Convergence Performance of Information Theoretic Similarity Measures for Robust Matching

Multimodality in The Medical Field

APPHIAH JUSTICE KABATA
Convergence Performance of Information Theoretic Similarity Measures for Robust Image Matching

Multimodal Imaging in The Medical Field

Apphiah Justice Kabata

Degree Project in Computer Science, DD143X
Supervisor: Atsuto Maki
Examiner: Örjan Ekeberg

CSC, KTH 2016-05-11
Acronyms

CCRE  cross-cumulative residual entropy
CT    computed tomography
MI    mutual information
MRI   magnetic resonance image
SAD   sum-of-absolute difference
SCV   sum of conditional variances
SSD   sum-of-squared difference
Acknowledgement

I would like to express my deepest appreciation to my supervisor Mr. Atsuto maki, professor at the Royal institute of Technology (KTH). His expertise, support and encouragement were essential for the completion of this thesis. My appreciation also goes to my father Balion, without whose love and morale support I will not have completed this thesis. My gratitude also goes to my friends Jamila and Chaimae for their support and presence during long experiment hours.

Apphiah
Abstract

Image matching is an area of significant use in the medical field, as there is a need to match images captured with different modalities, to overcome the limitation that can occur when dealing with individual modalities. However, performing matching of multimodal images may not be a trivial task. Multimodality entails changes in brightness and contrast that might be an obstacle when performing a match using similarity measures.

This study investigated the convergence performance of information-theoretic similarity measures. The similarity measures analysed in this study are mutual information (MI), the cross-cumulative residual entropy (CCRE), and the sum of conditional variances (SCV). To analyse the convergence performance of these measures, an experiment was conducted on one data set introducing the concept of multimodality, and two single images displaying a significant variation in texture. This was to investigate the impact of multimodality and variations in texture on the convergence performance of similarity measures. The experiment investigated the ability for similarity measures to find convergence on MRI and CT medical images after a displacement has occurred.

The results of the experiment showed that the convergence performance of similarity measures varies depending on the texture on images. MI is best suitable in the context of high-textured images while CCRE is more applicable in low-textured images. The measure SCV is the most stable similarity measure as it is little affected by the variation in texture. The experiment also reveals that the convergence performance of the similarity measures identified in the case of unimodality, can be preserved in the context of multimodality.

This study gives better awareness of the convergence performance of similarity measures. This could improve the use of similarity measures in the medical field which could yield better diagnosis of patients’ conditions.
Abstrakt

Matching av bilder har en stor betydelse inom det medicinska området. Detta eftersom det finns ett behov av att matcha bilder som tagits med olika modaliteter för att övervinna begränsningarna som kan uppkomma när man arbetar med en modalitet. Dock är det inte alltid trivialt att utföra matchning av multimodala bilder. Multimodalitet medför förändringar i ljussstyrka och kontrast som kan vara ett hinder vid utförande av matchning med likhetsmått.

Denna rapport undersökte konvergens prestanda av informationsteoretiska likhetsmått. Likhetsmätten som behandlades i denna studie var mutual information (MI), cross-cumulative residual entropy (CCRE), och the sum of conditional variances (SCV). För att analysera konvergens perstandat av likhetsmätten, utfördes ett experiment på en dataset som introducerar multimodalitet, och två enskilda bilder som har en betydande variation i struktur. Detta för att undersöka om konvergens av likhetsmätten påverkas av multimodalitet och variation i struktur. Experimentet fokuserades till att undersöka möjligheten för de olika likhetsmätten att hitta konvergens på MRI och CT medicinsiska bilder, efter att en manipulering av deras positioner har skett.

Resultaten av denna studie visade att konvergens prestanda av likhetsmätten varierade beroende på mängd struktur i datat. MI visade sig vara bäst lämpad i samband med tydlig struktur medan CCRE var mer lämplig när strukturen var mindre. SCV visade sig vara det mest stabila likhetsmåttet eftersom det inte påverkades märkbart av struktur variation. Studien visade också att konvergens beteendet av likhetsmätten som identifierades i samband med unimodalitet, kunde bevaras i multimodalitet.

Denna studie ger ökad medvetenhet om konvergens prestanda hos likhetsmätten. Detta kan leda till en förbättrad användning av likhetsmätten inom det medicinska området vilket kan ge möjligheter till att ställa bättre diagnoser av patienter.
# Table of contents

1. Introduction ................................................................. 7  
   1.1. Problem definition ...................................................... 8  
   1.2. Scope and constraints .................................................. 8  
   1.3. Literature study ....................................................... 8  
   1.4. Outline ................................................................... 9  

2. Background ........................................................................ 10  
   2.1. Image matching ........................................................... 10  
      2.1.1. Area-based methods .............................................. 10  
      2.1.2. Feature-based methods ......................................... 10  
   2.2. State of art of similarity measures ................................. 11  
      2.2.1. Mutual information (MI) ....................................... 11  
      2.2.2. Cross-cumulative residual entropy (CCRE) ............... 12  
      2.2.3. Sum of conditional variances ................................. 13  
   2.3. Multimodality ............................................................. 13  
      2.3.1. Morphological methods ....................................... 14  
      2.3.2. Functional methods ............................................ 15  
   2.4. Medical imaging ......................................................... 15  
      2.4.1. Magnetic resonance imaging (MRI) ......................... 15  
      2.4.2. Computed tomography (CT) .................................. 16  

3. Method ............................................................................ 17  
   3.1. Data collection ........................................................... 17  
      3.1.1. Data acknowledgement ......................................... 18  
   3.2. Technical approach ....................................................... 18  
      3.2.1. Matlab .............................................................. 18  
      3.2.2. Experiment ....................................................... 18  
         3.2.2.1 Gradient descent approach .............................. 19  
         3.2.2.2 Simulation .................................................. 19  

4. Results ............................................................................. 21  
   4.1. Simulation on identical data .......................................... 21  
      4.1.1. High-textured data .............................................. 21  
      4.1.2. Low-textured data .............................................. 22  
   4.2. Simulation on multimodal data ........................................ 23
4.3. Simulation on negated multimodal data ................................. 24

5. Discussion ................................................................................. 26
   5.1. Note on the method ............................................................... 26
   5.2. Note on the data ................................................................. 26
   5.3. Result analysis ................................................................. 27
       5.3.1 Variation in texture .................................................. 27
       5.3.2 Multimodality ......................................................... 28
       5.3.2.1 Misalignment .................................................... 29
       5.3.3 Negation ................................................................. 30
       5.3.3.1 Negation for high-textured data ......................... 30
       5.3.3.2 Negation for low-textured data ......................... 31
       5.3.3.3 General comments on negation ......................... 31

6. Conclusion ............................................................................. 33

7. References ............................................................................. 34
1. Introduction

Image matching is an area of great significance as there are several fields of application such as face detection. Image matching is also used as a tool for image-guided surgery. Image matching is necessary for computer vision as there is a need to map images to each other based on, for instance, features for matchmaking of similarities [1]. Image matching is used to determine if images are similar, independent of differences that might occur due to various types of disturbances. There are different types of measures for image matching depending on situational and contextual factors, where the disturbances are taken into consideration. All of them are referred to as similarity measures (For more information about computer vision see reference [34]).

There are several definitions of similarity measures used extensively depending on the context. One severally accepted definition is that a similarity measure can be regarded as a tool to quantify the dependency or independency between images [2]. Similarity measures are not always robust to transformation and noise in images. However, depending on the algorithm that each similarity measure uses, they provide different ways of dealing with outliers, noise, occlusions and other perturbations in data. Considerable efforts have been made to find the appropriate similarity measures for matching images. Information theoretic similarity measures are those similarity measures that make use of a statistical dependency of intensity between corresponding pixels [33]. Information theoretic measures are likely to be appropriate when handling multimodal images, because of their strong mathematical foundation and ability to detect non-linear changes, thus enabling matching of multimodal images (see section 2.3). The concept of information theoretic similarity measures was introduced by the use of mutual information (MI) as a similarity measure, in which emphasis in measurement is put on statistical dependence between variables. In an attempt at making similarity measures more robust, other information theoretic measures have been proposed such as cross-cumulative residual entropy (CCRE) and the sum of conditional variances (SCV).

Within information theoretic similarity measures, the concept of multimodality refers to images captured using different techniques or modes. The problem of establishing correspondences between two or more images is fundamental in computer vision. Images acquired through similar modes can be realigned by comparing the intensities in regions on images [3]. However, the technique of comparing intensities is challenging when images are from different modes. This is related to the fact that different modes may produce non-linear changes in brightness and contrast on images, which make it impossible to compare intensity regions. The difficulty in matching images from different modalities affects the medical field where there is a need to realign images from several medical modalities, for an accurate junction of complementary information for diagnostics.

In the medical field, many types of images are used to investigate, analyse and understand different anatomical and functional parts of the human body; this is referred to as medical
imaging [4]. These images are used to diagnose, monitor and treat various medical conditions. The aim of medical imaging is to reveal an object or a condition within the human body invisible to the observer [5]. Clinicians can, but have a limited view of the human body with visible light, because visible light has limited ability to penetrate tissues of the body. Due to that limitation, technologies used in medical imaging must make use of energy within the electromagnetic spectrum, to penetrate tissues and unveil precise conditions and objects. Different modes can be used to capture an image. These modes vary in the type of energy they use and reveal distinctive information about the human body. The modes are referred to as modality, and each modality has its application in the field [6]. Some medical images are obtained from magnetic resonance imaging (MRI), computer tomography (CT) and X-radiation (X-ray).

1.1. Problem definition

The focus of this study is to analyse the convergence performance of information theoretic similarity measures, namely CCRE, MI, and SCV. The study, therefore, aims to investigate the differences in convergence at variance of measurements, in the circumstances of the non-linear change of brightness and contrast emerging from the use of multimodality. In order words, this study will identify the convergence performance of similarity measures when multimodality is involved in medical imaging. It will also investigate how variations in texture on MRI and CT medical images, can affect the convergence of the similarity measures.

1.2. Scope and Constraints

The investigation concentrates on exposing the convergence performance of three different similarity measures with an information theoretic approach: MI, CCRE and SCV. These measures will be analyzed in the context of image matching concerning transformations arising from non-linear changes of intensity, that is, changes in brightness and contrast. The study will also investigate the impact of texture on the convergence performance of similarity measure. The images to be matched will be medical images, where, in this case, MRI/CT are used as capturing techniques.

1.3. Literature study

To gain understanding and information about the area of study, a literature study was conducted. The sources used for this report consisted of mainly previous research made in the field, reports of researchers within the area and websites with reliable information. The aim of the literature study was to gain valuable information and facts to create an accurate background for this report. The focus was placed on information theoretic similarity measures, but the literature research also included the optimisation algorithm and the programming tool used during the experiment.

To get relevant sources, keywords of focus were image matching/registration, medical imaging, information theoretic similarity measures, multimodal imaging and such in combination with each other.
1.4. Outline

This study investigates the convergence performance of the different similarity measures in different contexts. Initially, the major matching techniques are presented along with the different similarity measures used in this study. In the method section, the data used are accounted for, and the experiment explained. Results are presented in two contexts: multimodality, and variation in texture. In the discussion section, the results are illustrated by diagrams for better understanding, and to facilitate the interpretation of the results. The analysis of the results is included in different subsections depending on the results being addressed. Lastly, the conclusion of the study is presented.
2. Background

This section introduces image matching and the similarity measures used in this study. Multimodality is also explained with a focus on medical imaging.

2.1. Image matching

Image registration is the process of aligning different images of the same scene so that they spatially correspond to each other, basically making a match between images. The differences can be of many forms such as those that could arise when imaging from various angles, at various times and with diverse sensors or devices. Improvements in computer science have contributed to medical diagnosis with reliable and efficient image processing methods, thus, permitting clinicians to visualise objects within the human body. Image matching has its application in the medical field so as to obtain more concrete information about medical conditions of patients. There are several techniques for the actual matching but they could be categorised into two major groups: area-based methods and feature-based methods.

2.1.1 Area-based methods

The area-based method also known as the template matching method is a matching algorithm that consists of identifying a part of an image that matches a defined template. A template is a patch cropped from an image, in this case, the original image. This method defines the correspondence between images by determining the similarity of their gray level values. Additionally, the method performs feature matching on images without attempting to detect salient objects on these images. Therefore, the feature detection and the matching part are combined in the image registration process. The area-based method uses pixel intensity on the image to measure the similarity between a template and a search window. This can be done by analysing quantities of the mutual information or statistical correlation between the search window on the image and the template. Area-based methods exploit image intensities. These methods are consequently sensitive to intensity changes caused by noise, brightness variation or sensor variation. Examples of area-based similarity measures are cross-correlation and mutual information [1,7].

2.1.2 Feature-based methods

Feature-based methods are based on the extraction of salient structural features on images. Feature-based matching determines the correspondence between two images by comparable features such as significant regions, lines or points. In contrast to the area-based methods, the feature-based methods do not work directly with image intensity values. Features represented on the images exhibit information on the higher level and are therefore preferred. This attribute
makes feature-based methods suitable to situations where illumination changes are expected, or multisensor analysis is demanded [8].

Consequently, feature-based methods are more applicable to images congregated with salient and distinctive objects such as images produced by remote sensing. Nevertheless, area-based methods are frequently used and more applicable in the medical field because in the captured texture are not distinctively remarkable.

Matching can be very challenging when several modalities are used to extract information. Although it may be difficult dealing with information acquired with different modalities, it may be a good way to extract more information e.g. about a patient’s health [9]. The key issue connected with image matching is a choice of a matching entity. It is also connected to similarity measures, which are quantitative measures evaluating the match of entities.

2.2. The similarity measures

Similarity measure plays a crucial role in the image matching process because it is used to measure the spatial correspondence between images [2]. Various similarity measures have been formulated throughout the years, each with its strengths and weaknesses. Some measures use coarse image intensities, some normalise the intensities before using them, some use the ranks of the intensities, and some use joint probabilities of corresponding intensities [10]. An important distinction when using similarity measures is the modalities involved. When a single modality is involved, intensity-based similarity measures such as the sum of squared differences (SSD) or the sum of absolute difference (SAD) are more convenient, since the image intensity at corresponding points should be similar as images are of the same type or source [11]. However, when images are from different modalities, information theoretic similarity measures are more applicable, because these measures are sufficiently robust to cope with transformations arising when different modalities are involved such as non-linear illumination changes. Information theoretic similarity measures focus on the statistical dependence between pixel intensities and can easily disclose non-linear changes. Some of the information theoretic similarity measures are Mutual information (MI), cross-cumulative residual entropy (CCRE) and the sum of conditional variances (SCV). These measures are each further analysed in the following sections.

2.2.1 Mutual information (MI)

The use of mutual information (MI) was originally introduced by Viola and Maes. It has been shown to be an accurate and suitable similarity measure for registration of multimodal images. This similarity measure relies on the statistical dependence between two random variables in images, which is the amount of information that one image contains about the other [12, 32].

MI is closely related to the joint entropy, it analyses the joint entropy of the considered variables (images) and assumes no prior functional relationship, but rather a statistical relationship. The
measure derives from the computation of the entropy, which is a function describing the unpredictability of a variable. The $MI$ between two random images $A$ and $B$ are given by the following [12]:

$$MI = H(A) + H(B) - H(A,B) \quad (1)$$

where

$$H(A) = -E_A(\log(P(A)))$$

and

$$H(A,B) = -E_{A,B}(\log(P(A,B)))$$

in which $P$ is the probability function of a variable.

When maximising $MI$ between two images, more complex overlapping regions may be covered in the computation. The maximisation can be done by augmenting the individual entropies while diminishing their joint entropy. The $MI$ between two images is maximised when both images are geometrically aligned [12].

$MI$ has been a successful similarity measure in the medical domain, as it has proven to be robust to outliers and efficient in calculations. However, a significant limitation with the measure is that it does not acknowledge the spatial relationship between images. Furthermore, even with the extensive use of $MI$, it has been proved that there is a need to improve it due to misregistrations [13,14].

### 2.2.2 Cross-cumulative residual entropy (CCRE)

*Cross-cumulative residual entropy* (CCRE), also called *cumulative mutual information* was introduced by Wang and Rao and generalised by Drissi [16]. This measure has proven to be an alternative to $MI$ for multimodal registration [15]. It relies on the concept of measuring the entropy by using cumulative distributions and derives from the cumulative residual entropy (CRE). The key strength of the measure over $MI$ is its large noise immunity and rapid convergence over the field of transformation. It also supports large non-overlapping regions and variation in contrast and brightness [16].

Given two images $X$ and $Y$, CCRE ($C$) is defined as the following [17]:

$$C(X, Y) = \xi(X) - E[\xi(X|Y)] \quad (2)$$

where

$$\xi(X) = - \int_{R^+} F(\lambda) \log F(\lambda) \, d\lambda$$

and

$$F(X) = P(|X| > \lambda)$$
In which $\xi$ is the cumulative residual entropy of a variable and $F$ the cumulative residual distribution replacing the density function $H$ (see equation 1) used in MI. This replacement is used because the cumulative residual distribution is more prevalent than the density function. Also, it enables the preservation of the principle that the logarithm of the probability of an event should represent the information content in the event [17]. CCRE has been proposed as a similarity measure in the medical field, and it has proven to be more robust than MI [19].

2.2.3 Sum of conditional variances (SCV)

The Sum of conditional variances is invariant to non-linear illumination variations, and it is multimodal and computationally inexpensive. It can handle a large convergence radius with computational ease compared to other traditional similarity measures (e.g. the sum-of-squared difference and the sum-of-absolute difference) [18]. The measure utilises the joint of probability function like MI but is readily differentiable. Unlike MI, SCV assumes a statistical relationship between pixels on the basis that adjoining intensities are similarly allocated on both images. This ability makes the measure applicable in the context of multimodality [18].

SCV works as follows: For a given pair of images $X$ and $Y$, we refer to $X$ as the reference image and $Y$ as the target image. The measure partitions the pixels of the target image $Y$ into $n_b$ disjoint bins $Y(j)$ where $j = 1, ..., n_b$, and the same is done for the corresponding intensity regions of the reference image $X$ into $X(j)$. The calculation of the matching value is then done by summing the variances of the intensities of each bin $Y(j)$, this with the formula;

$$S_{SCV}(X,Y) = \sum_{j=1}^{n_b} E[(Y_i - E(Y_j))^2 | X_i \in X(j)]$$

where $X_i$ and $Y_i$ with $i = 1, ..., N_b$ give the pixel intensities of $X$ and $Y$ with $N_b$ being the total number of pixels.

SCV was originally developed in the context of medical registration. It was anticipated to be robust to nonlinear illuminations variations such as those arising in multimodality situational context [18].

2.3 Multimodality

Multimodality is a concept that involves the incorporation of two or more imaging modalities. In the medical field, there is a need for images from different modalities, because different modalities reveal different types of information. Furthermore, using medical images, integration of useful data obtained from separate images is often desired. There are several situations where images must come from different modalities to allow a more accurate diagnosis by the fusion and comparison of complementary information [20].
There is no perfect modality in the medical field; each modality is constrained by the physical energy it consists of and by its interaction with the human body. Allowing an integration of the strengths of each modality will overcome their limitations [19].

The performance of an imaging modality is indicated by its sensitivity and its specificity. Sensitivity and specificity are terms used in medical tests to identify the presence of a disease or further diagnostic process. Sensitivity in this context is the ability to detect true information i.e. the ability to correctly diagnose patients with a disease. Specificity in this context is the ability not to detect information when it is not there - it is the ability to correctly identify patients without a disease [19,20].

The role of multimodality in the medical field is to provide the exact localisation, extent, and metabolic activity of the target tissue. Furthermore, the role of any multimodal imaging is to yield the tissue flow and function or functional changes within the surrounding tissues, and, in the topic of imaging, highlight any pathognomonic changes leading to eventual disease [21]. Many clinicians believe that multimodality imaging will improve the ability to diagnose, guide therapy or predict outcomes. For the reason that it is almost impossible to capture all details with a single modality, that would ensure robustness of analysis and resulting diagnosis. Monitoring images from multiple modalities will ensure a more reliable and accurate assessment [22].

However, the use of multimodal sources of images entails non-linear changes in brightness or intensity. Considering the fact that images acquired through similar sensors present the same intensity range, then they have a linear change in brightness and contrast. Whereas images from different modalities present different intensity ranges, sometimes on different areas of the image. This is known as non-linear change in brightness and contrast.

Some of the major modalities in medical practice include magnetic resonance imaging (MRI), computer tomography (CT), ultrasound and others. These imaging modalities find a range of application in diagnosis of medical conditions affecting, e.g. the brain, breast, lungs, liver and bone marrow [22]. Imaging modalities are organised in two categories: morphological and functional techniques. The categories are described below.

2.3.1 Morphological methods

The morphological methods are applied to assess a mass according to the size, morphology, location and infiltration of the surrounding region. In other words, morphological methods provide anatomical information but provide little insight on the functional information. Morphological imaging technologies include CT and MRI [23]. Lately, MRI has gained interest due to its ability to obtain noninvasive contrast images of morphological structures [24]. Morphological methods, however, have a limitation. They are unable to detect a disease until structural changes within tissues are large enough to be morphologically identified [25].
2.3.2 Functional methods

Unlike morphological methods, functional methods offer the possibility to detect molecular and cellular changes caused by a disorder at the structural level. This ability implies that functional methods can be used when the structural level within tissues cannot be detected morphologically.

2.4. Medical imaging

The human body is a complex system. Acquiring information about its statistical and dynamic properties produce massive amounts of information [26]. Representation of this information as images has been the most efficient way to address the matter of how to acquire a vast quantity of information, so that the information can be interpreted and presented in a way that will yield better diagnostics. The most efficient way is known as medical imaging. Medical imaging can be defined as images of tissue characteristics that influence the way energy is emitted, transmitted, and reflected on the human body. As we humans rely on our sight to relate to the world around us, scientists rely on images to relate to the human body [26]. The biological properties that are accessible through image acquisition vary spatially in response to the structural and functional changes in the examined body. Depending on the type of energy used in the process of acquiring images, different properties are revealed. As a result, the obtained images will exhibit different properties. Intelligent interpretation and analysis of medical images require an understanding of their acquisition and results [19]. Some medical images are obtained by magnetic resonance imaging and computer tomography, which are described below.

2.4.1. Magnetic resonance imaging (MRI)

Magnetic resonance imaging (MRI) is a medical image acquisition process that uses a magnetic field and radio-frequency signals to produce images of the biological and anatomical functions in the human body. The imaging method uses a combination of external and internal energy sources to acquire more accurate information about tissue characteristics, and also physiological responses and functions [27]. MRI imaging is known to be a sophisticated imaging method as it provides images with high-resolution anatomical details and excellent soft-tissue characterisation. The images are identified with particular tissue characteristics or blood conditions that are predominant sources of contrast in the image. Usually within the range of black and white, MRI images are often used to display soft tissues like brains, where tissues with a high density of protons on the tissue will have a brighter appearance on the image. Fluid and blood may appear on the image as a source of contrast depending on the amount of water in them, and the extensive methods used during the image acquisition. Hard bones and air do not yield signals and therefore, appear as black on MRI images [28].

MRI images enable radiologists to analyse the distribution of brightness in the obtained image, to determine the health of tissues. Bright areas are tissues that emitted a high intense signal,
whereas dark areas are those that did not produce any signal. Since there is a range of signal intensity, the gray area may be spotted to show difference among various tissues [28].

MRI imaging is used to detect diseases like the brain tumour, spinal tumours and even blockage of blood vessels. MRI is a powerful tool in the medical field because the process can be optimised to display a broad range of medical conditions through images.

2.4. 2. Computed tomography (CT)

Computed tomography imaging (CT) is a medical image acquisition method that is derived from the X-ray method and makes use of the slice-technique known as tomography: visualisation by slices, layer by layer. It can be concretely defined as an imaging technique that produces cross-sectional images representing in each pixel the local X-ray attenuation properties of the human body. CT imaging uses only external energy sources and measures the attenuation coefficients of X-rays that are based on the density of the tissue or the part of the body being imaged [29].

As a result of the slice-technique, CT images display a better contrast between different tissues. A CT image can be seen as an image of densities of tissues. However, these images reveal bones much better than MRI images, but they can also display soft tissues and fluids. The quality of a CT image is measured regarding contrast among others factors such as spatial resolution. The contrast in CT images is dependent on the differences in X-ray attenuation by absorption in different types of tissues, and thus resulting in differences in intensities. The differences in attenuation is due to the electron density of different tissues. Soft-tissue contrast comes from the differences in physical density. Small changes in tissues are easily captured on CT images because of the nature of the image (two-dimensional dimensional layer) [30].

CT imaging help clinicians to detect brain damage after a head injury, brain tumours, and spinal stenosis. It can be also used to view sinuses as the inner ear is made of delicate bones. An advantage of CT imaging is the ability to visualise structures of low contrast in a subject.
3. Method

This section describes the data used in this study, and the technical experiment on the measures previously presented.

3.1. Data collection

As mentioned earlier, the scope of this paper only includes CT and MRI images in the medical field. Consequently, the dataset collected was to include both MRI and CT images displaying the same object, to be qualified as a multimodal dataset. There were no restrictions on the object displayed, provided that both imaging technologies were displaying an anatomical part to remain within the restraint of the medical field. An important but challenging selection factor during the collection of the data was the requirement for images from both modalities to differ but slightly.

Two types of images were considered during the collection: images with distinct texture and images with minor texture. The data collected in this context were allowed to display different objects, provided there was a significant distinction of displayed texture. The purpose of including texture variation was to investigate the performance of similarity measures when variations in structure were included in the data. All images were in gray format with a maximum size of 512x384 pixels. The gray scale was preferred as it solely entails intensity of monochrome images.

Because of the restrictions on the images, no image database seemed to provide eligible images. At first, a collection of ten different datasets from different sources was made, but eventually, only one set and two single images were found eligible. The data that was used during experimentation are presented below.

**Fig 1.** CT image of a brain. (Source: University of Hawaii, 2002)  
**Fig 2.** MRI image of a shoulder. (Source: Cincinnati children's center, 2015)  
**Fig 3.** CT and MRI images of lungs. (Source: Amber diagnostics, 2016)
3.1.1. Data acknowledgement

The collected data were from different but reliable sources. The sources are presented below. The data presented in Fig 1. was collected from the University of Hawai, John A. Burns School of Medicine. The image was captured on an infant to analyse aseptic meningitis. The analysis was conducted by medical students K. Higashigawa and L. Yamamoto, analysing the radiology Cases in Pediatric Emergency Medicine. The data presented in Fig 2. was collected from Cincinnati Children’s Center, an academic medical centre. The image was captured to analyse shoulder dislocations of a patient and was provided by Dr. Alexander J. Towbin. The data presented on Fig 3. were collected from Amber diagnostics, an American company that supplies medical equipment of various types. The image was captured in relation to a comparison between MRI and CT imaging technologies.

3.2 Technical Approach

The technical approach includes the programming language used for the experiment, and details of how the experiment was performed.

3.2.1. Matlab

The programming language used during experimentation was Matlab. Image processing toolboxes are used broadly in computer vision. An image processing toolbox, written in Matlab, is a collection of functions that extend the capacity of the programming language in a numeric computing environment. Thus, this makes Matlab heavily applicable as a computation tool for mathematical purposes. Hence, this causes the language to be appropriate in the context of similarity measures that make use of probabilities and statistics. The functions in toolboxes make image processing operations easy to write and also offer the ideal software prototyping environment for solutions to image processing problems.

3.2.2. Experiment

The experiment was based on performing a simulation of data using the gradient descent algorithm. The aim was to perform a registration task on two images to evaluate the convergence performance of the different measures. In other words, given a location on one image referred to as the reference image, and a displacement of the other image referred to as the target image. We evaluate the ability for the different measures to maneuver the target image as to reach the right location on the reference image following the steepest trajectory. The reason for using the gradient descent search is the constant demand to choose locations where the similarity is the lowest along the trajectory, the local minima. This procedure was continuously executed until convergence, or the maximum number of iterations was reached. The obligation of the steepest
search was entailed by the fact that the lower the similarity is between patches on the images, the more similar are the patches. Additionally, the gradient descent is one of the simplest algorithms that can be used in the context of the steepest path, and the algorithm is often used in image processing.

3.2.2.1 Gradient descent approach

Gradient descent is a popular algorithm used to perform optimisation. The algorithm is used to minimise objective functions. A simple definition of the gradient is the partial derivative for each dimension in the input space. Given a function defined by an initial set of parameters \( J(\Theta) \), the gradient descent update parameters as to interactively move in the opposite direction of the gradient of the function to the parameters \( \nabla J(\Theta) \). In other words, the gradient moves toward a set of parameters that minimises the function, also, the direction that is the steepest. In term of similarity measure, the gradient descent implies following the trajectory where the similarity value is lowest.

3.2.2.2 Simulation

Images were converted into two-dimensional matrices where each element represented a pixel. One of the images was chosen to be the reference image and the second as the target image. A location on the reference image was randomly selected to create a region or a patch, referred to as the template. The template is merely an extracted window on the reference image. A random displacement was performed on the target image before extracting a patch, referred to as the current. The size of both windows was set to 50x50 pixels, while the maximum distance for a displacement was set to 15 pixels.

For each measure, the similarity is computed between the template and the current using different formulas (see section 2).

The gradient descent is introduced by the following: the current window is modified so as to include an extra pixel on the four different cardinal directions respectively. Computing the similarity value of the modified window for each direction, the original position of the current window will be shifted towards the direction that reduced the similarity value. This procedure was repeated and could terminate in the following two ways: The maximum number of trials was reached or the current and template were correctly aligned. It was difficult to determine when the patches were perfectly aligned, because the lowest similarity for every movement in the trajectory does not necessary imply the lowest in the whole trajectory.

In a case of failure, the procedure was terminated once the number of maximum trials was reached, which was set to 30 in the experiment.

Convergence in this context refers to the ability for the current to track and reach the same position as the template following the steepest path using similarity measures. The use of similarity measure as a tool for direction implies that convergence is reached when the similarity
between patches is the lowest, which means when both windows are most similar and therefore display the same object on both images.

To remove the effect of noise and get an accurate result, 30 trials were conducted for each displacement of the current window. The convergence rate for each similarity is simply the number of times convergence was successfully reached, represented in proportion to the number of trials, which was 30.
4. Results

This section presents the result of the study obtained by using the previously described method. The results are presented as plots and will be assessed in the discussion section.

The results of the experiment are presented as representative plots, showing the outcome after the maximum number of trials was reached. All plots consist of the convergence rate in relation to the displacement of the current window. The convergence rate represents the proportion of succeeded convergence with respect to the number of trials (For better understanding see section 3). The result is presented for the different collected data in the following subsection.

4.1 Simulation on identical data

Self-simulation was conducted, that is, using the same data as both reference and target image. This simulation is performed to analyse the convergence in the trivial scene where tracking should be guaranteed.

4.1.1. High-textured data

The data introduce the context of distinct texture.

![Fig 4.1. The convergence rate in the context of high-textured data. Simulation conducted on identical data (self-simulation).](image)

The figure displays the convergence performance of each similarity measure in the context of distinct texture. On the figure, the convergence rate is put in relation to the displacement of the...
current patch. The curve representing SCV seems to decrease faster than those representing MI and CCRE. From a displacement of six pixels, although MI curve has a similar trend to CCRE curve, MI curve seems to decrease slightly faster. However, MI curve reaches convergence at lowest rate 0.4 while CCRE curve reaches at lowest 0.3 and SCV 0.1.

4.1.2. Low-textured data
The data introduce the context of minor texture.

As displayed on the figure, the CCRE curve decreases slower than MI and SCV curves. However, MI and CCRE curves seem to have similar trend while SCV curve represent a slight different trend. The lowest rate of CCRE is at 0.35 while MI and SCV have their lowest convergence at 0.1.
4.2 Simulation on multimodal data
Simulation was performed on two images from different modalities, CT and MRI. The CT image was selected as the reference image and the MRI image as the target image. The arrangement of images does not affect the result.

As illustrated on the figure, all similarity measures have difficulty in converging. At a displacement of one pixel, MI and SCV curves show higher convergence than CCRE. From a displacement greater than two pixels, all measures are likely not to converge. However, when convergence is reached, it is at random displacement positions with a rate of 0.1.
4.3 Simulation on negated multimodal data
Simulation performed on two images. One image is a real image used as the reference image. In order to illustrate multimodality, a negated version of the reference image is used as the target image. This is obtained by negating the real picture.

Simulation on high-textured data

Fig 4.4. The convergence rate in the context of negation. Simulation conducted on high-textured data.

Similar to the context of identical data, MI and CCRE curves have similar trends while SCV curve decreases faster. MI is able to converge with the lowest rate at 0.6 while CCRE converge at 0.5 and SCV 0.2 at their lowest.
Simulation on low-textured data

CCRE reach higher convergence compared to the other measures, with the lowest rate at 0.2 against 0.1 for the other measures. It can also be identified on the curves, that CCRE holds a higher convergence than the other measures for all displacement position.
5. Discussion

This section discusses the results presented in the previous section.

Image matching is an important task in the medical field. The comparison of both morphological and functional images can lead to a better diagnosis of a patient’s body. In the said field, there is a need to identify the correspondence between images so as to follow changes in structure, function, etc. Similarity measures are tools used for the purpose of measuring the correspondence between images. Making use of these measures, the convergence performance of each measure, which is the ability to find correspondence even after a displacement has occurred, may be used as a guide when selecting a measure to perform matching.

In this study, we wanted to investigate the convergence performance of different similarity measures used in different contexts. The contexts included in this study are multimodality and variation in texture. The purpose of choosing this context was to investigate the convergence performance of similarity measures when multimodality is involved and also to investigate how variation in texture may affect the convergence of measures.

To analyse the convergence performance of different measures, the experiment was conducted using different data, the reason for that is to identify changes in the convergence performance of measures.

The following subsections comments the method and data used in the experiment, and finally, the results are discussed.

5.1. Note on the method

Image processing functions used in the experiment were part of an image processing toolbox. The toolbox was made available by Mathworks and could, therefore, be considered as reliable. The formulas for each similarity measure were computed without inversion of values which implies that values were expected to decrease as the similarity increased. Consequently, the gradient descent optimisation was applicable in the experiment, as the steepest trajectory will imply the trajectory where the patches become more similar.

The simulation was conducted for 30 trials for each displacement, with a total of 450 trials for 15 pixel displacement. A trend that was noticed was the further the displacement, the longer time each trial took. For a displacement of 1 pixel, the time of computation was 0.218 CPU time on a Windows 7 Enterprise, 3.40 GHz Intel Core i7. While for a displacement of 15 pixels, the computation time was 2.58 CPU time on the same device. Further research could consider conducting the simulation on a wider range of data.

5.2. Note on the data

The results obtained in the study were based on the experiment conducted on one dataset and two single data. As mentioned before, restriction and requirements on the data rendered data collection difficult, and, as a consequence, the experiment involved only a few data.
The situation in which images are captured using CT and MRI entails itself a limitation. Images captured by these modalities are often multitemporal, which means they are captured at different times. In many cases, physicians can predict the information that can be retrieved on the patient at a specific time. Based on that, they will select one appropriate imaging technology. Consequently, available data were multitemporal, and, as a result of capturing at different times, the images displayed different information. Thus, although there is a wide range of MRI/CT data available, very few qualified for the purpose of this study.

5.3. Result analysis

Both the context of texture and that of multimodal have been experimented on using similar parameters so as to obtain more valuable results. The discussion of the obtained results for each context is presented in its respective section below.

5.3.1. Variation in texture

Convergence was analysed in the context of the variation in texture. This was done by measuring the convergence when the texture on image is distinct and when texture is minor. An important fact to remember about this study is that the convergence rate is expected to decrease the further the displacement. This is due to the fact that similarity is expected to deteriorate as the distance between patches increases. Because the further the displacement of the current window, the less information it shares with the template window.

Each similarity measure quantifies similarity differently. Consequently, convergence is also reached differently for the similarity measures. To make the convergence performance more comprehensive and explicit, the following tables and diagrams have been drawn. The values on the representations are gathered from a fair reading of the result in Fig 4.1 and Fig 4.2.

<table>
<thead>
<tr>
<th>Tool</th>
<th>1 px</th>
<th>5 px</th>
<th>10 px</th>
<th>15 px</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>100%</td>
<td>47%</td>
<td>25%</td>
<td>15%</td>
</tr>
<tr>
<td>CCRE</td>
<td>100%</td>
<td>95%</td>
<td>60%</td>
<td>35%</td>
</tr>
<tr>
<td>SCV</td>
<td>100%</td>
<td>54%</td>
<td>30%</td>
<td>15%</td>
</tr>
</tbody>
</table>

*Table 5.1. Convergence rate on low-textured data.*
The result in Fig 4.1 and 4.2 together with the diagrams and tables above reveal that CCRE outperforms MI and SCV, when the texture are unremarkable. CCRE is constantly tending to reach a higher convergence rate compared to the two other measures. However, its performance decreases when the texture on data are noticeable, while MI appears to have improved. MI is an area-based method, its improvement in the context of distinct texture is therefore expected. The area-based method focuses on performing a feature matching rather than feature detection. In other words, we do not look for salient objects; we simply look for a match of a feature in windows, which in the case of low-textured data, can be difficult to accomplish due to the absence of features (see section 2). SCV appears to outperform MI in the context of low texture, but its performance is inferior to other measures in the context of distinct texture. SCV is less robust in the context of local variations, which is the intensity variation of a pixel compared to its nearest neighbor. Consequently, its matching ability was affected, just like the convergence capability of the measure [31].

5.3.2. Multimodality

The result is presented on Fig 4.3. It shows that convergence is difficult to reach for all measures. On 1 pixel displacement, the convergence rate can be reached at best, in 30% of trials. Also, different measures converge at different displacement positions without a clear logic, which makes a clear performance hard to detect. The figure and the table below illustrate the convergence performance of the different measures in the context of multimodality. The values on the representations are gathered from a fair reading of the result in Fig 4.3.

<table>
<thead>
<tr>
<th>Tool</th>
<th>1 px</th>
<th>5 px</th>
<th>10 px</th>
<th>15 px</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>100%</td>
<td>90%</td>
<td>68%</td>
<td>45%</td>
</tr>
<tr>
<td>CCRE</td>
<td>100%</td>
<td>80%</td>
<td>77%</td>
<td>30%</td>
</tr>
<tr>
<td>SCV</td>
<td>100%</td>
<td>50%</td>
<td>25%</td>
<td>10%</td>
</tr>
</tbody>
</table>

*Table 5.2. Convergence rate on high-textured data.*
As illustrated in the figure, the convergence is hardly reached for all measures as it varies drastically at different displacement distances. Additionally, the result cannot be generalised as the simulation only consists of one dataset. Due to restrictions and requirements on the dataset (see section 3), finding another proper dataset was unsuccessful. The possible reason as to why the result was poor and unclear may be due to the misalignment in the dataset.

5.3.2.1. Misalignment

Using different modalities when capturing images may entail alignment difficulties, caused by the different information that can be captured. MRI has the capacity of capturing tissues and fluids, while CT may barely reveal the same information; it better displays bones (see section 2.4). The information captured by these techniques can lead to significant distinctions of images, thus making a dataset ineligible for simulation (see section 3.2.2). In some cases, there may be variation in certain parts of the object displayed, causing misalignment. Misalignment between images can be an obstacle when performing similarity computation. When geometric
deformation or misalignment exists in the images to be matched, the success rate of matching is low, and matching may eventually fail. The figure below illustrates areas of misalignment on the dataset used in the context of multimodality:

![CT and MRI images with misalignment](image)

Fig 5.1. Misalignment on data set.

As misalignment affects the convergence rate of similarity measures, it should be avoided. In order to deviate from the alignment issue while persisting in the context of multimodality, an alternative is to make use of negation.

### 5.3.3 Negation

As the use of different modalities entails a difference in intensity on images, negation, which implies inverting the color on the data, can be used to introduce variation of intensity. Thus, negation was used to illustrate the context of multimodality.

The values on the representations are gathered from an approximate reading of the result in Fig 4.4 and Fig 4.5.

#### 5.3.3.1 Negation on high-textured data

The table and diagram below are used to illustrate the convergence performance of measures.

<table>
<thead>
<tr>
<th>Tool</th>
<th>1 px</th>
<th>5 px</th>
<th>10 px</th>
<th>15 px</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>99%</td>
<td>95%</td>
<td>85%</td>
<td>65%</td>
</tr>
<tr>
<td>CCRE</td>
<td>99%</td>
<td>90%</td>
<td>55%</td>
<td>50%</td>
</tr>
<tr>
<td>SCV</td>
<td>99%</td>
<td>45%</td>
<td>30%</td>
<td>10%</td>
</tr>
</tbody>
</table>

**Table 5.4. Convergence rate on high-textured data.**
As illustrated in the diagram above, MI outperforms CCRE and SCV by reaching a higher average convergence. However, there is a significant difference between CCRE and SCV, as CCRE seems to reach twice the SCV convergence rate.

### 5.3.3.2 Negation on low-textured data

The table and diagram below are used to illustrate the convergence performance of measures.

<table>
<thead>
<tr>
<th>Tool</th>
<th>1 px</th>
<th>5 px</th>
<th>10 px</th>
<th>15 px</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>100%</td>
<td>50%</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>CCRE</td>
<td>100%</td>
<td>98%</td>
<td>60%</td>
<td>20%</td>
</tr>
<tr>
<td>SCV</td>
<td>100%</td>
<td>55%</td>
<td>20%</td>
<td>20%</td>
</tr>
</tbody>
</table>

**Table 5.5.** Convergence rate on low-textured data

As illustrated in the diagram above, CCRE outperforms both MI and SCV. SCV and MI seem to have similar convergence rates.

### 5.3.3.3 General comments on negation

By comparing the diagrams presented in section 5.3.3., one can clearly see that all similarity measures appear to have a higher convergence rate when negation was introduced in the context of multimodality. In the case of self-simulation and for similar reasons, MI performs outstandingly when the texture on the data is accentuated and significant, and CCRE performs better when the texture is relatively unremarkable. However, in the case of negation, there is a slight improvement in the convergence rate of SCV when texture are less distinct. This is inconsistent with the case of self-simulation, where SCV performed better when distinct texture were involved. The changes in performance for SCV are much more stable compared to that of the other measures; the rate of convergence of SCV varied within a range of ±10% between
contexts. SCV seems to be little indifferent to variations in texture and also to the introduction of multimodality.

Negation is a way of creating simple multimodal images. Negation entails variation in brightness and contrast. However, negation may entail linear variation in brightness and contrast while multimodality entails non-linear variation. Despite this, negation remains one of the easiest ways to illustrate multimodality.

If the simulation was to run on real well-aligned multimodal images, the performance of measures may have varied widely from the one presented earlier. It is precipitated to conclude that the use of well-aligned multimodal images will have a significant effect on the convergence rate of any measure. But from the results obtained by negation, we can conclude that the convergence rate of any similarity measure in the context of multimodality is dependent of the texture present on the data. The performance detected in the context of self-simulation could be preserved even in the context of multimodality.
6. Conclusion

It is hard to draw any definite conclusion based on such small data; it is possible that the outcome could be different if more data are tested. But based on the executed experiment, the results obtained, and the diagrams used to show performance, we could conclude that: the area-based method, the mutual information (MI) has a prominent convergence performance on high-textured data. The cross-cumulative residual entropy (CCRE), being an improvement of the mutual information, reaches higher performance than the mutual information (MI) when the texture on the data is unremarkable. The sum of conditional variances (SCV), although it does not always achieve a high convergence rate as MI and CCRE, is the most stable when the different contexts are taken into consideration. Its value does not diminish or increase drastically between contexts. As illustrated by negation, the same performance can be retained in the context of multimodality, given that the data are eligible. However, multimodality entails variance in the information displayed, which can affect the performance of the similarity measures.

Research has been conducted to investigate the possibility to extend SCV in order to improve its convergence rate. An extension to SCV is the sum of conditional variances of difference (SCVD) which has been shown to reach a higher convergence rate than SCV [18]. Further research could consider analyzing its convergence performance in the different contexts. As mentioned earlier, further research could include a wider range of data as to generalise the performance of the similarity measures. Further implementation could consider how to improve speed of the simulation in order to run simulation on a larger number of trials.

This study is useful in the medical field, because it analyses the convergence performance of three information theoretic similarity measures, and could therefore entail a better use of these measures. In a matching or tracking task, clinicians need to consider variations in texture on multimodal images to know on what similarity measure to rely. With that in consideration, better diagnosis can be expected when there is a need to analyse a patient's condition by images.
7. References


