Irish Road sign detection and recognition based on CIE-Lab colour segmentation and HOG feature extraction

Etienne Rochette, DCU Master Student

***Abstract*—This paper presents a road signs detection and classification system for a database including still images of Irish road signs. The solution combines colour segmentation and shape analysis for detection of candidates. The detection divides the candidates into three channels: red, blue and yellow signs treated independently. The features are extracted using Histogram of Gaussians (HOG) and Linear Discriminant Analysis (LDA) performs dimensionality reduction. Several SVMs and Nearest Neighbours classifiers are used to evaluate the performance of the algorithm. Although computationally efficiency has been an objective, the solution developed with Matlab is not suitable for real time application in its current form.**

***Keywords:* Computer Vision, road sign recognition, Colour segmentation, shape analysis, classification, Matlab**

I. INTRODUCTION

IMPROVING the road safety has never been a more serious matter and is made possible mostly by Advanced Driver Assistance Systems (ADAS). This technology is becoming more and more dependable over the past years and is leading to a driving revolution through Self-driving cars. This study is focused on traffic sign recognition as it is a key feature of ADAS.

There is a wide variety of prior works concerning this specific field and the nature of every study is unique. Some works [1, 2] are based on motion as cameras mounted on the car capture the road and its signs. But this project uses still images as a base for the detection [3, 4, 5].
Comparison of the performances is made complex as the database used in each study is different and the number of classes for signs is also unequal.

The detection and recognition of specific objects in natural scenes by using still images is one of the most challenging tasks in computer vision. This project belongs to this last category and the reason why the task is challenging for road signs is explained in section II.

The detection solution retained combines colour segmentation and shape analysis methods described in section III. Once candidates have been extracted from the image, HOG features are extracted and reduced to a lower dimension via LDA as explained in section IV.

Database and sign classes

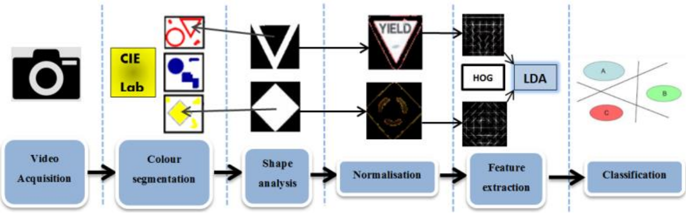

Figure 1, Algorithm flowchart from acquisition to classification

The database includes both pictures taken with a standard camera in the streets of Dublin and screenshots of Google Map Street View in Ireland. The pictures of Dublin's road sign constitute a set with various meteorological conditions and also signs taken from different angles with a wide variety of background objects. Screenshots of Street View were taken to obtain a more complete database with more classes of signs. The yellow colour class also contains orange signs that have to be detected by the colour segmentation.

Existing techniques for road sign recognition

The objective for all ADAS and self-driving cars using road sign segmentation is to surpass human performances. This objective is ruled by the need for an improvement of the road safety. The self-driving cars form a competing innovation only if it is considered safer than a human driver. A study has been carried out to compare the state-of-the-art computer vision solutions to human performances for detecting and classifying road sings [7]. The best solutions have the shared advantage to benefit from a very large database allowing a powerful training for the classifiers. The dataset is made of more than 50,000 images of German road signs divided into 43 classes.

A committee of Convolutional neural networks (CNNs) appear to come with the best solution with correct classification rates of 99.46% that outperforms human performances on the same dataset.

A more extensive survey of research on the existing solutions for road sign recognition may be found in [11] showing that many techniques can produce robust solutions to the problem.

The objective for this project is to design a solution with results comparable to the state-of-the-art but for the Irish

E. Rochette is with the Vision Systems Group, Dublin City University, Dublin, Ireland, in a double degree with association of INSA de Lyon, France (e-mail: etienne.rochette2@mail.dcu.ie).

Road Signs. The solution is also designed to be robust to scale, various physical conditions of the sign (tilted, slightly occulted, twisted), misalignment with the camera and meteorological conditions.

II. SIGN DETECTION

Colour Segmentation

Colour represents a precious source of information for detection of road signs as there are only few natural objects with colour as characteristic as the signs. Although some solutions [8, 14] retain the RGB, YUV or HSI colour space to perform the segmentation, alternatives were also considered.

The CIE-LAB colour space was chosen because it was intended to produce a more perceptually linear space than other colour spaces.

The colour-opponent dimensions ***a*** and ***b*** are calculated using the CIELAB model. The resulting "a,b" space is not perfectly invariant to illumination changes. Thus a set of conditions allow initialising starting centre points (markers) based on extreme values in the "a,b" space. The initialisation concept is introduced on figure 2. A flowchart for the colour segmentation is available in Appendix C of this paper [12].

Orange and red signs are not from the same colour class but are located in the same region of the "a,b" space. Another extreme point analysis focused in this region makes possible the distinction between these two colours.

A first segmentation is performed computing the Euclidian distance (in a,b space) between the centres and each point in the image.
New markers are chosen to be the mean value of the resulting points of the first segmentation.

A second segmentation similar to the first using Euclidian distance allows a better merging of perceptually similar colours.

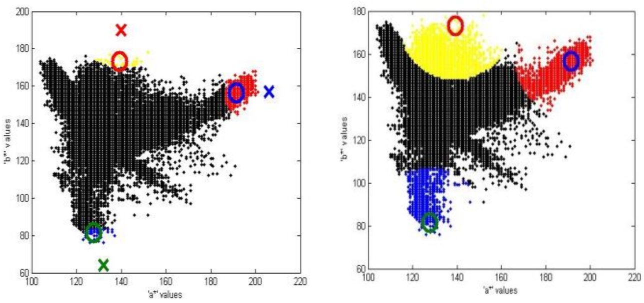

Figure 3, two staged segmentation. The crosses are the first markers; the circles are the mean values used for the second markers

In order to avoid a too wide segmentation for one colour channel, markers are shifted away on the ab space when too many pixels have been assigned to a colour. This operation is performed before the second segmentation.
Unlike the example of figure 2, some pictures don't include all three colours of interest. In this case the markers are pushed away so that the segmentation ignores the colour.

The three channels generated by the segmentation are then used for the shape analysis. The channels corresponding to the segmentation of figure 2 are those of figure 3.

Figure 2, the three channels generated by the colour segmentation displayed on figure 2.

Shape Analysis

The process of selection of candidates for classification is based on a blob by blob analysis and its flowchart available in Appendix C of the paper [12].

Blobs are processed if their surface is included between two extremums, percentages of the total size of the input image. Additionally, blobs with one of their dimension (height or width) superior than 2.5 times the other dimension are ignored. If such a blob is a sign, it is considered to be too angled to be detected thus it is ignored.

Once a candidate is retained, it is resized to 50x50 pixels and is compared to the set of templates. The decision to keep a candidate is ruled by similarity percentage threshold. The easiest and fastest way to compute the similarity percentage is to perform a pixel wise comparison. Templates and candidates are binary images and every identical pixel in the two images is counted. If the ratio of identical pixels is higher than the threshold the blob is treated as a road sign. The candidate is saved both in its original 50x50 colour format and in another form where the background has been removed and only the pictogram is visible.
The operation of pictogram isolation is achieved using reconstruction by dilation on the output blob of the pattern matching.

III. Classification of Candidates

According to the Driver and Vehicle Licensing Agency (DVLA), the signs are designed to be of use for colour blind people as the illustration or text contained in the sign explains the meaning rather than colour itself [9].

This is why colour information is ignored for classification and the patterns are analysed using a greyscale version of the candidates, which provides a computational advantage to a colour-based analysis.

HOG feature extraction

HOG features are useful to describe rigid objects and are used for digit recognition and also pedestrian detection [6]. As road sign pictograms present limited variation in pose, appearance and can contain digits, using a HOG descriptor provides a good description with meaningful separation.

Tests were designed to determine the combination that offers the best size of the cells and the number of orientations to obtain maximum performance of classification.

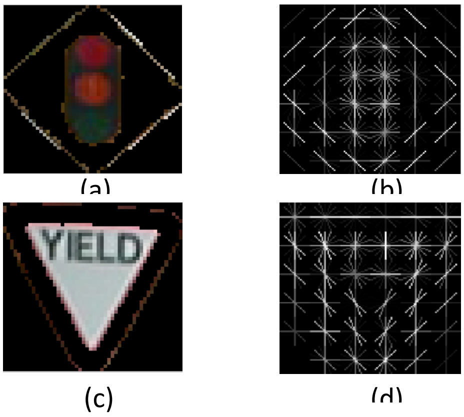

Figure 4, (a) and (c) : input images of the HOG feature extraction (b) and (d) : corresponding HOG images

The results show with the appropriate HOG descriptor 1008 features are used using cells of 8 pixels and 8 orientations. This number is too big to be used directly by a simple classifier such as Nearest Neighbours or Linear SVM.

Dimensionality reduction

Reducing computational complexity is an essential issue to efficiently handle a large number of features for classification. The dimensionality reduction algorithm chosen is Linear Discriminant Analysis (LDA).
The LDA algorithm chosen performs multiclass analysis. It is applied three times for each colour channel as the detection previously discriminated the data into three colour classes. The output dimension for LDA depends on the colour class as each of them has a different number of sub-classes.
The choice has been made to keep as many features as possible because the number of classes of the channels is between 9 and 14 and removing any feature result in a decrease of the classification rates.

Classification

Both the training and test are done according to the colour of each candidate region so that every candidate blob is only compared to those signs that have the same colour as the blob to reduce the complexity of the problem.

Both original and background removed versions of the candidates are tested sequentially for the classification.

For the experiments, only the following classification methods are considered:

- Nearest Neighbors Classifier (NN): Assign points to the class of the closest point based on the minimum least squares error.
The code used for this classifier is the one present in the Statistics Toolbox of Matlab.
- Support Vector Machines Classifiers (SVM):
  - o With linear kernel (LINSVM)
  - o With polynomial kernel (POLYSVM)
  - o With radial basis function (RADSVM) of expression $\exp(-|x-y|^2)$.

The SVM classifiers are train using a one-vs-all (or one-vs-rest) strategy. It implies training a single classifier per class, with the points of the class considered as positive and all other points as negatives. This classification method is performed using the LIBSVM library [10].

To evaluate the classification performance, K-Fold Cross Validation is used. The choice of the number of folds depends on the size of the dataset and the script allows a data driven choice for the value of K.
Different one-versus-all SVMs classifiers are used so that the system can recognize every sign.

The solution on Matlab uses external codes for some of the techniques introduced. The VSG IPA Toolbox [17] is used as a base for the implementation of image processing algorithms.
The HOG features are computed using VLFeat, a library of Computer Vision algorithms [15]. LDA is performed as in [16] and a library for Support vector Machines LIBSVM [10] allows training one-vs-all SVMs.

IV. Detection and Classification Experiments

Detection results

On the final set of images made of 435 images, 342 road signs are extracted by the detection algorithm with a distribution of the data shown in the table below.

The presentation of the sign classes can be found in the Tests and Results Appendix [13].

	total of 342 candidates	
	number of classes	number of points
red	14	145
blue	9	92
yellow	13	105

Table, Properties of the detected candidates

The detection is able distinguish between a sign and another object with a similar colour to the sign and it adapts successfully to illumination changes in the scenes. It is possible to detect several signs on one image but the detection fails when the colour segmentation was not able to extract the sign as one single blob. In this scenario an additional merging step as in [14] would be required to restore the colour blob extracted as an entire sign. Splitting is also required when signs are overlapping. The latest version of the detection algorithm doesn't include the splitting as a feature. One solution to perform the splitting can be to analyse blobs with a height close to twice its width and detect when two signs are touching and then separate them.

Apart from the two previous scenarios, there is a last case where the detection fails even if the road sign is captured with good conditions. The colour segmentation is unsuccessful when a stronger source of the same colour of the sign is present in the image. Based on the dataset, this situation occurs only on the red channel. When the illumination conditions for the sign are bad and another red object benefits from good lightning or is glowing, the sign is either ignored by the segmentation or is partially segmented.
No solution has been found to solve this problem as the colour segmentation principle is based on capturing the strongest red blue and yellow objects of the scene. However this scenario occurs rarely, only when the brightness of the red sign is significantly far from the brightness of the glowing object.

When the colour of the sign is dark due to poor illumination conditions, the colour segmentation is harder to achieve and tends to segment partially the border of the signs. Taking a similar picture using the flash incorporated to the camera generally solve this problem. This hardware precaution of acquisition allows avoiding a more complex software solution to this problem.

Classification results

Several tests based on the candidates presented on the previous table were designed in an attempt to assess the performance of the solution.

Computing HOG features over the 50x50 pixels candidates allows extracting useful information for the classification but the size of the cells and the number of orientations used for each cell affects the performance of the classification. A test was designed to decide what parameters to use for the HOG.
The classification is performed using LDA and a SVM with radial basis kernel and the results are displayed on figure 5. In term of computational efficiency, it is unpractical to use too small cell sizes. The optimal parameters for the HOG extraction are a cell size of 8 pixels and 8 orientations, for a total of 1008 features extracted.

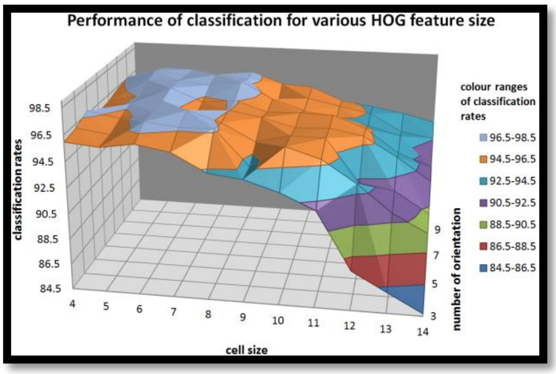

Figure 5, Performance of classification for various HOG feature sizes, classification rates in %

Results of other tests are gathered on figure 6. The combination of techniques for each test is indicated by the following abbreviations:

- **HOG**: Histograms of Gradients feature extraction
- **Raw pixels**: raw pixel values used as features
- **LDA**: Linear Discriminant Analysis for dimensionality reduction
- **NN**: Nearest Neighbors classifier
- **LSVM**: Linear Support Vector Machine Classifier
- **POLYSVM**: Polynomial Support Vector Machine Classifier

RADSVM: Radial Basis Function Support Vector Machine Classifier

	raw pixels + LDA + SVM	HOG + RADSVM	HOG + NN	HOG + LDA + NN	HOG + LDA + LSVM	HOG + LDA + POLYSVM	HOG + LDA + RADSVM
red pictogram	42,5	89,5	85,5	94,5	97,2	94,3	97,2
red signs	66,8	89,5	90,1	95,9	96,5	93,8	**98**
blue pictogram	18,9	94,8	92,2	94,6	95,3	91,3	96,3
blue sign	73	100	97,8	100	100	100	**100**
yellow pictogram	19,3	92,4	91,2	94,3	95,3	92,4	**97**
yellow sign	65,4	93,4	90,4	94,3	94,3	94,2	95,3

Figure 6, Results of all tests for both pictogram and original versions of the candidates, classification rates in %

V. Analysis

Evaluating the performance of the detection is harder than the classification because of the wide variety of conditions of the pictures. Some pictures contain signs that are not supposed to be detected for reasons elaborated in [13]. But the solution also fails to detect signs for reasons presented in the previous section of this paper. Considering all this and the fact that 342 signs are extracted from the 435 images, the detection can be considered to be successful. The detection furnishes enough individuals of each ach sign class to enable the classification with good conditions.

The success of the template matching step is highly linked to the success of the colour segmentation. Each sign is not segmented with the same accuracy and the similarity percentage can be low if only a part of the sign has been segmented. Experiments showed that the YIELD sign is harder to detect by the similarity percentage calculation because of misalignments with the camera. This is why there are three templates for detecting the triangle shape. There is not one template by class of sign but some templates are designed to detect several classes of signs. The circles in the red channel can detect all signs except STOP, NO ENTRY and YIELD signs. The disc for the blue channel is also a generic template for all kinds of blue signs.

It is clear from the graph of figure 5 that the bigger the cells are the harder it is to extract useful information for the classification. The classification rates drop when increasing the cell size starting from a cell size of 9 pixels. Too big cell sizes make the extraction of details in the pictograms impossible. If one cell englobes a large part of the pictogram, some of the gradient information will not be captured in the histogram. For smaller cell sizes, the classification rates are higher but the rates decrease for the smallest cell size. In term of computational efficiency, it is unpractical to use too small cell sizes because the number of orientations to computes becomes too important and this process is slow.
Thus a compromise between a large cell size and a good classification rate represent the best combination for the solution.

For the number of orientation, the test shows that using too many orientations does not improve the classification rates and can even negatively affect the rates. Logically, using too few orientations is not a good idea as the information is lost and the rates drop.
This is with these parameters that the highest classification rate has been found and the computational cost of the extraction is acceptable as 1008 features are extracted in total.

Using raw pixel values with LDA and SVM is the less appropriate technique as the classification rates are very low compared to the rest of the experiments. The rates for the classification of pictograms are significantly lower than for the original signs. This is due to the fact that most pixel values are 0 for the pictograms and the exact position of the contour and the pictogram itself is very unlikely to be exactly at the same pixel location for every candidate of a same class. This test shows that extracting HOG features is essential to capture meaningful information about the pictograms.

HOG +RADSVM and HOG + NN tests were designed to discover if LDA is essential to the solution. Removing the dimensionality reduction to the process affects negatively the classification because too many features are used. But even if the rates are lower than for the HOG and LDA combined, the results are surprisingly high. For example every blue sign is correctly classified for HOG+RADSVM and the smallest rate is obtained for the red pictograms with 89.5% correct classification.

Thus both HOG extraction and LDA are useful and contribute to improve the classification success. Concerning the classifiers, the worst rates are obtained for the POLYSVM classifier although the expectations were that the NN classifier would be less efficient because it is sometimes referred to as a naive approach to classification. The Nearest Neighbors shows results close to the other classifier.

In all experiments HOG+LDA+RADSVM produces better results than the other techniques. But the pictogram isolation improves the results only for the yellow colour class. In this case the classification rate is 2% higher than with the original version of the signs. But for the two other colour classes, using only the pictogram doesn't improve the classification. Therefore the red and blue signs have to be classified in their original format with the background but the pictogram isolation has to be performed on yellow candidates.
Knowing this, the best correct classification rates are 98.1% for red signs, 100% for blue signs and 97.1% for yellow signs. This makes an overall correct classification of **98.4%.**

The results are perfect for the blue colour class as the 92 candidates are correctly classified into the 9 sign classes. The result is higher than for the other colours because there are less classes and also they are visually easier to distinguish as the pictograms are not alike.
For the red and yellow colour class, some signs have similar pictograms and it makes the classification harder. Some speed limit signs, forbidden red signs, or merging traffic yellow signs can be very alike. The classification can fail if the candidate was extracted with a very low resolution or with another distortion presented in [13].

Background removal is not perfect and tends to fail mostly for blue signs. This can be explained by the fact that the colour of the illustration inside the sign is corrupted by the blue of the sign itself. This causes the detection to include the pictogram with the colour blue and then the conditional dilation removes parts of the illustration. In a potential latter release of an improved solution, the algorithm used for the background removal has to be modified. Another colour segmentation focused only on the candidates seems to be easy to implement and should provide a good separation for the sign. Alternatively, Hough transforms could be useful tools to locate the boundaries of the various shapes of signs.

VI. Conclusion

The detection algorithm allows finding signs under various conditions with a total of 342 road signs and only a few false candidates. The solution is not rotation invariant because of symmetry properties of different signs but it was designed to be invariant to scale, lighting conditions, small occlusions, rain and distortions of the signs. These conditions for the signs were captured in the dataset and the solution has proven to successfully overstep these complications. But the solution's robustness was tested and detection failures were identified and explained.

The detection could have been improved by integrating additional features such as merging and splitting of the blobs generated by the colour segmentation.

This project and the state of the art publication presented in a more extensive survey of research on the road sign recognition and segmentation found in [11] don't rely on the same aspects and resources. The database in the best studies of the competition [7] regroups more than 50,000 images of German road signs in 43 classes. Thus the information available for training is highly superior to this project counting 342 candidates for classification. A committee of Convolutional Neural Networks (CNN) showed the highest classification accuracies. The high performance of the CNN is made possible by 37 hours of training using 4 GPUs on dedicated hardware. Despite benefiting from a strong capacity for training, the classification for the German dataset is challenging because of a large number of classes of signs and very poor conditions of the signs.

The corresponding results of classification for the state of the art solution are close to perfect with 99.46% correct classification rate. This not only outperforms the rates of this project with Irish signs of 98.4% but also the best individual in the human performance experiment who got 99.22% correct classification.

The project can nevertheless be considered as a success because of the high classification results on standard and simple classifiers. The solution identifies signs from 36 classes with rates comparable to the latest solutions released.

References

[1] M. Wada, K. S. Yoon, H. Hashimoto, "Development of Advanced Parking Assistance System" IEEE Intelligent Vehicles Symposium, pages 255 – 260, June 2005

[2] R. Marcin, L. Eichner, P. Breckon, "Integrated Speed Limit Detection and Recognition from Real-Time Video", Presented at IEEE Intelligent Vehicles Symposium Conference, The Netherlands, June 2008

[3] S. Maldonado-Bascón, S. Lafuente-Arroyo, P. Gil Jiménez et al. "Road-Sign Detection and Recognition Based on Support Vector Machines" IEEE Transactions on intelligent transportation systems, Vol. 8, No. 2, June 2007

[4] Y. Liu, D. Duh, S. Chen "Scale and Skew-Invariant Road Sign Recognition", International Journal of Imaging Systems and Technology archive, Vol. 17 Issue 1, June 2007, Pages 28 – 39

[5] S. Vitabile, A. Gentile and F. Sorbello "A neural network based automatic road signs recognizer" Neural Networks, 2002. Proceedings of the 2002 International Joint Conference on Neural Network, Vol. 3, 2002, Pages 2315 – 2320

[6] J. Stallkamp, M. Schlipsing, J. Salmen, C. Igel, "Man vs. Computer: Benchmarking Machine Learning Algorithms for Traffic Sign Recognition", Institute für Neuroinformatik, 2010

[7] E. Rochette, A Survey of research on Irish road sign recognition and segmentation, MEng Final Portfolio, Appendix A, September 2014.

[8] H. Fleyeh "Color Detection And Segmentation For Road And Traffic Signs" Proceedings of the 2004 IEEE Conference on Cybernetics and Intelligent Systems, Singapore, 1-3 December, 2004

[9] A. Broggi, P. Cerri, P. Medici "Real Time Road Signs Recognition", Proceedings of the IEEE Intelligent Vehicles Symposium, Istanbul, Turkey, June 2007

[10] E. Rochette, Project Design and Implementation on Irish road sign recognition and segmentation, MEng Final Portfolio, Appendix C, September 2014.

[11] Driver & Vehicle Licensing Agency https://www.gov.uk/government/organisations/driver-and-vehicle-licensing-agency

[12] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection" In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 886–893, 2005

[13] Chih-Chung Chang and Chih-Jen Lin (2011), "LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology", Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm , <Accessed on: 01-09-2014>

[14] Paul F Whelan, "VSG Image Processing and Analysis (VSG IPA) Matlab Toolbox Manual", Technical Report, Centre for Image Processing & Analysis, Dublin City University, 2013

[15] A. Vedaldi and B. Fulkerson "VLFeat: An Open and Portable Library of Computer Vision Algorithms",2008, http://www.vlfeat.org/ , <Accessed on: 01-09-2014>

[16] Deng Cai, Xiaofei He, Jiawei Han, "SRDA: An Efficient Algorithm for Large Scale Discriminant Analysis", IEEE Transactions on Knowledge and Data Engineering, 2007

[17] E. Rochette, Tests and Results on Irish road sign recognition and segmentation, MEng Final Portfolio, Appendix D, September 2014.

[18] H. Bay, T. Tuytelaars, L. Van Gool, "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359, 2008

[19] X. Hu, X. Zhu, D. Li, H. Li "traffic sign recognition using Scale Invariant Feature Transform and SVM"

[20] G. Azzopardi, N. Petkov "Trainable COSFIRE Filters for Keypoint Detection and Pattern Recognition", Actions on Pattern Analysis and Machine Intelligence, Vol 35., No. 2, February 2013

[21] Paul F Whelan (2011) "VSG Image Processing and Analysis (VSG IPA) Toolbox", Technical Report, Centre for Image Processing & Analysis, Dublin City University.

[22] Paul F Whelan (2011) "VSG Image Processing and Analysis (VSG IPA) Matlab Toolbox - Software"", http://www.cipa.dcu.ie/code.html, Centre for Image Processing & Analysis, Dublin City University.

[23] Paul F. Whelan and D. Molloy (2000), "Machine Vision Algorithms in Java: Techniques and Implementation", Springer (London), 298 Pages. ISBN 1-85233-218-2.

[24] Paul F Whelan (2011), "EE544: Computer Vision Course Notes", Centre for Image Processing & Analysis, Dublin City University.

[25] Paul F Whelan (2011), "EE425/EE453: Image Processing and Analysis Course Notes", Centre for Image Processing & Analysis, Dublin City University.