Seventh Framework Programme
Research for the benefit of SMEs

SAFEMETAL

Increasing EU citizen security by utilising innovative intelligent signal processing systems for euro-coin validation and metal quality testing

FP7-SME-2010-1-262568-SAFEMETAL

Deliverable number: D2.4

Responsible partner: Telecommunications Systems Institute

Intelligent Signal Processing Algorithms

Status: Final

Dissemination:
Confidential, only for members of the Consortium (including the Commission Services)

Date: 31 October 2011

D2.4 Intelligent Signal Processing Algorithms

TABLE OF CONTENTS

This project has been funded with support from the European Commission.
This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

List of Figures

List of Tables

1. Introduction

This report focuses on the investigation of numerous signal processing algorithms that can prove to be effective in the context of SAFEMETAL. The aim is to thoroughly examine a range of suitable algorithms and select one that promises to deliver results at minimum expense both in effort and resources. The motivation was to dedicate this report on the development of an underlying theory that will lead to suitable models responsible for processing the signals coming from the data validator. Initially, the intention was to base this on the chromatic analysis methodology, however, as the project evolved; it seemed more prudent to direct our attention to state-of-the-art algorithms associated with computer vision and specifically computer vision. Hence, this report thoroughly analyses a number of key object recognition algorithms, assesses their pros and cons and presents the most promising candidate suitable for the task currently under investigation, i.e. processing of images captured from video cameras operating within the validator and placed on either side of a coin that has come to rest.

Object recognition (OR) is an integral part of image processing and computer vision. Currently, both those areas represent an application domain and research area of immense proportions. Similarly, the term object recognition based on the definition that *"given some knowledge about the appearance of certain objects, one or more images are examined in order to evaluate which objects are present and where"*[1] spans over a significant number of applications. All this has led to the development of many different algorithms. This diversity in object recognition algorithms has spawned an equivalent multitude in documented material, publications and literature in general.

Often, the problem tends to be the highly specialised nature of the documented material related to the topic of object recognition. This is understandable considering the enormity of its application domain, however, this makes it difficult to come across a well documented and complete account on the subject. A worthwhile account on the topic of object recognition can be found in several sources[1,2,3,4,5].

---

[1] M. Treiber, *An introduction to Object Recognition*, Springer-Verlag London Limited, 1st edition, 2010

[2] J. C. Russ, *The image processing handbook*, Taylor & Francis Group LLC, 5th edition, 2007

[3] R. Szeliski, *Computer vision: Algorithms and applications*, Springer-Verlag London Limited, 1st edition, 2011

[4] R. Brunelli, *Template matching techniques in computer vision*, John Wisely & Sons Ltd, 2009

[5] R. C. Gonzalez et al., *Digital image processing*, Prentice-Hall Inc, 2nd edition, 2001

In this work, a generalised approach is adopted that provides a "concise" summary of the algorithms used in object recognition at the expense of in-depth and application-specific analysis of state-of-the-art problems. Hence, it offers an overview of the field but lacks information on the detail.

This document will follow loosely the structure assumed in Treiber[1] and it will complement it wherever necessary such that important information related to the work carried out in SAFEMETAL are not omitted. For instance, addressing in a lot of detail the topic of 3D-object recognition is not as important as being quite thorough with respect to 2D-object recognition.

2. Overview

Prior to discussing in detail the various object recognition algorithms, it would be worth summarising a few important aspects. First, we can categorise the areas of application into the following:

- *Position Measurement*: Industrial applications many times have to do with accurately locating the position of objects such as Integrated Circuits (ICs) before placing them onto a PCB (Printed Circuit Board).

- *Inspection*: Many applications require inspection of manufactured parts for defects in order to ensure quality control in production environments

- *Sorting*: These are situations where the correct sorting of objects is of significant importance such as the case of parcels in postal automation environments

- *Counting*: In certain applications, it is highly important to measure the number of occurrences a specific object appears in an image. Typical application area is that of molecular biology.

- *Object Detection*: In this case, an image is compared against a model database leading to a successful or a non-successful match. The model of the object contained in the database is often built during a training phase prior to detection.

- *Scene Categorisation*: This category refers to the case where an object is to be matched to a class of objects rather than with a specific type of object, i.e. car, person, tree etc.

- *Image retrieval*: An input image that contains a certain object, an image database is searched in order to determine the images containing the same object or class of objects.

It is quite evident that the application area that corresponds to the SAFEMETAL project requirements is that of object detection. Now, in addition to the various types of applications, there also exist a number of key specifications related to object recognition. Naturally, the algorithm employed in an application has a direct consequence on those parameters and that is why selecting the most efficient is such as important task. The key parameters are the following:

- *Evaluation Time*: This is a very important parameter since it can influence cost so much. For example, in industrial environments, image data had to be processed in real time, e.g. 10-50 ms, to ensure high production speed. It is only natural that the object recognition algorithm plays a key role in ensuring that application specifications are met.

- *Accuracy*: Many applications require object detection placement within a fraction of a pixel and this poses a significant challenge. Obviously, high resolution cameras help in this but there also exists a trade-off between speed of detection and accuracy of recognition.

- *Recognition Reliability*: This is a challenge related to misclassifying objects to belong to a reference object or a certain class.

- *Invariance*: Invariance refers to the ability to identify an object regardless of variations in the image. In other words, the object recognition algorithm used ought to be insensitive to variations of the object to be detected. This is a very challenging task especially since the algorithm has to be highly sensitive to information discrepancies between objects of different classes but insensitive to information discrepancies between objects of the same class.

Finally, although a clear and concise categorisation of the different object recognition methods is prohibited by the sheer volume of different applications and their corresponding specialised methods of detection, it remains possible to provide one based on some fundamental criteria. Those criteria are found at the basis of object recognition analysis and it, therefore, helps to be aware of them since they are also relevant to the SAFEMETAL demands.

2.1. Object related criteria

It is very important to be clear about a number of aspects related to the information that describe the object. First, it is possible to for the information that describe the object to be based on either its *geometry* or its *appearance*. Using intuition, it is relatively easy to work out what each stands for. Geometrical data refer to information about the shape of the object, its dimensions, its boundaries or the position of specific shape characteristics on the object itself. A lot of the time, model creation is performed using CAD software.

On the other hand, appearance-based models represent characteristics of image regions covered by the object. In this case, model creation goes through a training phase , i.e. the model is built automatically using one or more images.

Second, based on the complexity of the application, the model data themselves can be of two types. Firstly, they can be of a *global* nature, i.e. global features such as area and perimeter are summarised in a global feature vector[1]. This approach is usually appropriate for simple objects on a 2D image. Secondly, model data can be more specialised and refer to *local* object properties. This method is more complex and, therefore, suitable for highly structured objects. In this case, the model is comprised of data that refer to different areas of the same object. Understandably, the approach based on local object properties is more robust against missing data due to occlusion.

Third, resilience to variations, due to change in viewpoint or illumination, is of paramount importance when it comes to object recognition algorithms. Objects that belong to the same class are not always represented identically in an image and, therefore, the algorithm has to be able to identify all objects that belong to the same class but are subject to inter-class variations and reject the rest.

2.2. Data quality and matching criteria

Depending on the application, it is sometimes easy to ensure that the image data are of high quality and in other times almost impossible. Naturally, this has an impact on the type and complexity of the object recognition algorithm. For instance, in industrial applications, it is possible to minimise the "disturbing" background material by presenting the objects upon a uniform background, which is something that on different application such as surveillance applications is impossible to do.

An important step during object recognition is the matching phase. During that phase, the object model is aligned with the obtained image, facilitating the decision on similarity or dissimilarity. The object recognition algorithm is responsible for the matching strategy, which can range from a simple correlation of data to optimising the parameters of a transformation characterising the relationship between model and the scene image. Furthermore, the type of data used during the matching step can be divided into three categories.

- Raw intensity picture values

- Low-level feature values
- High-level feature values

With low-level features, such as edge data, the borders of an object are often indicated by rapid changes of grey value intensities (in the case of a grey-scale image). The image pixels that correspond to those high grey value gradients are referred to as "edge pixels" and they can be grouped in order to high-level features, e.g. lines that can be grouped again to create line groups and so on. All this leads to an increasing information content giving greater accuracy in results at the expense of processing time. Therefore, once again, it becomes apparent that the level of accuracy is based upon the application demands.

3. Intelligent ElectroMagnetic Signal Processing

Electromagnetic (eddy current) sensors with corresponding signal processing (combined some times with image, optical and acoustical sensors, but also usable individually) could be a cost-efficient and precise method for coin recognition and validation[6], and are widely used in vending and other coin handling (e.g. sorting, counting) machines.

Various (by shape, magnetic core types and construction) electromagnetic sensors and signals (excitation frequencies and waveforms) could be used, sometimes separate excitation (transmitting) and receiving coils are used, in reflexive or transmitting through the coin modes, as described for example in the patent specifications (US 2010170766, US 6471030, US 6056104, GB 2393840, GB 2422941, EP 1172772, EP 1241636, EP 1411480,EP 1445739,EP 1577844). Most typically the solutions are working in the frequency range of 40-500 kHz.

Most simple electromagnetic (called also as “eddy current” or “inductive”) sensor could be a simple air-core or magnetic-core coil (or array of coils), from which the metal coin is passing (as done[7] in Fig 1). Using of several frequencies simultaneously or sophisticated (“rich spectrum”) signals gives of course more information about the electrical properties of the coin materials, as various frequencies penetrate to different depth of the material. Example of using of a single coil sensor and two simultaneous (“high and low”) frequencies (e.g. 100 kHz and 2 MHz) is given in the solution in Fig.2 [8]).

A solid theoretical and numerical simulation models are available for air-core single-coil cases[9], “above the (even multilayer) metal” starting from 1960-s[10]. Typically such models are

---

[6] A. Carlosena, A.J. López-Martin, F. Arizti, A. Martínez-de-Guerenu, J.L. Pina-Insausti, and J.L. García-Sayés, “Sensing in coin discriminators”, Proc. IEEE Symp., San Diego, CA, USA, Feburary 6-8, 2007, pp. 1-6.

[7] J. Harris, J.Churchman, and D. Sharman (Marconi UK), “Coin-validation arrangement”, US Patent 7 243 772, July 17, 2007.

[8] G. Howells, (Scancoin Ind AB), “Coin discriminators”, US patent 7 584 83, August 9, 2009.

[9] C.V. Dodd, and W.E. Deeds, “ Analytical solutions to eddy-current probe-coil problems”, Journal of Applied Physics, vol. 39 (1968), No. 6, pp. 2829-2838

[10] T. Theodoulidis,and M. Kotouzas, “Eddy current Testing Simulation on a Personal

considering “infinite size” of area of the “metal plate”, but as theory and experiments show, often a size of the coil, being a few millimetres smaller of the metal plate (coin), is enough for precise modelling the coil complex impedance. So, by measuring of the complex (phase-sensitive) impedance (ratio of the voltage across the coil to the current passing through) of the measurement coil, the electrical conductivity of the coin could be found exactly. One nice thing of (even for a single frequency measurement) is the possibility to find the lift-off and electrical conductivity of the material under test in parallel by using of the complex (real and imaginary part measurement results), as proposed for example in the solution[11].

In our current work the dynamical (in time, when coin passes the coil) measurement of the coil impedance for coin validation has been demonstrated, over a frequency range 20-500 kHz. Inverse eddy current impedance of the coil model has been used to estimate the electrical conductivity of the coin. Ways of adaptive subsampling in time domain and using adaptive number of frequency “bins” of DFT (discrete Fourier' transform) has been shown. The most time- and frequency resolution is used only at the very limited time-period, where coin is just passing the coil, so allowing to relax the requirements of the average processing power of the system.

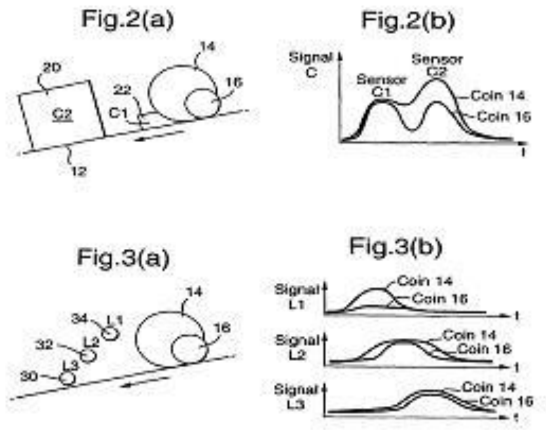


Figure 1 : Solution, proposed in US patent 7243772 [2].

Computer”, Proc. 15th World Conference on Nondestructive Testing, Roma (Italy) 15-21, Oct. 2000.

[11] J.H.Maindonald, “Using R for Data Analysis and Graphics: Introduction, Code and Commentary Centre for Mathematics and Its Applications”, Australian National University, 2008, 96 pp,

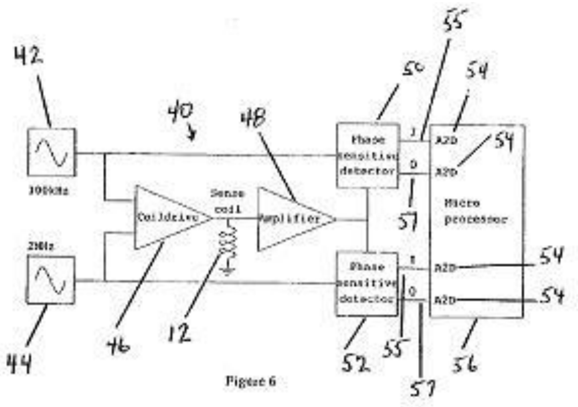


Figure 2 Solution, proposed in US patent 7584833 [3].

3.1. Solution for the coin conductivity measurement

The following solution has been implemented and tested. A coil, as flat as possible, has been made of a copper wire (external diameter of the coil - D=20mm, number of turns N= 288, ohmic resistance R=33,0 Ohms) and a provisional setup to rotate (various) coins over the coil has been implemented (Fig.3).

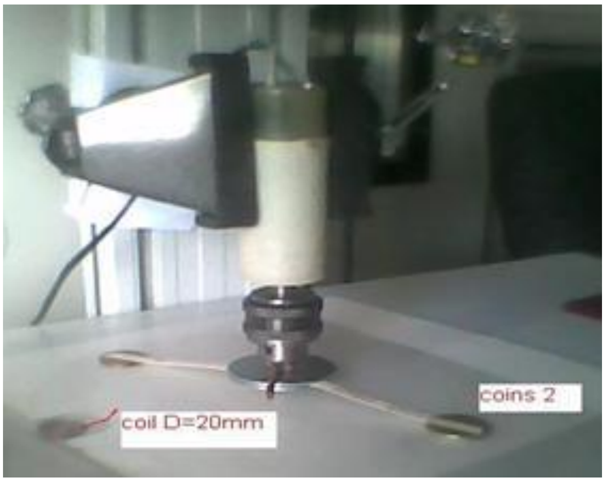


Figure 3 : Implemented test arrangement with D=20mm coil

An arbitrary waveform generator AFG-3252 has been used to generate excitation signals (10, 20, 50, 100, 200 and 500 kHz through a 100 Ohm shunt- resistor) and a scope as a digitizer at 20MSa/s rate to log the voltage drops across the coil and shunt resistor.

The time domain analysis of log files has been implemented in R-language environment[12]. A window size was 1ms (20 thousand samples) and step in analysis was 0,5ms, further "in-phase" (real-part) was found by formula:

*Zreal = aver (U(i) * I(i)) / aver (I(i) * (I(i))   (1)*

and for "quadrature" (imaginary) part values of the measured current were shifted by 90 degrees

I2 ← I (90 degrees)   (2)

and then found in the similar way

*Zimag = aver (U(i) * I2(i)) / aver (I(i) * (I(i))   (3)*

Such analysis have been done over all logged frequencies (20-500 kHz) and results of the measured impedance are shown, for 100 Khz example, for coins of 5 cents, 50 cents and 2 Euro coins on Fig. 4-6.

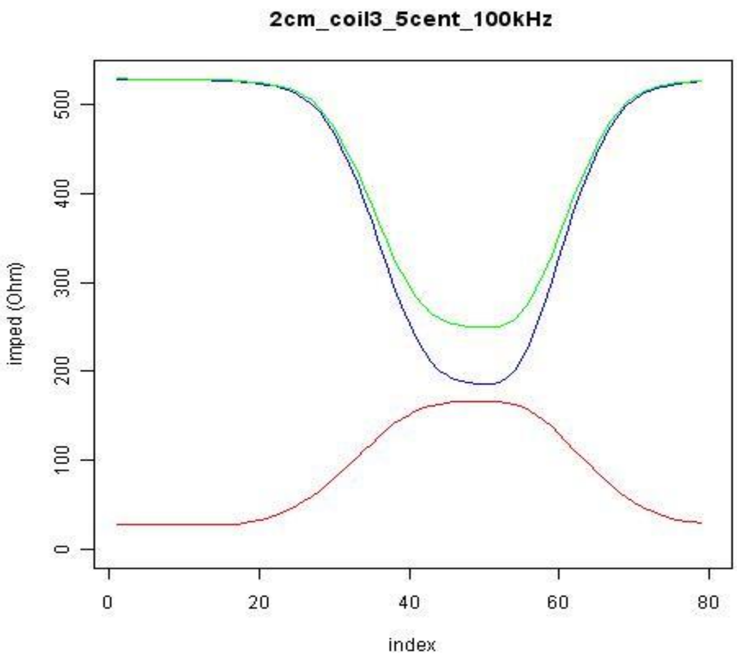


Figure 4 : Changing (in time) measured impedance with D=20mm coil, 100 kHz, 5 eurocent coin (from up to down- (green) module of the impedance, next line (blue)-imaginary part and lowest line (red)- real part, all in ohms). Index-(X-)axis is showing time, in 0,5 ms steps (units)

[12] Snyder, Patrick J. (The Boeing Company), "Method and apparatus for reducing errors in eddy-current conductivity measurements due to lift-off by interpolating between a plurality of reference conductivity measurements", US Patent No 5 394 084, February 28, 1995.

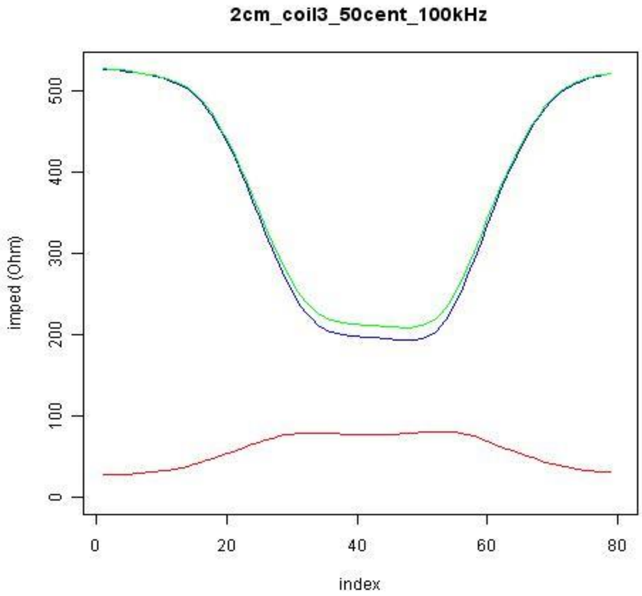


Figure 5 : The same with figure 4, for a 50 eurocent coin

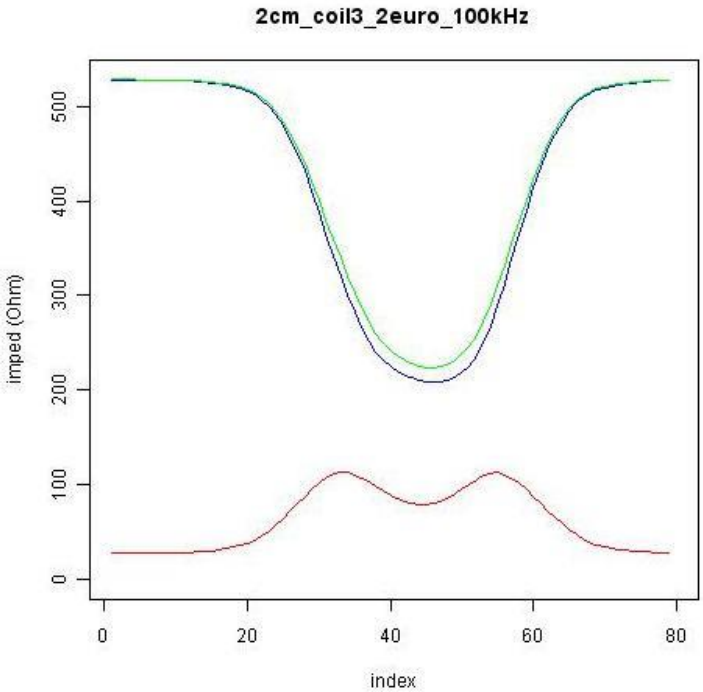


Figure 6 : The same with figure 4, for a 2 Euro coin.

Obtained curves can be used directly as reference values ("baselines") for coin recognition and validation. The impedance value, corresponding to exact passing of the coil by coin, in time domain, could be at the minimum value of imaginary impedance part.

Furthermore, for more precise validation, alternatively to use of the got results, at this point the change of the imaginary and real components of the coil impedance (compared with the situation, without the coin- or "coil in the air" in other words), corresponding to the "most left" values on the graphs and used in the (reverse) model of the "coil above the metal layers" as stated in the introductory subsection. Such "reverse model" to estimate the lift-off and conductivity (for single layer case, so far) has been developed by authors on the basis of model of Theodoulidis described above, in C/C++.

This reverse model gave good results for 50-eurocent case (9.4 to 9.8 MS/m conductivity in the frequency range 20-200 kHz and 7,3 MS/m at 500 Khz, against expected 9,6 MS/m value, at lift-off of 0,45 mm) and further corrections and improvements are probably possible.

For 5 eurocents, the eddy current models are not so simply applied, as these coins have significant magnetic properties, needed to be taken into account into eddy current models. For 2 Euros (as well as for 1 Euro coin), the coin is consisting of outer and inner rings and inner core has coating, so such composition needs more sophisticated modelling (and reverse modelling), what is currently under the way.

One improvement, in accuracy, motivating the using of the precise inverse eddy current model, is possible effective lift-off compensation, in the software.

4. Intelligent Image Processing

Image processing (combined with electromagnetic, optical and acoustical sensors) is an efficient method for coin recognition and validation[13]. While electromagnetic (eddy current) sensors can allow relatively simple, precise and efficient solution, still some different coins, worldwide[14], can have similar electrical or magnetic properties, and so image processing-based solutions can have their place[15]. Of course, complexity of image processing part for coin validation can vary from simple image-acquisition and processing solutions[16] up to more

[13] Snyder, Patrick J. (The Boeing Company), "Method and apparatus for reducing errors in eddy-current conductivity measurements due to lift-off by interpolating between a plurality of reference conductivity measurements", US Patent No 5 394 084, February 28, 1995.

[14] The European Central Bank, official web site: *http://www.ecb.int/home/html/index.en.html*

[15] J. Karlsson (Scan Coin Industries AB), "Coin discriminating device, coin handling apparatus including such a device, and coin discriminating method", US Patent 6,761,257, July 13, 2004.

[16] M. Tresanchez, T. Pallejà, M. Teixidó, and J. Palacín, "Using the Optical Mouse Sensor as a Two-Euro Counterfeit Coin Detector", Sensors, 2009, no. 9(9):pp. 7083-7096.

sophisticated image acquisition[17] and processing solutions[18].

Generally, image matching systems consist of at least two phases: image processing and object classification. One of the first tasks are image pre-conditoning and localizing the objects from the background.

The aim is to identify eurocoins and their location in image scale space that is invariant with respect to image translation, scaling, and rotation, and is minimally affected by noise and small distortions.

There are many factors to consider in formulating an image (coin) matching strategy including acceptable precision/error limits, scene characteristics, instrument parameters, and the method of image acquisition[19]. The final goal of this work is to obtain a confident coin image match with minimal overlap at a reasonable speed.

From the image processing viewpoint - a universal method is finding and checking the coin image by cross-correlating it with expected (reference) image(s). Correlation is widely used as an effective similarity measure in matching tasks[20]. However, traditional correlation based matching methods are limited by various ways. Main challenge here (additionally to "pre-conditioning" of compared images) is the significant reduction of the computations, needed for finding cross-correlation values over 2-D space and for possible rotation values.

One general approach here could be using of "pyramid schemes", where variable resolution image versions are used in parallel. Alternative could be also "adaptive subsampling"[21] or using of some kind of "rotation invariant" approaches or "multi-match" techniques, to match

---

[17] M. Kampel, and S. Zambanini, "Coin Data Acquisition for Image Recognition", 36th Conference on Computer Applications and Quantitative Methods in Archaeology, Budapest, Hungary, April 2008.

[18] A. Khashman, B. Sekeroglu, and K. Dimililer, "A Novel Coin Identification System", Intelligent Computing in Signal Processing and Pattern Recognition: Lecture Notes in Control and Information Sciences, Vol. 345, 2006, Springer Berlin/ Heidelberg, pp. 913- 918

[19] Ch. Heipke, "Overview of image matching techniques", OEEPE Workshop on the Application of Digital hotogrammetric Workstations, OEEPE Official Publications, no.33, pp.173-189, 1996.

[20] F. Zhao, Q. Huang, and W. Gao, "Image Matching by Normalized Cross-Correlation", IEEE International Conference on Acoustics, Speech and Signal Processing, 2006, Proc., Toulouse, ICASSP 2006, 14-19 May 2006, pp.729-732.

[21] A. Ricardo ,F. Belfor, M. P. A. Hesp, R. L. Lagendijk, and J. Biemond, "Spatially Adaptive Subsampling of Image Sequences", IEEE Transaction on Image Processing, vol. 3, no. 5 (Sept), 1994, pp.492-500.

the object with any of many “candidate” image-versions in parallel[22].

Still, after practical experiments with various algorithms, cross-correlation has been taken as one good candidate for image matching, and ways to drastically simplify the calculation has been looked for.

An important feature of the proposed solution is that it imposes only minimal requirements on the image acquisition process. The scale detection is not restricted to a specific type of ruler and is robust with respect to the placement and orientation of the ruler[23]. Similarly, the coin segmentation works for a wide range of coins and backgrounds and is independent of the position of the coin.

4.1. Approach 1: adaptively subsampled cross-correlation

In the current work usage of variable-rate sub-sampled (by pixel blocks) reference images has been proposed and evaluated.

The aim of segmentation an image into parts in this system is to differentiate the region of interest from other region of the image. A special case have been considered, where most of the image blocks are sampled with zero (none) sample-values, while a few blocks (some to some tens) of the image has been sampled at the full accuracy (240 x 240 pixels, in the used examples). The criteria, used for selection of the “specific (unique) blocks” of the reference image has been the minimum value of the maximum cross-correlation of an block of samples (eg 20 x 20 or 30 x 30 pixels) against any other (shifted or rotated) block of the same coin image. So, blocks with relatively high value of “secondary “correlation peaks at shifting and rotating are not considered as “good” or “unique” ones.

This can be computed very efficiently by building an image pyramid with re-sampling between each level. Figure 7 shows the regions with “specific (unique) blocks” of a 2-Euro coin, as an example.

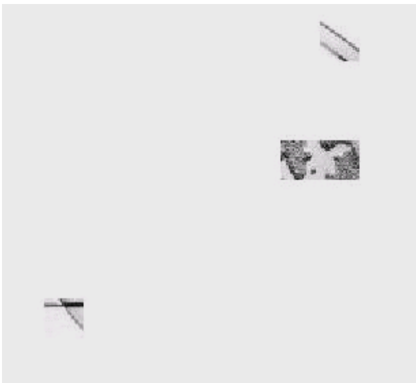
[22] Y. Lin, C. Chen, and C. Wei, “New method for subpixel image matching with rotation invariance by combining the parametric template method and the ring projection transform process“, Optical Engineering , vol. 45 (2006.), no.6 (June), pp. 067202.1- 067202.9.

[23] M.S. Islam, L. Kitchen, “Nonlinear Similarity Based Image Matching”, Intelligent Information Processing III, IFIP International Federation for Information Processing, vol. 228 (2007), pp. 401-410.

So, sub-sampled reference images for various euro-coins have been proposed, and corresponding algorithms has been evaluated. The computational complexity of proposed method is significantly reduced (with full-image correlation calculation), as only a small part of the image is considered for calculations.

a) 

b) 

c) 

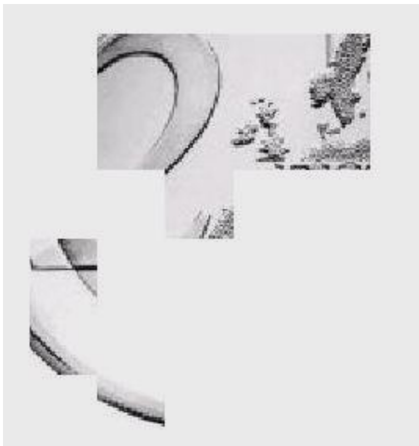
d) 

Figure 7 Adaptively subsampled images of 2 Euro coin (a- 10 blocks of 20x20 ,b – 20 blocks of 20x20 -both of side A, and c-4 blocks of 20x20, 20-blocks -side B).

4.2. Approach 2: SIFT analysis

Alternatively, using of set of feature-points of the images, as reference for cross-correlation, has been evaluated, for various EuroCoins (with corresponding determination of reasonable feature points, for various coins). SIFT (Scale Invariant Feature Transform) feature points[24] have widely been used for image matching and stitching. SIFT method is invariant to image scaling, illumination changes and rotation. This makes it very suitable for coin matching (recognition).

Method is based on unique feature point calculation. For each point also features like scale and orientation are being calculated. Based on these features points on two images will be matched.

Figures 8 and 9 demonstrate image matching using feature points. On the first image (Fig.8) all coins have zero degree angle. Second image (Fig.9) represents two-euro coin which has been rotated 33 degrees. Sub-images on the right display 8 different euro coins, among others also two-euro coin. SIFT feature points have been calculated for both images. Best matches between two point sets are then calculated. Blue crosses on both images represent feature points on both images. All matched feature points are located on two euro coin. This indicates sufficient uniqueness of each point.

Feature point calculation could be used as alternative approach for cross-correlation matching algorithm. Rotation invariance makes it suitable for coin matching. More complicated processing makes it more demanding in computational power sense compared to correlation method. This method could be especially useful in cases where many coins must be recognized on a same image.

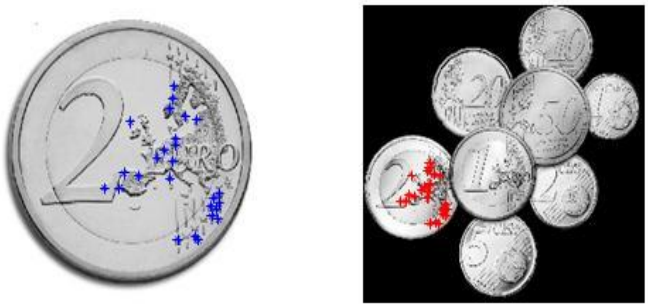


Figure 8 : Euro coin detection with SIFT feature method.

[24] D.G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision, vol.60 (2004), no. 2, pp. 91-110

Figure 9 : Rotated 2 Euro coin detection with SIFT feature method.

4.3. Approach 3: Local maximum and minimum difference based interesting pixel block selection

One of the simplest methods to select and determine if the pixel block has any interesting or unique features on it is the local maximum and minimum difference method. It is known that image pixel intensities in an 8 bit image vary between 0 to 255 values. If all the pixel intensities would be with the same values or with minimal pixel intensity variance then the contrast inside the pixel block is not very large and therefore the pixel block is not very characteristic or interesting. However, if the difference between the intensities inside the pixel block is relatively large then the pixel block is definitely characteristic and offers us more of an interest.

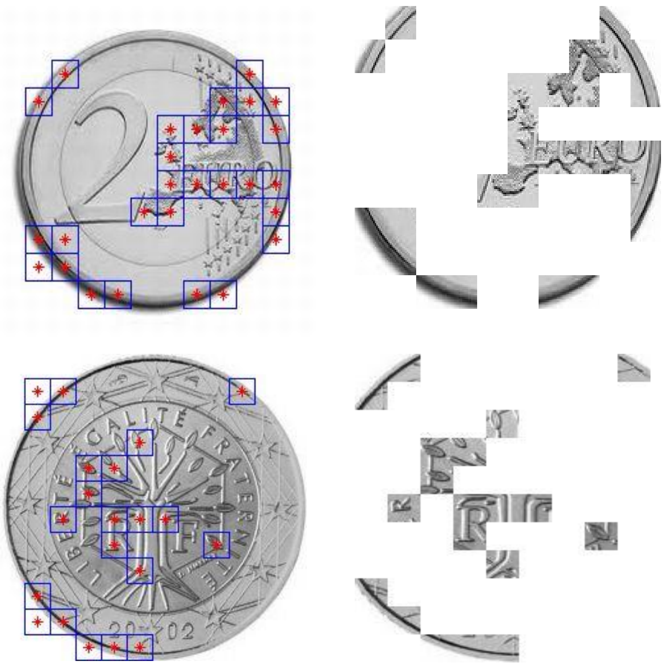


Figure 10 Adaptively sub-sampled images of 2 Euro coin based on local

maximum and minimum differences (left upper image- selected pixel blocks on original image front side, left lower image- sub-sampled image front side, right up - selected pixel blocks on original image front side, right lower - sub-sampled image back side).

In order to find an interesting pixel blocks the reference image was divided into 20 x 20 pixel blocks. In every pixel block the maximum and the minimum pixel intensity values were found. The difference between the maximum and the minimum intensity values were calculated.

Threshold level for the pixel intensity difference was set to distinguish the most interesting pixel blocks from all the pixel blocks. In figure 10 a threshold level of 0.8 was used. Local maximum and minimum difference based interesting pixel block selection method is the simplest method to determine interesting pixel blocks. Also the calculation power that is needed compared with the cross-correlation based pixel block selection is marginal.

5. Object Recognition Algorithms

This section summarises the most popular contemporary object recognition algorithms. Emphasis is placed upon algorithms that are used in 2D object recognition and would therefore be suitable for the requirements posed by the particular set-up met within the context of SAFEMETAL.

5.1. Global Methods

Before moving onto listing the principal object recognition algorithms that are categorised as global methods, a brief explanation for the meaning of this term should be provided. When working with global methods, the object model represents the object to be recognised as a whole. The example image of the model, or *"template"* as is many times also referred, is reated as a single data set containing several global characteristics of the object. Object detection occurs by correlating the template with the content of a scene image. This method is relatively simple, which makes it easy to implement but vulnerable to a number of challenges. This section describes algorithms that work according to the premise of global methodology along with some algorithms that have been proposed as solutions to the limitations posed by this kind of approach.

5.1.1. Correlation & 2D Correlation

This can be considered as one the most straightforward approaches to object recognition. 2D cross-correlation is a slight expansion to original cross-correlation. The correlation technique is an effective method for comparing signals with one another[25] since it can offer a quantitative measure for the shared property between the two. The two signals are superimposed and slid against each other one sample at a time, **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**1. In other words, correlation is a repetitive process, while one series is delayed in time relative to the other. For each time shift, a multiply and accumulate operation occurs through all adjacent data points. The end product is a measure of similarity between the two signals.

[25] K. G. Beauchamp, Transforms for engineers: A guide to signal processing, Oxford University Press, 1987

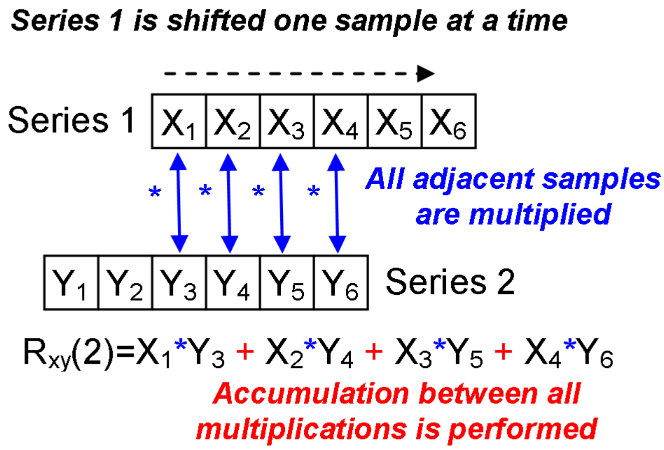


Figure 11 : The cross-correlation function

Specifically, given two discrete signal series, x_i and y_i, their cross-correlation function, $R_{xy}(n)$, is given as:

$$R_{xy}(n) = \frac{1}{N}\sum_{i=1}^{N-n} x_i \cdot y_{i+n}$$

where $n = 0,1,\dots, M-1$ and M is the total number of correlation delays to be used. N is the total number of samples in each data series normalising the correlation coefficient to a maximum value of 1.

Apart from cross-correlation there are other variations such as circular correlation, for periodic signals, and auto-correlation. Auto-correlation is the same process as cross-correlation only this time a signal is correlated with itself:

$$R_{xx}(n) = \frac{1}{N}\sum_{i=1}^{N-n} x_i \cdot x_{i+n}$$

The cross-correlation and auto-correlation functions reveal information about the frequency domain of discrete time signals[25]. For example, the auto-correlation function of a noise signal x_i is nearly zero at all times apart from t = 0. The noise signal x_i is uncorrelated and with a small time shift the similarity between x_i and x_{i+t} is destroyed, Figure 12

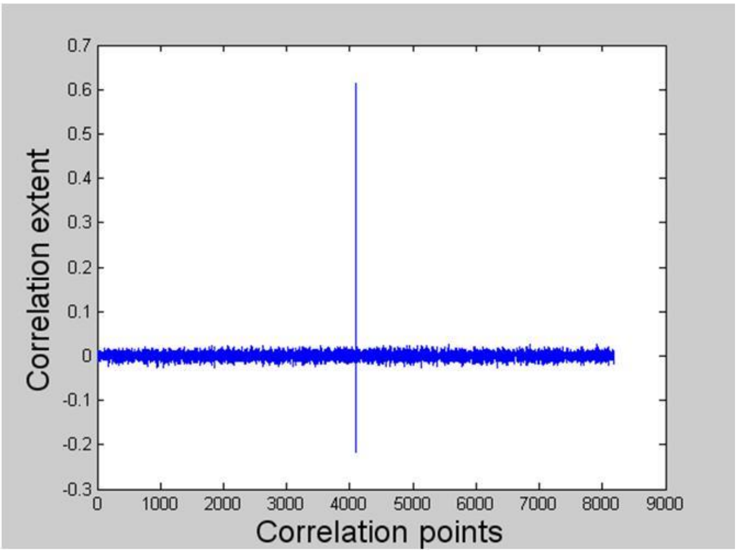


Figure 12 : Auto-correlation of noise

Similarly, the auto-correlation function of a periodic function, such as a sinusoidal signal, persists over a range of time delays, showing cyclic similarity between different values. An additional characteristic of the correlation technique is that it can improve the Signal-to-Noise Ratio (SNR) of a signal obscured by noise[25,26]. Therefore, correlation is also used for filtering purposes, since it simplifies signal detection.

Two-dimensional (2D) correlation operates along the same lines but this time in the spatial domain. For instance, in the case of image processing, a template image, representing the object to be detected or recognised, is compared against a scene image that has been obtained. Depending on the application demands, one could superimpose the two images and calculate the correlation coefficient. In case a slightly more detailed approach is required, the template represents a part of the whole object and is therefore cross-correlated with different parts of the scene image, **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**. For instance, the example shown in Figure 13 uses a 3x3 pixel template with an 8x8 pixel image. The starting position is located at the top left of the scene image. The template is shifted one pixel to the right and each time the 2D correlation coefficient is calculated. When the end of the row is reached, the template is brought back to the leftmost position but one row down and the process repeats itself until the bottom right image corner is reached.

[26] Y. W. Lee et al., Application of correlation analysis to the detection of periodic signals in noise, Proc, IRE38, 1950

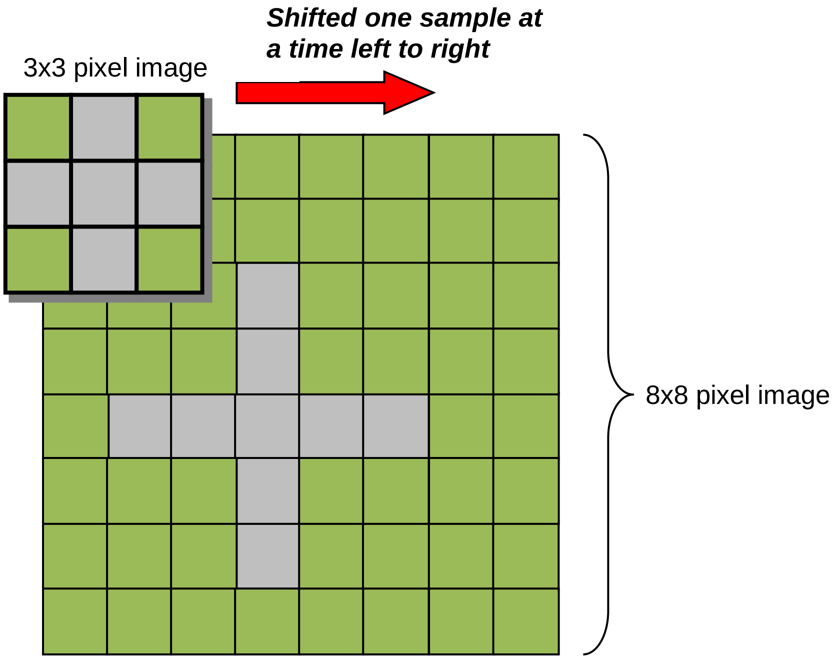


Figure 13 : Two-dimensional (2D) Correlation

Mathematically, the 2D correlation operation can be described as follows:

$$R_{xy} = \frac{\sum_{i=0}^{N}\sum_{j=0}^{M}\left(x_{i+a,j+b} - \bar{x}\right)\cdot\left(y_{i,j} - \bar{y}\right)}{\sqrt{\sum_{i=0}^{N}\sum_{j=0}^{M}\left(x_{i+a,j+b} - \bar{x}\right)^2 \cdot \sum_{i=0}^{N}\sum_{j=0}^{M}\left(y_{i,j} - \bar{y}\right)^2}}$$

Variables x and y represent pixel brightness for the collected image and template respectively and the cross-correlation coefficient becomes very high when pixels of increased brightness overlap. In the simple case of cross-correlation, variable N acted as the normalisation factor. In the 2D correlation analysis, the denominator of the equation above is responsible for normalising the result and all too often, but not necessarily always, this corresponds to values for R_{xy} between -1 and 1. Naturally, a local maximum of this matching function would indicate a possible occurrence and usually we set a threshold value for the correlation coefficient. Any value above that is then categorised as an occurrence and any value below as a lack of such an event.

This method can be used with relative ease and compared to the majority of object recognition algorithms it is one of the most straightforward. However, 2D correlation is hindered by a number of drawbacks that become apparent as the application complexity increases. As a result, several improvements have been developed such as that of *edge-detection filtering* prior to correlation. This type of filtering helps in bringing forward object outlines and, in general, parts of the image where the pixel intensity gradient is big. Hence, the original method is expanded by introducing a filtering stage prior to the 2D correlation phase.

Another approach is that of the *image pyramids*. In this method, the original template is sub-sampled into a number of new images that contain less and less information but, for the same reason, are faster to correlate with the scene image. This way, parts of the scene image that contain entirely unrelated information to that of the template can be excluded relatively quickly allowing for a time consuming 2D correlation to take place only over image regions that are likely to give out a match.

5.1.2. Global Feature Vectors

Global characteristics of an object can also be described using feature vectors. These multi-dimensional vectors contain information related to an object and, hence, allow for a unique object class to be extracted. Feature vectors can represent unique points in a feature space but with the ultimate aim being the best distinctiveness between object classes, various global characteristics may be used. These can be area, moments, mean grey value, Fourier descriptors and others. At this point, it is beneficial to expand a little on *moments* and *Fourier descriptors*.

Naturally, with Fourier descriptors, feature values are derived in the spectral domain and after a Fourier transformation has been applied on the object image. Fourier descriptor refers to the feature vector calculated from the spectral domain data representation. It is usually more robust against noise and minor boundary modifications.

A good example is found in Treiber[1] based on one of the fiducial shapes used in Printed Circuit Board (PCB) design. Those shapes are used for accurately positioning the PCBs and the approach here is to extract the geometry for one type of fiducial shape and then generate a corresponding classification vector. The complete process is also shown in Figure 14.

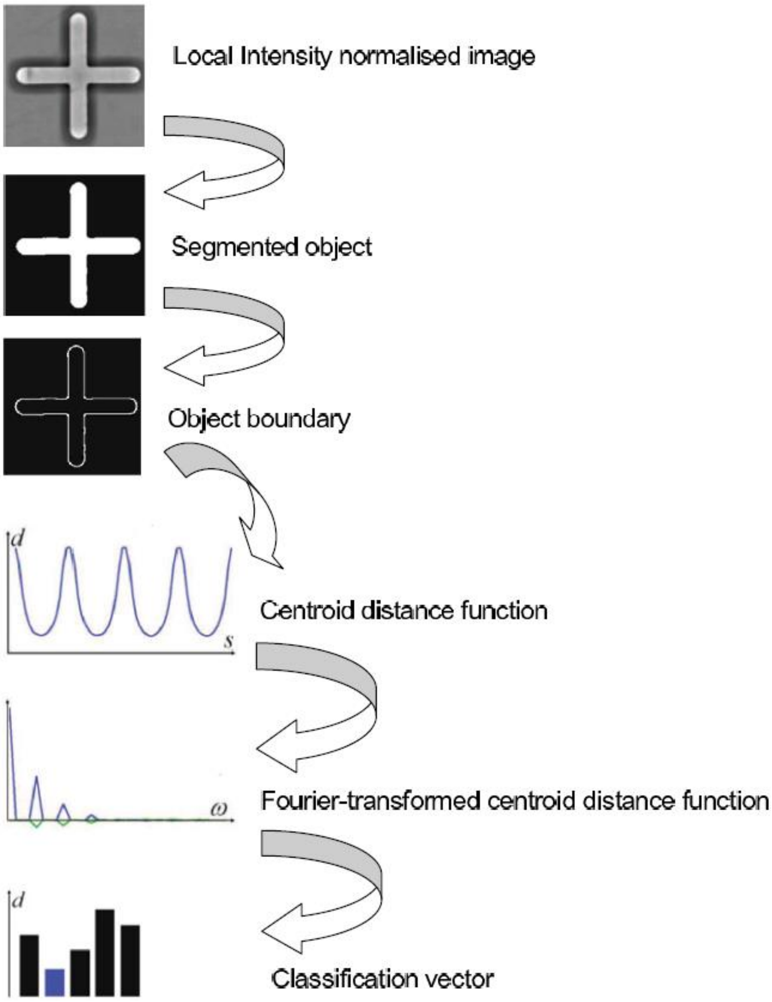


Figure 14 : Extracting feature values using the Fourier Transform

First, a distinction is made between the object to be classified and its background. The aim is to establish the distance between all contour points of the object from its centre. This gives rise to a "distance function" shown on the bottom right of Figure 14. Next, this one-dimensional representation of a two-dimensional image, is represented in the frequency domain using the Fourier Transform. Note that by doing this, a very compact representation of the data is obtained. Details on how the Fourier Transform operates are provided in a later section (Transformation-based methods). In this particular example, the shape of the object has helped in generating a relatively straightforward frequency spectrum representation due to the almost ideal sinusoidal behaviour of the distance function.

The fixed-length distance function is one feature vector and one can appreciate how complex objects can in fact be represented by a number of different feature vectors of this sort. Back in the example of Figure 14, the Fourier-transformed feature vector is compared against a number of "prototypes", Fourier-transformed feature vectors representing a number of geometric forms, and a match is established. It should be noted that an improved modification to the Fourier descriptors worked out from an [x,y] representation method is that of polar Fourier Spectra but this is beyond the scope of this report.

Region or grey-value moments are also used to characterise and classify objects according to their global features. The analysis is applied on an image based on the following prototype equation:

$$m_{pq} = \sum_{(x,y)\in R} x^p \cdot y^q$$

Coefficients *x* and *y* allow to investigate low-order, medium and high-order (defined by coefficients *p* and *q*) moments with respect to all image pixels. The previous formula is usually applied in a more complex fashion, such that the moment coefficients are normalised or they are calculated relative to their centre of gravity, however, this is beyond the principal aim of this report. The key issue here is that all image moments related to a specific object, i.e. from m_{00} to m_{xy}, can be integrated into a feature vector which can then act as reference during an object recognition process.

5.1.3. Principal Component Analysis (PCA)

This analysis is based upon a fundamental observation made by Murase and Nayar[27] stating that *"the appearance of an object in an image depends on the object shape, its reflectance properties, the object pose, and the illumination conditions"*. By keeping track of those parameters, we gain an advantage especially with respect to minimising the risk of false mismatch between a template object that is also located in a scene image due to a change in pose or illumination. Hence, one way is to work out the more stable features and focus on them, alternatively, develop a model that contains all expected appearance variations.

Now for a thorough model to be developed an exhaustive training phase has to be performed. Nevertheless, in order to minimise redundant data a better suited representation is needed. This is where the PCA transformation comes in and it aims at representing the variations of the object's appearance with as few dimensions as possible.

Specifically, with PCA, image patterns can be identified allowing for data compression to take place without much loss of information. The rather complex mathematics are once again beyond the scope of this report, however, the steps followed throughout this method are as

[27] H. Murase et al., Visual learning and recognition of 3D objects from appearance, Computer

follows. First, for each dimension, its mean value is subtracted, e.g. for a two-dimensional image, the two means x and y are taken out giving two data sets with an average of zero. Second, the covariance matrix is calculated, in our case a 2 x 2 matrix, and its eigenvectors and eigenvalues are also calculated. The eigenvectors are then ordered by eigenvalue, highest to lowest, and this gives the components (eigenvectors) in order of significance.

Finally, the low eigenvalue components (eigenvectors) can be left out giving a data set with fewer dimensions than the original. The retained eigenvectors can then be ordered in a vector matrix to produce the *feature vectors* required during object recognition analysis.

5.1.4. Summary

This section gives a short summary, Table 1, of the techniques investigated in previous sections that are commonly known as Global Methods used in object recognition analysis. Focus has been shown on the main advantages and disadvantages of those methods, which will assist in deciding on the one used in SAFEMETAL.

Table 1: List of Global Methods

	Pros	**Cons**
2D correlation	*Simple, fast and easy to implement* *Can be applied to any kind of object*	*Not invariant to rotation and scale* *Sensitive to clutter and occlusion*
Global Feature Vectors	*Fast* *Compact object recognition*	*Requires a pre-processing stage* *Compactness lost when dealing with complex objects*
PCA	*Strong invariance to same-object appearance variations* *Efficient over a wide range of objects*	*Consider number of sample images needed* *Unable to detect multiple objects in an image* *Scene object has to be separated from background*

5.2. Transformation-Search Based Methods

The Global Methods listed in the previous section can be thought of as the easiest to implement, however, they suffer from serious drawbacks when faced with serious occlusion and background clutter. It is difficult for those methods to be invariant to such challenges and

Vision, 14:5-24, 1995

that is why more sophisticated techniques have been used. One such technique is the transformation-search based method and in place of the transformation algorithm used there are more than just one option. An added benefit to using the transformation-search based method is that image data can be checked locally and not necessarily only as a whole, i.e. in Global Methods any discrepancy between model and scene image affects the global features. Hence, it is possible to concentrate on characteristic parts of the object we want to detect.

The way this method works is relatively straightforward to summarise. This time object models consist of a finite set of points and their position. What is then needed is a model point set for the object to be recognised and a second point set from a scene image. A transformation algorithm is then applied in order to assess the similarity (or difference) between the two point sets through the generation of a set of transformation coefficients. The good thing about this method is that a *degree* of similarity can be established that guarantees recognition on one hand but allows for a level of dissimilarity to exist on the other. Such dissimilarity could simply be due to the presence of occlusion and/or background clutter, which means that the unwanted effect of those two parameters is in a sense filtered out.

5.2.1. Affine Transformation

Typically, non-complex applications, i.e. three-dimensional objects with very short depth, are treated as two-dimensional objects using a transformation class known as the *affine transformation* shown in the following equation:

$$X_{S,i} = \begin{bmatrix} x_{S,i} \\ y_{S,i} \end{bmatrix} = \mathbf{A} \cdot x_{M,i} + \mathbf{t} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \cdot \begin{bmatrix} x_{M,i} \\ y_{M,i} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

It is evident that this is a linear transformation, which greatly simplifies the mathematical complexity and calculations needed. Matrix **A** and vector **t** hold all six parameters used in order to establish the closeness between the position of a point or feature $x_{S,i}$ in the scene image and its corresponding model position, i.e. $x_{M,i}$.

A significant reason for using a transformation-based approach has to do with computational complexity. In some applications, object detection can be performed in the spatial or image space, an approach, however, that may prove to involve a significantly large computational effort. Some of those techniques work in a different domain called the frequency space such as the Fourier Transform and the Wavelets. Frequency-based transforms are used extensively in computer vision (and not just as a means of detecting objects), for instance, data compression and image enhancement. The most famous of all is the Fourier Transform, which is partly owed to the highly efficient algorithm developed by Cooley and Tukey in 1965[28]. Other versions exist, such as the Fast Hartley Transform (FHT)[29] or the Fast Cosine/Sine Transform (FCT/FST)[30].

5.2.2. Fourier Series & Fast Fourier Transform

In this report, the Fast Fourier Transform, shall be first described using a one-dimensional waveform followed by an additional example of a two-dimensional image. The Fourier series representation and the Fourier transform can be defined as the mathematical tools to map between the time and frequency domain[31]. The Fourier series representation is used for periodic signals only while the Fourier transform is used for both periodic and non-periodic signals. The Fourier transform can represent any periodic time domain signal by a summation of harmonically related sinusoids. Thus, a perfect sine wave signal must have only one frequency component with a frequency equal to that of the input sine wave. The representation of a time domain signal by a summation of sine or cosine components is referred to as the spectrum of that particular signal. The spectrum is drawn as discrete lines on a graph, with the position of a line on the x-axis representing the particular frequency component of the signal and the height of this line representing the amplitude, Figure 15.

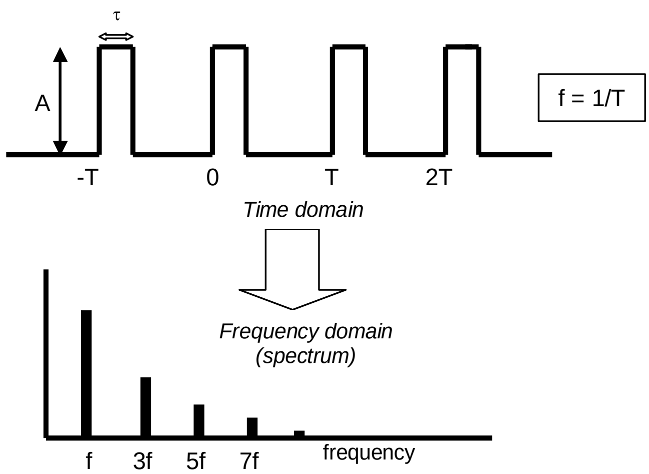


Figure 15 : Amplitude – Frequency spectrum analysis

[28] J. W. Cooley et al., An algorithm for the machine calculation of the complex Fourier series, Mathematics of computation, 19:297-301, 1965
[29] R. N. Braswell, The Fast Hartley Transform (FHT), Proc. IEEE, 1984
[30] W. H. Chen et al., A fast computation algorithm for the Discrete Cosine Transform (DCT),

With the introduction of inexpensive digital signal processing devices, the FFT has become a commonly used measurement technique. The FFT is a fast way of computing the discrete finite Fourier transform of a signal. The FFT determines the amplitude of a particular frequency sine wave or cosine wave in a signal by multiplying the signal, point by point, with a unit amplitude sine wave. The result is averaged over an integer number of sine wave cycles and if the sine wave is not present in the signal being analysed, the average will tend to zero. The Fourier transform of a signal *x(t)*, continuous for all time, is defined by the following equation:

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-j \cdot 2 \cdot \pi \cdot f \cdot t} dt$$

whereas the discrete finite Fourier transform (DFT), used for sampled data, is:

$$XD(f) = \sum_{m=0}^{M-1} X(m\Delta t) e^{-j \cdot 2 \cdot \pi \cdot f \cdot m\Delta t} \Delta t$$

The difference between *X(f)* and *XD(f)* is that of spectral resolution. *X(t)* has infinite spectral resolution while *XD(f)* has a discrete frequency resolution of

$$f = \frac{1}{M \cdot \Delta t}$$

due to the finite number of points in the data record. However, a problem emerges when employing FFT for signal analysis. Since all hypothetical sine and cosine frequencies in the FFT are multiples of the reciprocal of the waveform length, the analysis is of equal resolution in the frequency domain. Furthermore, the FFT assumes that the waveform being analysed is periodic, with a period equal to the length of the data record being analysed. Hence, sharp discontinuities at the points where the start of one record joins the end of the preceding record, cause the spectral components of *X(f)* to be spread or smeared in *XD(f)* [32,33,34], **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**.

IEEE Trans. Communications, COM25:1004-1009, 1977

[31] A. Bateman, Digital communications, Addison-Wesley, 1999

[32] R. C. Cabot, Fundamentals of modern audio measurement, Audio Engineering Society, 47, 1999

[33] M. Mahoney, DSP-based testing of analogue and mixed-signal circuits, IEEE Computer Society, 1987

[34] B. E. Peetz, Dynamic testing of waveform recorders, IEEE Trans. Instrumentation and Measurement, IM-32:12-17, 1983

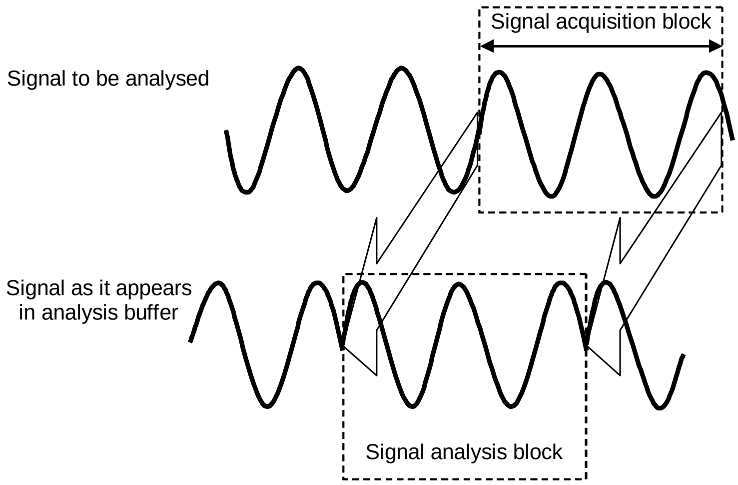


Figure 16 : The problem of discontinuity

To overcome the effect of discontinuity between data records, window functions have been developed[35]. The windowing function basically drives the data record values at the end points to zero. There are a large number of different window functions but only a few are most commonly used. Amongst those regularly used are the Hann, the Hamming and the Blackman-Harris windows. Naturally, each different windowing technique has its own advantages and disadvantages but these are beyond the scope of this section. Another major characteristic of this technique concerns the Nyquist criterion. If *X(f)* contains components that exceed the sampling frequency, i.e. the Nyquist frequency *Fs/2*, then these components are folded back, or aliased, onto spectral lines below *Fs/2* causing aliasing errors. Therefore, the sampling frequency *Fs* is always chosen to be at least twice the bandwidth of the input signal, *2*BW*, to avoid aliasing errors and the spectrum of *XD(f)* is displayed only from DC to *Fs/2*.

Extending the one-dimensional Fourier Transform to two (or more) dimensions, one would simply have to change from *x(t)*} to *x(i,j)* and perform the integral (or summation) over two (or more variables) instead of one. For demonstration purposes, a simple example is used showing two sinusoidal variations in brightness in the image space with their corresponding two-dimensional Fourier Transform response in the frequency domain, Figure 17.

[35] F. J. Harris, On the use of windows for harmonic analysis with the discrete fourier transform, IEEE Proc., 66:51-83, 1978

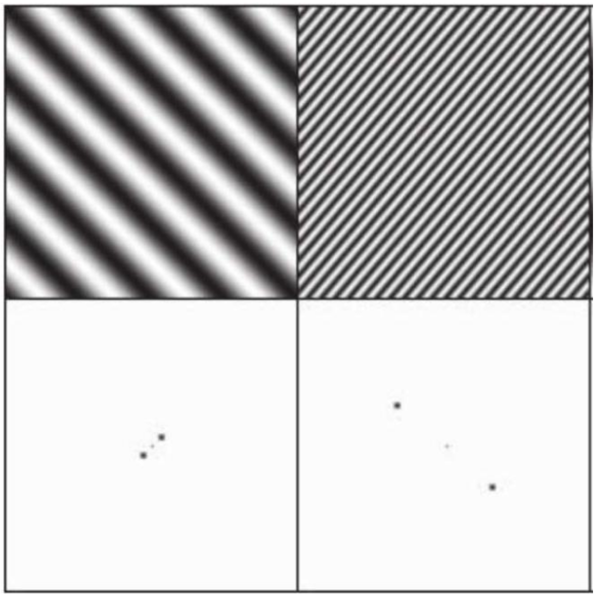


Figure 17 : Two sinusoidal brightness patterns differing in frequency and direction and their corresponding 2D Fourier Transform

The two sinusoids of Figure 17 are different in terms of their frequency and direction. What they have in common, however, is that both can be reconstructed using a single tone. This is clearly evident in their corresponding frequency domain response plots, where each sinusoidal variation is represented by a single point. The nature of the Fourier Transform, however, generates redundant data and typically this shows as an inverted mirror image of the top half of the frequency domain response. Hence, a second point present in the frequency domain. Note that the position of the points reveals both the direction of the brightness variations and their frequency, i.e. the closer/farther to the centre (DC) of the frequency domain response the lower/higher the frequency.

It must be noted that Fourier is simply a tool for image processing and, therefore, in the context of object recognition it must be complemented by some sort of template matching functionality. Generating the frequency domain response for a scene image is only the first part of the object recognition process since the frequency domain response of a model image would then have to be compared against it.

5.2.3. Wavelet Transforms

Wavelets are also extensively used in image processing and as part of object recognition schemes. Hence, the underlying theory is described in this section. Essentially, the wavelet is a transform with basis functions that are localised in frequency, i.e. vary with frequency. The big advantages of this method over the FFT are mainly the fact that the temporal locality of its basis functions offers good time resolution. Additionally, it is of lower computational complexity. For a data stream of size *N*, the arithmetic operations needed by wavelet transform and FFT are proportional to *N* and *NlogN* respectively.

The novelty about wavelets is that unlike windowed Fourier analysis, where the length of the sampling window is kept constant, wavelet methodology keeps the number of oscillations in the window constant, and instead stretches or compresses the length of the window. The two fundamental properties that define wavelets are *dilation* and *translation*. The dilation property allows the wavelet to be stretched or compressed to satisfy the fixed number of cycles constraint. The translation property allows the wavelet to shift in time, and obtain a matching phase with the signal. The relationship between these two properties and wavelets is defined by the following equation:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \cdot \Psi\left(\frac{t-b}{a}\right)$$

The scaling parameter, *a*, measures the degree of compression or scale, and the translation parameter, *b* determines the time location of the wavelet. As expected, parameter *a* determines the size of the wavelet. Also, there is an association between the time-width of a wavelet, i.e. its scale, and its frequency. This can be defined as a wavelet time-width adaptation to their frequencies. If the absolute value of *a* is smaller than unity, $|a| < 1$, the wavelet, $\Psi_{a,b}(t)$, becomes a compressed version of the mother wavelet, $\psi(t)$, and corresponds mainly to higher frequencies. For the opposite case, where $|a| > 1$, the wavelet obtains a larger time-width compared to the mother wavelet and corresponds to lower frequencies.

Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε. illustrates a typical mother wavelet together with two of its possible modifications, depending upon the already-mentioned parameters *a* and *b*.

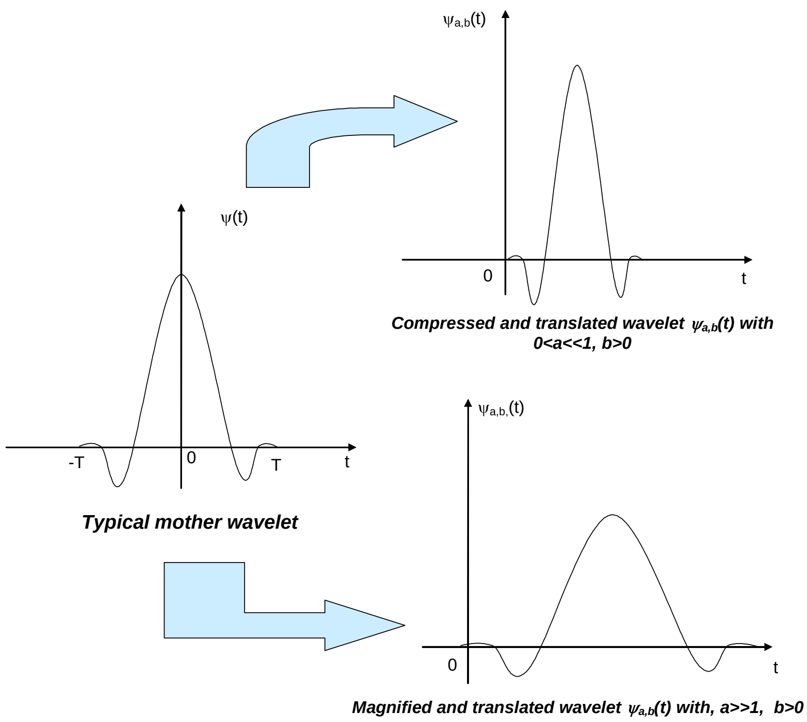


Figure 18 : A typical mother wavelet with a compressed and translated wavelet and a magnified and translated wavelet

Different values for parameter *b* represent the time localisation centre, and each resulting wavelet, $\Psi_{a,b}(t)$, is localised around the centre *t* = *b*. As the scale parameter *a* varies, the resulting wavelet covers different frequency ranges. Small values for $|a|$ result in very narrow windows and correspond to high frequencies. On the other hand, very large values of $|a|$ result in very wide windows and correspond to low frequencies.

Since the discovery of the first wavelet transform, a large set of Discrete Wavelet Transforms (DWTs) has been developed. Some of the most well known wavelet basis functions are the Morlet wavelet, the Mexican Hat and the Daubechies wavelet[36]. The procedure followed for the implementation of a DWT can be mathematically expressed as:

$$y[n] = \sum_{k=-\infty}^{+\infty} h[k] \cdot x[2n-k]$$

where *y[n]* are the resulting wavelet coefficients, *x[n]* is the input signal and *h[k]* is a filtering function[37]. The DWT analyses the signal at different frequency bands with different

[36] L. Debnath, Wavelet transforms and their applications, Birkhauser, 2002
[37] R. Polikar, The wavelet tutorial: Parts i to iv, 2001 (http://users.rowan.edu/

resolutions. It recursively applies a low-pass and high-pass filter on the time domain input signal, achieving signal decomposition into different frequency bands. The two filters, high-pass *g[n]* and low-pass *h[n]*, are initially applied on the respective half-bands of the input signal. After the filtering, the high-pass filter coefficients are stored and attention is drawn to the low-pass coefficients. The signal is then sub-sampled by 2 by discarding every other sample, hence, constituting the first level of decomposition in the DWT expressed by the following pair of equations[37]:

$$y_{high}[k] = \sum_n x[n] \cdot g[2k - n]$$
$$y_{low}[k] = \sum_n x[n] \cdot h[2k - n]$$

where *y[k]* high and low are the outputs of the high-pass and low-pass filters after sub-sampling by 2. By using half the number of samples to characterise the signal on the low-pass filter output the time resolution is halved and, hence, the frequency resolution doubled. The above procedure is then repeated for further decomposition. At every level the filtering and sub-sampling results in half the number of samples, i.e. half the time resolution, and half the span of the frequency band, and hence doubles the frequency resolution. Figure 19 shows diagrammatically the above procedure. The *h[n]* and *g[n]* blocks represent the low-pass and high-pass filtering of the DWT coefficients respectively. F_b represents the frequency bandwidth in radians.

polikar/wavelets/wttutorial.html)

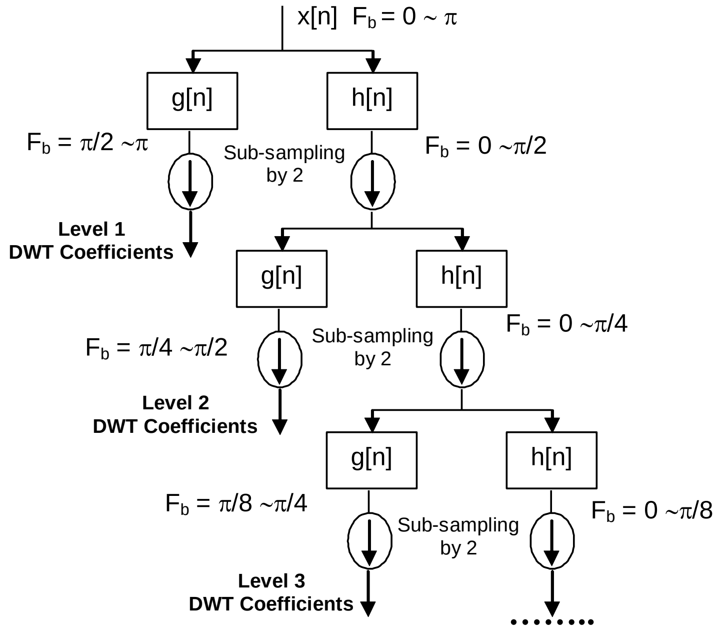


Figure 19 : A typical Discrete Wavelet Transform (DWT) algorithm

The number of levels of decomposition is dependent upon the size of the signal to be analysed. For example, a 256 sample signal would have 8 levels of decomposition and a 16384-length input would have 14 levels of decomposition. Generally, it obeys the relationship $N = 2n$, where N is the size of the input data and n the number of decomposition levels.

Like the FFT, it converts time-domain data into the frequency domain. However, unlike the FFT, the wavelet transform assumes that the frequency spectrum is changing over time. The result of a wavelet transform is harder to interpret, but provides more information. The algorithm does not yield a straightforward frequency spectrum, i.e. double resolution, in frequency and in time. This forces the user to find the trade-off where the resolution is satisfactory in both time and frequency domains. In this work all coefficients have been set to focus on the frequency domain.

5.2.4. Hough Transform & Hausdorff Distance

A discussion on transformation-based methods for object recognition, however, remains incomplete without mentioning its best two representatives, i.e. the generalised Hough transform and the Hausdorff distance. A pre-condition for the Hough transform is that the object shape to be detected must be known in advance. The Hough transform works by transforming complex patterns of pixels in the image domain into compact features in a chosen parameter space. The result is such that many points in the image space map to single points in the parameter space. To explain this further we shall use the Straight Line Hough Transform, which is the simplest version for this transform since there are also the Circle Detection Hough Transform and the Generalised Hough Transform that are used in more complicated applications.

A parameter space *(m,c)* is defined and in this space any line image can be described using the slope-intercept equation:

$$y = mx + c$$

Hence, the image space is transformed into the *(m,c)* space and now line detection becomes a problem of point detection. The transformation is then applied to each individual point in the image and this way a divergent concentration of points is observed in the parameter space. Specifically, for each point in the image space a line is generated in the parameter space. Local maxima are defined as the points of intersection between lines in the parameter space. Finally, points of intersection (or points of local maxima) in the parameter space reveal lines in the image space.

The original form of the Hough transform was proven to be computational intensive and for that reason it has evolved through the years under different variants. A notable variant is that of circle detection using the Hough transform. As opposed to the previous case where straight lines within an image are detected, this time the presence of circles is targeted. The method of operation for this technique is better explained using the aid of **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**.

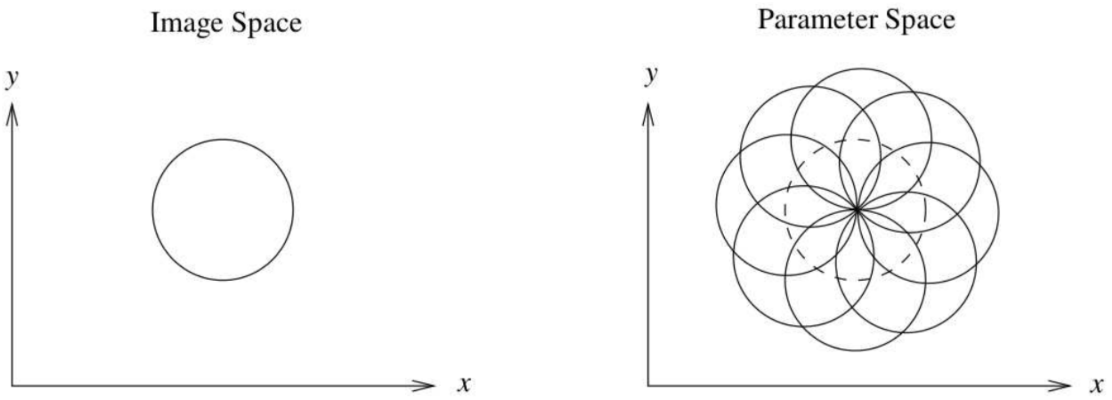


Figure 20 : The Hough transform for circle detection

In order to detect a circle of a particular radius, we plot a circle of the same radius for all edge segments in the image space. In Figure 20, there exist a number of intersections, however, the point of local maxima can be clearly distinguished, pointing to the centre of the original circle we were trying to detect.

Nevertheless, both previous methods have nothing to offer when the object of interest for detection is neither a straight line nor a circle. The most common method based on the Hough transform for detecting arbitrary shapes is the Generalised Hough transform[38]. The generalised Hough transform is an expansion on the Circle Detection Hough transform. This time a shape is plotted around a point in the parameter space not at a fixed but at a variable distance *R(θ)* along a line that is angularly displaced from the normal by a variable angle *a(θ)*, **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**.

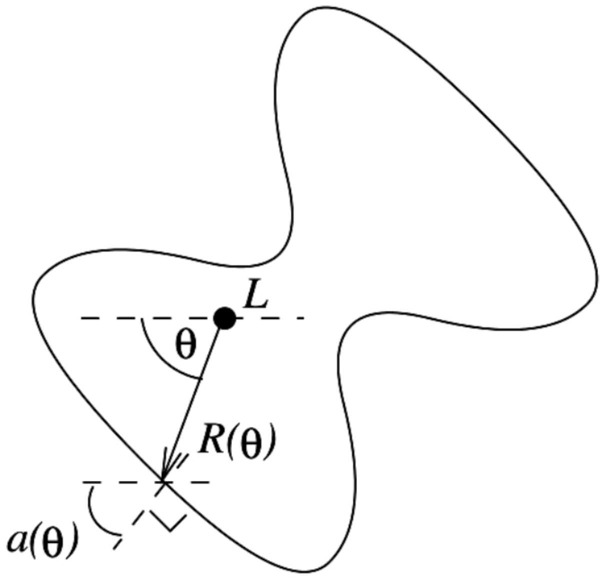


Figure 21 : The generalised Hough transform for arbitrary shape detection

Values for θ are stored in what is commonly known as the R-table and it is up to the values for θ, in other words on how big the table is, to successfully detect an arbitrary object.

The Hausdorff distance is an alternative transformation-based method to object recognition. It is a mathematically complex technique and beyond the scope of this report to go into details about its functionality. It will suffice to state its major principles of operation and its strong and weak points summarised at the end of the transformation-based methodologies section.

The concise definition for the Hausdorff distance is *"the shortest distance of a set to the*

[38] P. V. C. Hough, Method and means for recognising complex patterns, 1962, US Patent No. 3069654

nearest point in the other set"[39] . When applied to object recognition, one point set represents the model whereas the second represents the content of a scene image.

5.2.5. Summary

This section gives a short summary, Table 2, of the Transformation-Based techniques for object detection analysis. It should be stated that with respect to the SAFEMETAL demands using such a technique would not be entirely necessary since their algorithmic complexities do not reflect the level of complexity/difficulty of the application.

Table 2: List of transformation-based methods

	Pros	**Cons**
Affine Transformation	*Simple, fast and easy to implement*	*Highly limited with respect to shapes that can be detected*
Fourier Transform	*Fast and flexible*	*Needs to be complemented* *Can be computationally demanding*
Wavelet Transform	*Fast and more flexible than FFT*	*Needs to be complemented*
Hough Transform	*Can compensate for occlusion and data outliers*	*Significant memory requirements* *Long execution times*
Hausdorff Distance	*Significant range of objects that can be addressed* *Robust to occlusion and clutter*	*Dependency on illumination* *Prone to "false alarms" i.e. false identifications*

5.3. 3D Object Recognition and Other Approaches

This last section of the review is concerned with approaches that are fit for applications quite divergent from the particular case found in SAFEMETAL, i.e. two-dimensional image detection of a coin photographed under controlled ambient conditions and with no additional objects apart from the one of interest. As will be shown, the techniques outlined in this section, are developed with specific applications in mind and the reason they are mentioned in this report has more to do with presenting a complete study rather than considering employing one of those methodologies for the SAFEMETAL demands.

The approaches summarised in this section belong to the relatively wide categories listed

[39] G. Rote, Computing the minimum Hausdorff distance between two point sets on a line

below:

- Three-dimensional object recognition
- Geometric correspondence-based approaches
- Flexible shape matching
- Interest point detection and region descriptors

5.3.1. Three-dimensional Object Recognition

Sometimes, being able to locate an object in a three-dimensional space, is an application requirement, e.g. lifting and moving objects. The way to work through such a situation is to either use groups of sensors that help in reconstructing the 3D image or, if the application allows it, project the three-dimensional image onto a two-dimensional plane and work on this thereon.

The methods presented up to now perform matching of a 2D model to the 2D camera image plane, i.e. there is an effort to match a 2D scene image onto a 2D model image and work out the degree of resemblance. Adding an extra dimension to this complicates things quite significantly. This time a mapping takes place between a 3D scene to a 3D model. A few methods exist for achieving this, however, one highly popular involves estimating the projection of an object's location in 3D space onto a 2D image.

The transformation (referred to as *projective* transformation) that maps a $[X,Y,Z]^T$ position of coordinates onto the [x,y] image plane is given by:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \mathbf{R} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \mathbf{t}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \cdot a/c \\ f \cdot b/c \end{bmatrix}$$

where **R** is a 3×3 rotation matrix, **f{t}** a 3D translation vector and *f* is determined by the camera's focal length. For the process to be complete, the 2D scene image must be matched wither to either a 2D model image or a 3D object model, depending on how the model data is saved.

under translation, Information Processing Letters, 38:123-127, 1991

It is quite common to meet conditions such as 3D object recognition from a 2D scene image in industrial applications where it is often possible to model parts based on their characteristic or prominent features. Also, these are conditions that can be controlled for the purpose of improving performance. The most popular methods used in the domain described just now are the *SCERPO System*[40], the *Relational Indexing*[41] and *LEWIS*[42].

5.3.2. Geometric Correspondence-Based Approaches

This is another category of approaches that are typically used in industrial applications. The object to be detected is known beforehand and its model image is based on key features specific to it, such as its silhouette. Detection takes place in **two stages** whereby all locations of the target features in a scene image are established and then a matching process between the features found and those of the model is executed. It should be noted that the matching phase is correspondence-based, i.e. one-to-one correspondences between model and image features are established and evaluated.

The detection algorithms make use of geometric features such as the object's contour which is modelled using a set of primitives, i.e.line segments, circular arcs and others. Specifically, the **feature detection** stage is mainly comprised of three steps. These are the *Edge point detection*[43], the *Edge point linking* and the *Feature extraction*[44,45]. There is a great deal more to be mentioned about those three steps, however, this is beyond the scope of this report and the interested reader is referred to the referenced bibliography.

When feature detection has been completed, a **matching process** commences with the aim of establishing the level of similarity between the extracted features and those of the model image. A typical approach used is what is known as Graph-Based Matching. This is based on the fact that the geometry of objects can be represented using graphs. Graph models do not allow for much of intra-class variation, however, this is sometimes desirable in industrial applications.

---

[40] D. G. Lowe, Three-dimensional object recognition from single two-dimensional images, Artificial Intelligence, 31:355-395, 1987

[41] M. Costa et al., 3D object recognition and pose with relational indexing, Computer Vision and Image Understanding, 79:364-407, 2000

[42] C. A Rothwell et al., Planer object recognition using projective shape representation, International Journal of Computer Vision,16:57-99, 1995

[43] J. F. Canny, A computational approach to edge detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 8:679-698, 1986

[44] U. Ramer, An interactive procedure for the polygonal approximation of plane curves, Computer Graphics and Image Processing, 1:244-256, 1972

[45] J. M. Chen et al., Segmentation of planar curves into circular arcs and line segments, Image and Vision Computing, 14:71-83, 1996

Matching then is performed using an *association graph* where the detailed description as to its operating principles can be found in Ballard[46]. The advantage of using geometrical graph matching lays with the fact that it is tolerant to occlusion and clatter. On the downside, it is computationally complex and, therefore, computational costs remain low only for simple objects.

5.3.3. Flexible Shape Matching

Flexible shape matching has been developed to address applications where considerable object class-variations of their shape are noted and algorithms based on rigid object models are no longer suitable. Deformable objects are, therefore, targeted using specialised algorithms which shall be discussed a little more extensively in this section.

A typical example of deformable objects is fruits. It is easy to appreciate how a human can easily identify a potato although its shape may vary considerably. A similar behaviour in computer vision can only be achieved using algorithms that can work with deformations in the object's shape. A popular method is based on *parametric curves* and it has spawned a number different algorithms. A parametric curve $v(s,\varphi)$ is written as:

$$v(s,\varphi) = [x(s,\varphi), y(s,\varphi)]$$

In the equation above, φ is a vector of model parameters that define the curve and s is a scalar that monotonically increases from 0 to 1 as the curve is traversed. Parameter vector φ offers the flexibility for the model curve to accurately follow the object's contour.

One method for applying the parametric curve approach is based on optimising the parameter vector φ such that the discrepancies between the curve and the object shape are minimised. This method is seen a segmentation technique since it does not classify an object found in a scene image, rather, it tries to locate the exact borders of a deformable object. Hence, they are usually combined with some additional signal processing method in order to construct a fully functioning object recognition scheme. Typical segmentation methods are the *Active Contour Models/Snakes* that can be found in Kass[47] and the *Contracting Curve Density algorithm (CCD)* described in Hanek[48].

---

[46] D. H. Ballard et al., Computer vision, Prentice-Hall, Englewood Cliffs, N. J., 1982

[47] M. Kass et al., Snakes: Active contour models, Inter. Jour. Computer Vision, pp. 321-333, 1988

[48] R. Hanek et al., The contracting curve density algorithm: Fitting parametric curve models to images using local self-adapting separation criteria, Inter. Jour. Computer Vision, pp 253-258, 2004

A second method, based on the parametric curve, follows a different approach whereby a similarity measure or distance metric is calculated between arbitrarily shaped curves. In other words, first, a curve approximating an object boundary is established, and then it is compared with curves that describe the contour of known object classes. The comparison is performed using different metrics such as the distance metric. Further information on this topic is given in the listed references[49],[50],[51].

5.3.4. Interest Point Detection & Region Descriptors

Real-life applications normally require an object recognition methodology that is more advanced than those that use either geometrical features or point sets characterising the object's contour (e.g. correspondence-based methods) or the global appearance of the object (e.g. correlation). In addition, global appearance methods are unsuitable since they cannot compensate for heavy background clutter and occlusion. What is needed in real-world applications is local evaluation of image information, which is a more challenging task, albeit, highly accurate at the same time.

The strategy that has been developed to meet those demands can be split into two stages[52]. The first step consists of the detection of characteristic "interest points". Secondly, "region descriptors", i.e. feature vectors, are derived representing the image information that is located around each interest point. Object recognition can then be performed by comparing both the region descriptors and their spatial configuration with the model database.

Naturally, a multitude of methods have been developed for the detection of interest points and the derivation of region descriptors. One key method that engulfs both tasks is called Scale Invariant Feature Transform (SIFT) described in the listed references[53,54]. Now in terms of region descriptor techniques, there are variants of the SIFT Descriptor[55,56], Differential-

[49] L. J. Latecki et al., Shape similarity measure based on correspondence of visual parts, IEEE Trans. Pattern Analysis and Machine Intelligence, 22:1185-1190, 2000

[50] P. Van Otterloo, A contour-oriented approach to shape analysis, Prentice-Hall Ltd., Englewood Cliffs, 1992

[51] F. Mokhatarian et al., Efficient and robust retrieval by shape content through curvature scale space, Intr. Workshop on Image Databases and Multimedia Search, Amsterdam, Netherlands, pp 35-42, 1996

[52] C. Schmid et al., Local grey value invariants for image retrieval, IEEE Trans. Pattern Analysis and Machine Intelligence, 19:530-535, 1997

[53] D. G. Lowe, Object recognition from local scale-invariant features, Intr. Conference on Computer Vision, Corfu, Greece, pp 1150-1157, 1999

[54] D. G. Lowe, Distinctive image features from scale invariant viewpoints, Intr. Jour. Computer Vision, 60:91-111, 2004

[55] Y. Ke et al., A more distinctive representation for local image features, IEEE Proc. Conference on Computer Vision and Pattern Recognition, pp 506-513, 2004

[56] K. Mikolajczyk et al., A performance evaluation of local descriptors, IEEE Trans. Pattern

Based Filters[57] and Moment Invariants[58]. Finally, alternative interest point detector solutions have been proposed and the most notable of those are the Harris and Hessian-Based Detectors[59], the FAST detector for corners[60] and the Maximally Stable External Regions (MSER)[61].

5.4. Conclusions

At the closing stage of this report, it is worth summarising the major characteristics of all key object recognition methodologies that have been addressed. This assessment, however, can only be offered at a coarse level since the actual performance of an object recognition algorithm is very much dependent upon the application characteristics. Nevertheless, this classification ought to offer a broad guideline as to which of the methodologies would be suitable for the application in mind, table 3.

Table 3: Overview of methods

Method	Computational speed	Object complexity	Intra-class variation	Occlusion/Clutter
Correlation	*Fast/Slow*	*Simple/Medium*	*Rigid*	*Low*
Feature vector analysis	*Very fast/Medium*	*Simple*	*Rigid*	*Medium*
Transformation-based	*Fast/Slow*	*Simple/High*	*Rigid/Medium*	*Medium/High*
3D recog./ invariants	*Fast/Medium*	*Simple/Medium*	*Rigid*	*Medium (Occl.)/ High (Clutter)*
Snakes	*Medium*	*Simple*	*Very flexible*	*Low/Medium*
SIFT	*Medium/ Very slow*	*Medium/ Very high*	*Rigid/Flexible*	*Very high*
Shape Contexts	*Medium*	*Medium*	*Flexible*	*High*

Analysis and Machine Intelligence, 27:1615-1630, 2005

57 J. Koenderink et al., Representation of local geometry in the visual system, Biological Cybernetics, 55:367-375, 1987

58 L. Van Gool et al., Affine/photometric invariants for planar intensity patterns, Proc. European Conference on Computer Vision, pp 642-651, 1996

59 C. Harris et al., A combined corner and edge detector, Alvey Vision Conference, pp147-151, 1988

60 E. Rosten et al., Machine learning for high-speed corner detection, Proc European Conference on Computer Vision, Graz, Austria, 430-443, 2006

61 J. Matas et al., Robust wide baseline stereo from maximally stable external regions, Proc. British Machine Vision Conference, pp 384-393, 2002

Global methods such as correlation are simple techniques that make no assumptions about the object to be detected. Therefore, they can be used in many applications. On the downside, their simplicity means that there are times many computations will be needed making it a slow algorithm. Also, these techniques are unable to account for intra-class variations and occlusion/clutter can easily degrade their performance. Nevertheless, this is not a problem in the case of SAFEMETAL since the image of the coin to be analysed will be obtained from within a controlled environment. Problems such as occlusion, clutter, viewpoint or illumination will be eliminated almost to their entirety. Hence, the simplicity of the correlation algorithm can be taken advantage of to lead to a straightforward hardware implementation.

Transformation-based techniques can also account for many different objects. In addition, they can counteract unwanted occlusion and clutter effects (up to a certain degree). The problem lies with the complexity of those algorithms that many times pose significant memory demands and generate long execution times.

To detect objects in 3D space using a single 2D image as input for recognition, the use of invariants or feature configurations has been proposed. This technique has been successful in many cases using just a single 2D image to localise objects in 3D space. The problem is, however, that this method can only be applied to objects that have prominent features by which they can be easily characterised. Nevertheless, feature detection is not a straightforward process especially since multiple features may be involved. The remaining three techniques are mentioned for completeness rather than for consideration. They are characterised by high computational complexity demands and are therefore unsuitable for the task in hand.