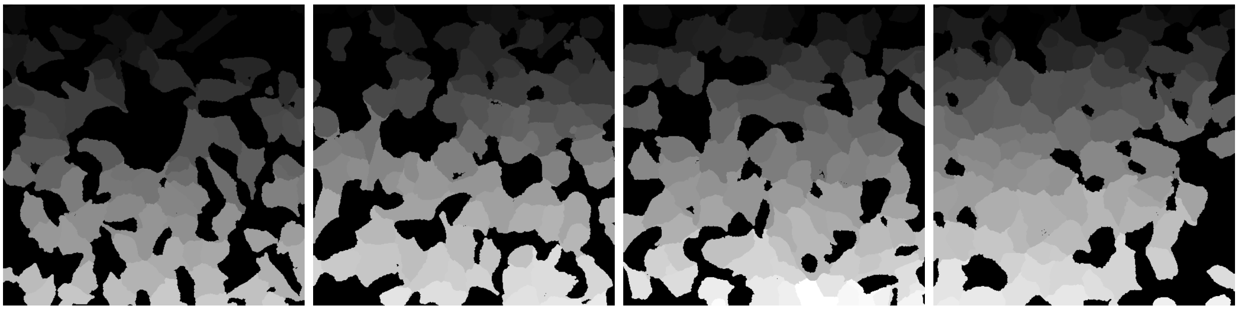


Figure 19: cmask/max(cmask)

Cells are outlined using paintObjects.

```
> res = paintObjects(cmask, img, col='#ff00ff')
> res = paintObjects(nmask, res, col='#ffff00')
```

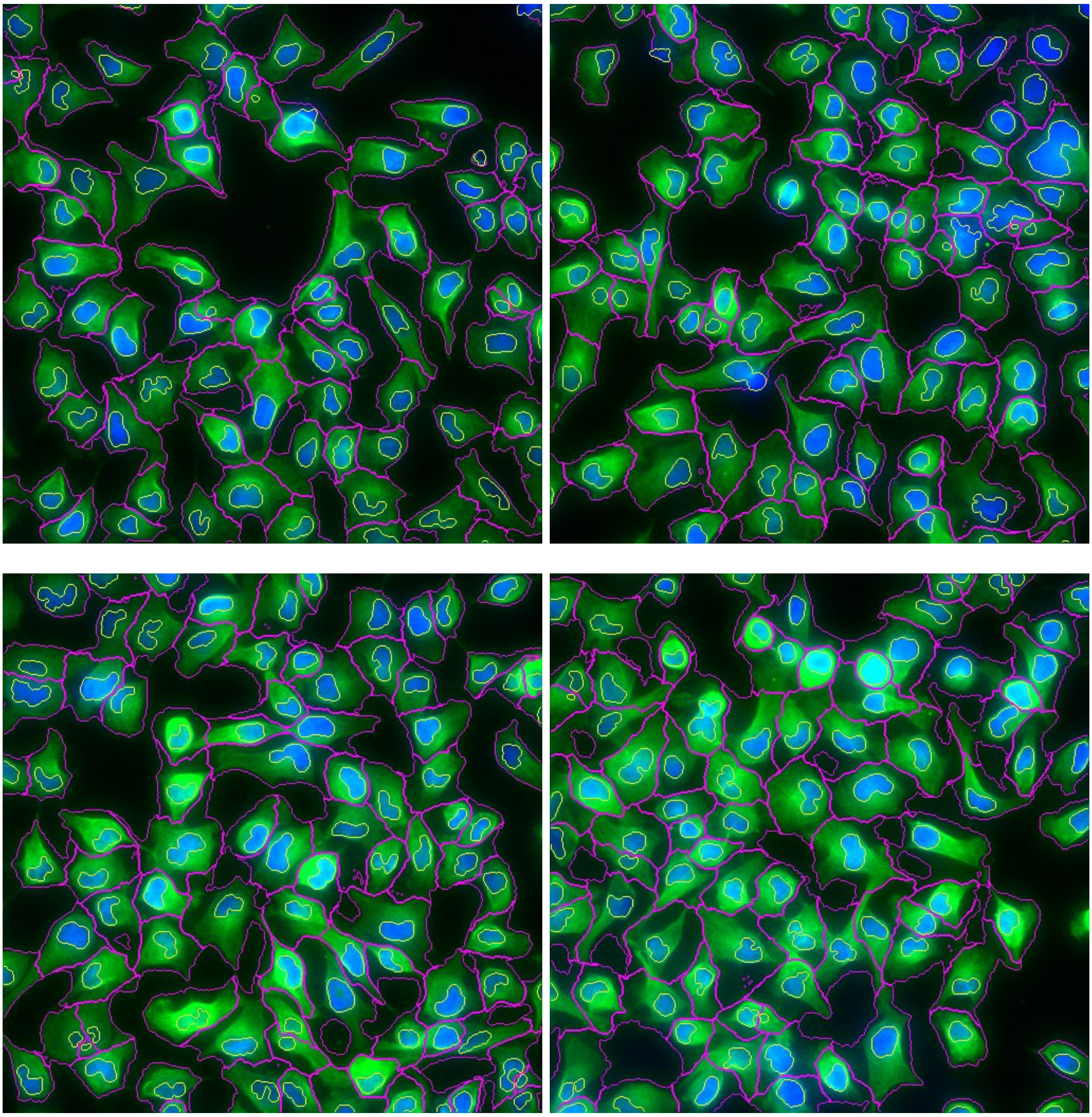


Figure 20: Final segmentation `res`

APPENDIX E: ANALYSIS TECHNIQUES EXPLAINED

(Killeen, 2012)

The ability to understand the Environment in which your business is operating is essential for business success.

In this article, Jim Killeen introduces a simple approach to ensuring that business owners and senior managers are aware of the key factors which must be taken into account when formulating business plans and business strategy.

Understanding your Business Environment

RESULTS

PEOPLE DEVELOPING PEOPLE

UNDERSTANDING YOUR OPERATING ENVIRONMENT

AN INTRODUCTION

Running a business is never easy. There always seems to be something happening that means you need to fit more into your day than you think you can handle. But somehow you do handle it. You keep going.
Not only that, but a large proportion of the things you have to handle seem to be things that you can't control – you have to react to the situation.
If this is you, than you have both my sympathy and my admiration.

That doesn't mean that it has to carry on though. It's absolutely true that some things to do with running a business are outside our control – we have to respond to external events. **But** some things we do have complete control over. And some things we have some influence over.
This article is designed to help you to come to terms with these realities. Also, you are not alone. Others have thought about this as well, including some of the most respected and well informed business leaders and business academics.

Don't worry. This is not a course and there are no exams. Don't let the word *academic* frighten you. The models and ways of thinking I'm going to introduce are not there for their own sake. They are there to help you think about ways of planning, organising and carrying out the activities of the business in order to be more effective and more efficient.
By effective I mean better at achieving the goals of the business more often. By efficient I mean doing this with the minimum of time, effort and money.
So it doesn't matter if you can't remember the names of this stuff. So long as the ways of thinking that I'm going to introduce help you to think about your business more rationally, and hopefully, successfully. Ultimately, this article is about being more informed when you construct or review your Business Plan.

Remember I said that some things are outside your control, some things are under your control and some things you can influence?
We can call that the 3 Environments Model of Business.

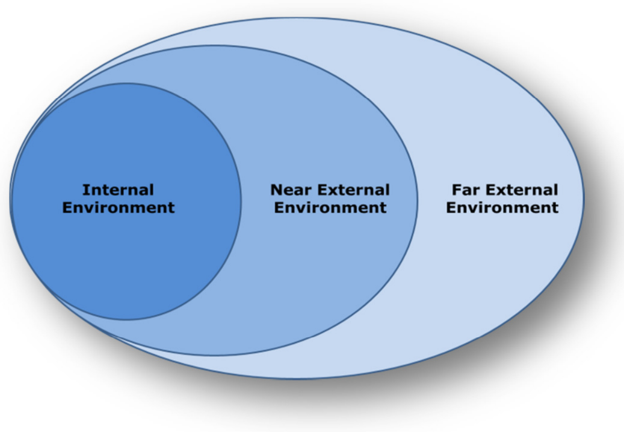


Figure 1 - The 3 Environments Model

Every business, every organisation, operates in these 3 environments, all at the same time. All the time, they are influencing the way you do business, and influencing the success of the business. So it makes sense that if we can understand these 3 Environments, we might have a better chance at being successful as a business.

The Far External Environment is the one we have No Control Over. We can only Respond to developments in the Far External Environment.

The Near External Environment is the one we can Influence. It exerts pressure on our business, but we can push back. (Sometimes more easily than others).

The Internal Environment is the one we can Control. This is our business itself. This includes our goals, how we aim to achieve these goals, and the way we set up the business – how we operate.

For each of these, there is a handy dandy way of analysing what's happening.

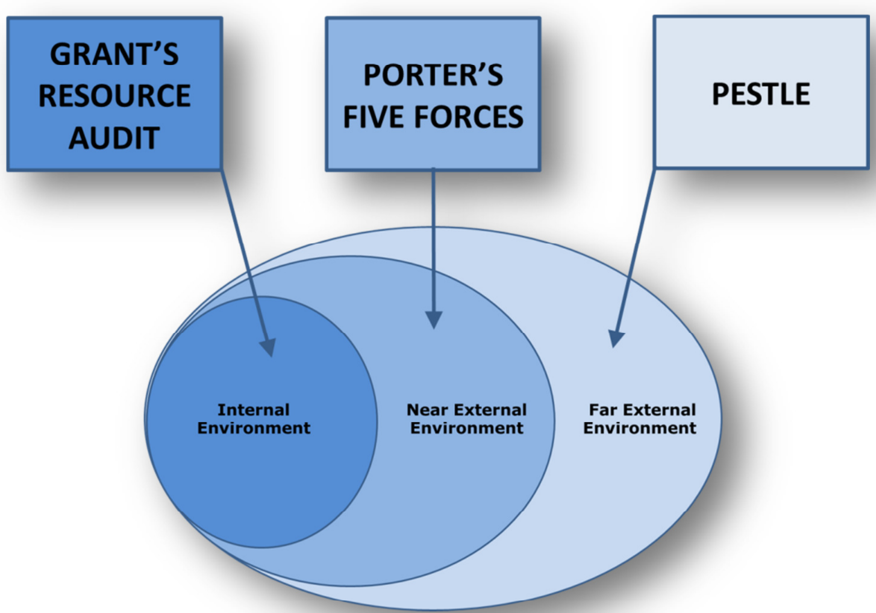
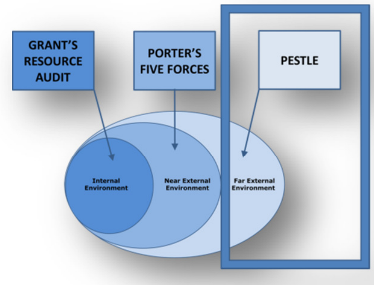


Figure 2 - The 3 Environments Analysis

In the following chapters, I will be going through each of these 3 Environments, and looking to see if we can find a way to understand them, and so be more able to do business within them. The important thing is to understand that every business is subject to influencing factors in its success. The more successful companies handle these influences better than others.
So whether you are looking to start up a business or are in the early years, if you are a manager interested in Management Training or a Director interested in Leadership Training, or indeed just looking for some business advice, I hope you find this article useful.

UNDERSTANDING THE FAR EXTERNAL ENVIRONMENT



In the previous chapter, I talked about the 3 Environments that affect a business and its chances of success.
In this chapter I'll be looking at the first – The Far External Environment.
For this, I'll be using an analysis model called PESTLE.

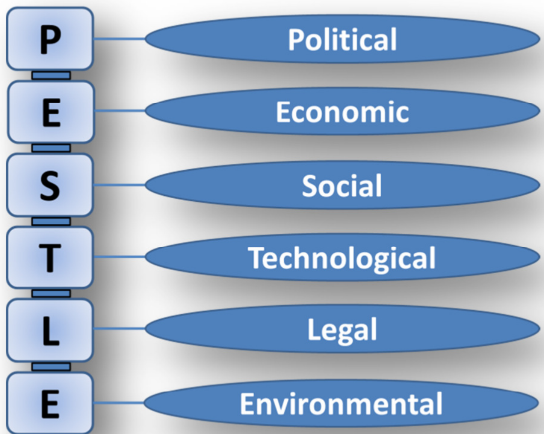


Figure 3 - PESTLE Analysis

The reason it's called PESTLE is because, as you can probably see, the first letters of the Key Elements that you have to think about spell out PESTLE
So, we'll go through them one by one. For each one, we look for the factors that may have an effect on us and our business, either now or in the future, but which we cannot control.
Usually, we know what's affecting us now, so it's important to see what changes may occur that will affect us in the future.
For this, we use Environmental Scanning. It sound very technical, but what it really means is reading newspapers, watching the news and trying to get a feel of changes that are occurring or are likely to occur that might affect us.

POLITICAL
This includes Political Parties coming to power – changes in Government. Changes to Political Parties in Government generally mean changes in Policy.
For example, cast your mind back to 2010, with a general election coming up. If you are a company that relies on Government funded contracts, you might be thinking "If the Conservatives get in, they might reduce funding, because they have been saying that the Government needs to reduce the amount it spends each year". So, thinking this might happen, you could start to plan what you would do.

ECONOMIC
This deals primarily with the overall state of the economy. At the moment the economy isn't doing too well. Most analysts believe it may be some time – years – before the economy picks up. So if you are a company that was thinking of spending a large "Launch Fund" on launching a range of luxury items, perhaps you might wait to see what happens. On the other hand, you might still want to launch the range any way, but at slightly lower prices, believing that people want a treat and you want to be the "Affordable Treat" provider.

SOCIAL

This deals with social trends – of many different types.
Say you run a Hairdressing Salon. Perhaps you have noticed that the local college is expanding, and its 100 yards away from your shop. The next intake is next month. So you approach the college and ask to have a spot during "Freshers Week" (The first week of the term when the new students find their feet). This way, you can make sure that they all know you are local, and have got really good rates for students. And so you increase your turnover. (No matter what you hear, students ARE poor and DO get their hair cut – you might just have found a very large new market)

TECHNOLOGICAL

This deals with any new technologies, and how they might be used to help your business. Or it might be becoming aware of threats to your business.
Things like the internet, which allow people to buy without having to go into a shop. If you decide to sell on the Internet it might help you. If you don't, it might hurt you.
Cast your mind back to the 90's. Bookshops generally just did not see AMAZON coming – until it was too late. Bookshops seemed to be convinced that people wanted to browse. They did want to browse – so Amazon included this feature in their Web Site.

LEGAL

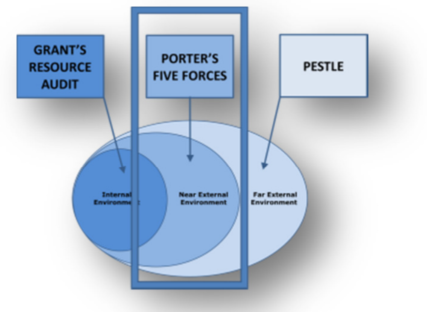
An offshoot of POLITICAL, these are changes in Law that might affect your business. Generally, we know about changes in Law that are coming into force before they actually do – that's how it works with Parliamentary Law making. So, knowing this, we can be prepared for it.

ENVIRONMENTAL

These are the things to do with what most people think of as Nature. So changes in the environment (nature) and the associated implications of these changes. For example, the growth of "green" technologies and products, such as organic foods.

With all of these, the key is to think how the factors out there in the work might affect our business. Not all of them will always apply at any given time to any given business. What's important is that you are generally aware of trends in these areas, how these trends might affect your business, how you plan to deal with them.

UNDERSTANDING THE NEAR EXTERNAL ENVIRONMENT



In previous chapters, I have introduced the idea that Businesses operate in Three Environments – The Far External, The Near External, and the Internal.

Here, I am going to examine the Near External Environment, and how it affects our ability to be successful in business. For this I will be using a very famous model (in business academic circles anyway) proposed by Michael Porter* – The Five Forces Model.

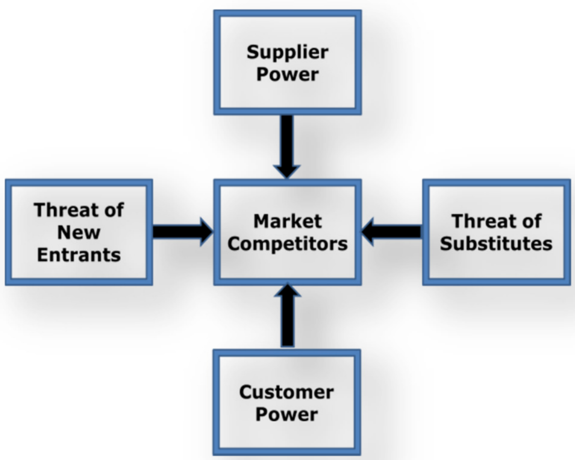


Figure 4 - Porter's Five Forces

Porter believes that for any business or organisation, there are five main factors that put pressure on a business and affect its likelihood of success. These are:

- Market Competition,
- Customer Power,
- Supplier Power,
- Threat of new Entrants,
- Threat of Substitutes.

We can go through each one of these in turn, starting with the most obvious –

Market Competition.

This relates to the competition you are up against to obtain and retain customers for your product or service. So if you are a hairdresser, your competition is other hairdressers. Porter believes that the key to surviving this competition is to identify, develop and maintain Competitive Advantage. This is an aspect of your product or service offering that is difficult for the competition to replicate or acquire, and which makes you more attractive to customers. The possibilities for achieving Competitive Advantage are dependent on the market you are in, and the imagination of the company. In short it is the thing or things that you do differently or better when compared to the competition. For example, what do you think the Competitive Advantage of Ferrari is. Or Porsche? To go back to the Hairdresser example – what might their advantage be? Perhaps it is that all of their stylists trained at Vidal Sassoon – and no one else in the area has those types of stylists. Or perhaps it's that they provide a crèche for children while parents are being seen to.

Whatever your industry or market, it is very helpful to identify something that you can do that is different or better than your competitors, and which (very importantly) will attract and help to retain customers.

Customer Power.

This relates to the size of the company in relation to the size of the market. It also relates to the ability of the customer to switch from one company to another to acquire the thing they want. Consider Supermarkets. In terms of size, Tesco's is very large and takes up a large part of the market. It is therefore less worried about ONE customer leaving and going somewhere else. However, it is very easy for customers to switch. So to be successful, Tesco has to try and decide how it is going to try to satisfy a large number of different customers with different needs and wants who can switch relatively easily – customers with little loyalty to the firm. So they sell lots of variety, both of product types and product identities (Brands).
On the other hand consider Rolls Royce Engines, who make engines for aeroplanes. There are very few companies who do this, but also very few customers (aeroplane builders). So Rolls Royce engines make few products, but all the products are very carefully tailored to suit the very few customers who order them.

Supplier Power.

Like Customer power in reverse, this tries to analyse the power of suppliers over the customer. For example, if you are the only petrol Station on the M1 within 50 miles all round, you can pretty much dictate prices (within reason). So if you are the only Train Operator between A and B, you can control pricing and, since the customer has little choice but to use you if they want to take the train, you also dictate service levels. But of course, no Government would allow that, would they?

Threat of New Entrants

While you may know how many competitors you have, and who they are, you also need to be aware of how likely it is that someone else will enter the market.
The easiest way to think about this is: remember how easy or difficult it was for you to enter the market and start trading. That's how easy it's going to be for someone else. It depends on how much money it takes to set up, any special technology that must be mastered, any special licenses that are needed, and so on. So entering the market of window cleaning is fairly easy, entering the market of making aeroplanes is very difficult.

Finally, the Threat of Substitutes.

Often overlooked, even by leading business figures, it is very important in business. Basically, in order to understand this threat – and it is a big one, you need to really understand the Need or Want that you are satisfying for your customers. Then ask yourself: How could my customers satisfy this need or want without my product service offering or one like it.

For example, Coke – its main competitor is Pepsi, but that is not a substitute, it's a competitor. In order to understand substitutes we need to understand Need or Want. So for Coke, its customers want something that will ease their thirst. Tap water does that, so tap water is a substitute product.

Now consider cars. Suppose you are Ford. Your Competitors are Volkswagen, Vauxhall, Toyota, etc. But they are not your substitute threat. If you ask yourself: "What is the need or want that is satisfied by Ford?" We can say "A means of getting from A to B". Buses do that, so do bikes, so do trains, so do aeroplanes. So these are the substitute threats.
Now, you might reply to the question of need or want: "A means of getting into work, so that I can sit at a computer all day and do paperwork". In other words "A way of me being able to work". But computers do that. Hence, why so many people work from home. They do not need a car because they do not need to travel. So the Internet is a Substitute Threat for cars in meeting this customer need.

The lesson here is that advances in technology often allow substitutes to take over from established markets. When Henry Ford started making cars for the masses, the established market was horses. The substitute was cars.
The same with typewriters. It used to be if you wanted to send an official letter, you typed it using metal keys and ink, onto paper and sent by post. Now, there is no need for ink or paper. We still type, but the product is not physical– the substitute for typewriters was a computer.
Remember Amazon from a previous chapter? – they are a substitute for Bookshops. Kindle is a substitute for Book Printers.
Consider a Barber. He cuts hair. For men. This is a service. 40 years ago there were very few substitute threats. Now, I can buy a home hair trimmer for the price of a decent haircut, and give myself a number 3 every month. No barber. A product has replaced a service.
Artists replaced by photographers and cameras and photographic film. Kodak Film replaced by digital cameras. Digital cameras replaced by mobile phones. Mobile phones replaced by?????

The reason I am spending so long on this one is because it is so important. While businesses often spend time deciding who their competition is and how they are going to out-compete them, they can often miss the substitute offering. If you can work out what the substitute might be, perhaps you can get in on the action. This aspect of business training – to look beyond what is immediately obvious – is essential and often overlooked.

Now some of you might be saying – "That's all very well and good, but it's based on hindsight, seeing something after the fact". Except that businesses can do it and do it successfully. Kindle is replacing books. That might be a problem for Amazon who sell books, except that Amazon and Kindle are one and the same. Amazon saw the technological advances, saw the threat of electronic books as a substitute for real ones, and saw an opportunity to get in on the ground floor. Not only that, but you don't need warehouses to store electronic books, which makes Amazon's Kindle e-book operation more profitable AND cheaper to customers. Everyone wins – except people who print books.

*Michael Porter is a Professor at Harvard University Business School. A leading business academic and consultant to large companies, he has written several books on Business and is highly regarded in his field.