UNSUPERVISED REGION OF INTEREST DETECTION USING FAST AND SURF

Abass A. Olaode[1], Golshah Naghdy[1] and Catherine A. Todd[2]

[1]School of Electrical Computer Telecommunication Engineering,
University of Wollongong, Wollongong, Australia
Abass.Olaode808@uowmail.edu.au
golshah@uow.edu.au
[2]Faculty of Computer Science and Engineering,
University of Wollongong, Dubai, UAE
CatherineTodd@uowdubai.edu.au.

ABSTRACT

The determination of Region-of-Interest has been recognised as an important means by which unimportant image content can be identified and excluded during image compression or image modelling, however existing Region-of-Interest detection methods are computationally expensive thus are mostly unsuitable for managing large number of images and the compression of images especially for real-time video applications. This paper therefore proposes an unsupervised algorithm that takes advantage of the high computation speed being offered by Speeded-Up Robust Features (SURF) and Features from Accelerated Segment Test (FAST) to achieve fast and efficient Region-of-Interest detection.

KEYWORDS

Region of Interest, Image segmentation, SURF, FAST, Texture description, PLSA, BOV, K-means clustering, unsupervised image classification.

1. INTRODUCTION

Image modelling has been recognised as an essential step towards recognition [1], thus an important components in image retrieval. Many image retrieval implementations adopt global features such as colour histograms in describing image contents [2]. Although this approach of covering the entire image space has proven to be successful for images with distinct colours, Tuytelaars and Mikolajczyk [2] noted that such approach cannot distinguish between foreground and background, and it is severely challenged by image clutter and occlusions [2].

Tuytelaars and Mikolajczyk [2] explained that an approach to tackling the challenges of global image features is to segment the image into a limited number of regions or segments, where each region corresponding to a single object or part of an object. Common methods of achieving this involves exhaustively sampling different subparts of the image at each location and scale. Such approach has the tendency to become computational expensive and inefficient [2, 3]. An efficient alternative is to determine the most important region of the image, commonly regarded as Region of Interest (ROI) [4].

David C. Wyld et al. (Eds) : CCSIT, SIPP, AISC, NLP - 2015
pp. 63–72, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.50406

The identification of an image's ROI using supervised learning is often challenged by the need for prior information regarding patterns within the image collection, thus making unsupervised learning a more attractive option [4, 5]. This paper presents a novel ROI detection approach that uses fast feature detection algorithms (Speeded-Up Robust Features (SURF) and Features from Accelerated Segment Test (FAST)) to identify likely regions of interest, and then compares the texture of these regions to complete the ROI detection.

The remainder of this paper is structured as follows: Section 2 provides background information SURF detector and FAST detector. Section 3 gives a brief review of recent works on ROI detection where related approaches have been applied, while Section 4 provides a detailed description of the implementation of the proposed SURF and FAST combination in the detection of ROI. Section 5 discusses the experimentations carried out in the evaluation of the effect of the proposed ROI detection on unsupervised classification using PLSA, and Section 6 highlights the direction of future works aimed at the use of ROI in semantic labelling of images.

2. IMAGE FEATURE DETECTORS

Feature extraction algorithms use digital image processing techniques to extract low level features from the high dimensional matrix representation of images. For reliable image recognition, it is important that the features extracted from images be detectable even under changes in image scale, noise and illumination. To satisfy this need, keypoints corresponding to high-contrast locations such as object edges and corners are often sought [6, 7]. Although traditional image feature extraction algorithms consist of feature detection and feature description components, this section focuses mainly on the detection of image features.

The most popular image feature extraction algorithm is the Shift Invariant Feature Transform (SIFT). SIFT uses the Difference of Gaussian (DoG) algorithm to detect image features, and has proven to be very successful in computer vision applications due to it resistance to common image transformations [6, 7]. However, the computational requirement of SIFT is very high [6, 7], which has made algorithms such as SURF and FAST preferred choice for real-time applications [6, 7].

2.1 SURF

Like SIFT, SURF features extraction algorithms can be regarded as sparse feature extraction algorithms because it only detect and describe features at keypoint locations. Rather than using DoG and image pyramid for the detection of keypoints as in SIFT, SURF uses the Hessian matrix in which the convolution of Gaussian second order partial derivatives with a desired image are replaced with box filters applied over image integrals (sum of grayscale pixel values) [9]. Given a point x = (x, y) in an image I, the Hessian matrix H(x, σ) in x at scale σ is defined as follows (Equation 1):

$$H(x,\sigma) = \begin{pmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{pmatrix} \quad (1)\ [9]$$

Where Lxx(x, σ) is the convolution of the Gaussian second order partial derivative $\frac{\partial^2}{\partial x^2}\, g(\sigma)$ with the image I in point x, and similarly for Lxy(x, σ) and Lyy(x, σ) [9]. The use of integral image representation at the keypoint detection stage of SURF ensures that the computational cost of applying box filter is independent of the size of the filter. This allows SURF to achieve much faster keypoint detection than SIFT by keeping the image size the same while varying only the

filter size [9]. Figure 1A illustrates the SURF keypoints detected on a sample image from Caltech-101 dataset.

Although, SURF's performance is mostly similar to SIFT, it is unstable to rotation and illumination changes [10]. Liu et al. [11] noted that although SURF is capable of representing most image patterns, it is not equipped to handle more complicated ones. However, Khan et al [9], implemented classification experiments on images from David Nister, Indoor, Hogween and Caltech datasets to yield results that confirms that SURF's performance is as good as that of SIFT, with both recording 97% accuracies on Caltech dataset. Therefore, this study considers SURF adequate enough to be considered for the purpose of detecting image features in the determination of ROI.

2.2 FAST

Corners are found at various types of junctions, on highly textured surfaces, and occlusion boundaries. With the aim of identifying a set of stable and repeatable features, they are typically detected using corner detectors such as the Harris corner detector, Smallest Univalue Segment Assimilating Nucleus (SUSAN) detector, and FAST detector.

The Harris detector was identified as the most stable one in many independent evaluations. Although SUSAN is more efficient than Harris detector, it is also more sensitive to noise. FAST is an improvement over the SUSAN detector with higher accuracy [2]. It fell just behind the Harris detector, but significantly faster than any other algorithm [13] making it the most appropriate for real time machine vision applications and for image retrieval purposes.

The FAST considers corners more intuitive than edges because they show a stronger two dimensional intensity change, and are therefore well distinguished from the neighbouring points [13]. FAST uses a corner response function (CRF) that gives a numerical value for the corner strength based on the image intensity in the local neighbourhood. This CRF was computed over the image and corners which were treated as local maxima of the CRF. Along with this, FAST also employs a multi-grid technique to improve the computational speed and supress detected false corners being detected. Figure 1 demonstrates the corners points detected using the FAST algorithm.

Undoubtedly, the main contribution of FAST was the increment of the computational speed required in the detection of corners [2]. SURF and FAST have been considered two of feature detection algorithms most suitable for real-time applications due to their speeds.

a.SURF Keypoints b. Fast corner points

Figure 1. Feature detection on a sample image

3. RELATED WORKS

Existing unsupervised ROI determination algorithms often require extensive and computational search within an image space before the identification of the desired region [4]. In the unsupervised image categorisation framework proposed by Huang et al. [4], the authors presented an unsupervised ROI detection approach that computes dense-SURF over the unlabelled image [4]. The high computation requirement of dense-features is further increased by the comparison of the images within the image set so as to identify common feature when the process is built on unsupervised learning [4].

The unsupervised ROI determination proposed by Huang et al. [4] incurs heavy computation, and is not suitable when managing large number of images. This study considers computational speed an important requirement in image retrieval especially when handling large number of images (1000 images and above), and proposes a ROI determination algorithm that takes advantage of the high computational speed offered by SURF and FAST keypoint detection algorithms.

Although the use of keypoints in the detection of ROI has been investigated by Kapsalas et al [3], who used Harris corner detector in identifying objects present in an image, the approach proposed in this paper differs, in that it attempts to achieve effective labelling through the identification of the important object within the image. It also differs from the ROI detection approach proposed by Huang et al [4] since it does not compared the image being processed with other images in the set during the ROI detection, thus reducing the computational requirement.

4. THE PROPOSED UNSUPERVISED FRAMEWORK

The identification of interest points present within the space of an image is important in the determination of the image's ROI [3], therefore the method being proposed in this paper maximises the number of the number of interest points detected within a sample image through the use of the combination of FAST corner detector and SURF detector as shown in Figure 2A. The use of several complementary feature detectors in such manner ensures good coverage of the image surface [2].

If FAST corner points and SURF keypoints are respectively represented by the sets $F = \{f_1, f_2, f_3 \dots \dots \dots f_L\}$ $and S = \{s_1, s_2, s_3 \dots \dots \dots s_L\}$, then the combined FAST and SURF feature points can be represented by the set P; where $P = F \cup S = P = \{p_1, p_2, p_3 \dots \dots \dots p_L\}$. The two key criteria which distinguish keypoints belonging to an ROI from those that do not belong to the desired region are location and description.

The algorithm proposed uses the coordinates provided by the SURF and FAST algorithm to satisfy its need for location information. It recognises texture as the most appropriate image characteristics by which regions within an image can be distinguished from one another, and employs Law's filter [13, 14] in developing a 9 dimensional descriptor for a rectangular mask centred at the coordinates of the location of the point.

The dimension of the mask use is made to be 0.33*(height * Width), thus responding to image size while capturing similarities between neighbouring keypoints. The choice of 9 dimensional texture descriptor ensures the avoidance of the heavy computations associated with the popular descriptors, thereby reducing the computation overhead. The proposed algorithm relies on K-Nearest Neighbour categorisation (KNN) for the categorisation of the texture descriptions of each keypoint into either foreground or background, therefore it requires training samples.

From a reference point $(\overline{x}, \overline{y})$ established to be the mean of all the x and y coordinates, the horizontal and vertical distances of each point are calculated. The pair of distances for all the keypoints are placed in the sets $X = \{x_1, x_2, x_3 \dots \dots x_l\}$ which has a mean of $\overline{X}$, and $Y = \{y_1, y_2, y_3 \dots \dots y_l\}$ with a mean of $\overline{Y}$represents the means of the sets. A keypoint (x_i, y_i) is chosen to be a likely foreground training sample candidate if it satisfies the conditions of Equation 2.

$$|x_i - \overline{x}| < \overline{X} \tag{2a}$$

$$|y_i - \overline{y}| < \overline{Y} \tag{2b}$$

Where $I_x$and $I_y$represents the image dimensions, any keypoints that do not satisfy the above criteria is considered to be a background training sample if does not satisfy at any of Equation 3a or Equation 3b.

$$|x_i - \overline{x}| < \frac{2}{5} * (I_x) \tag{3a}$$

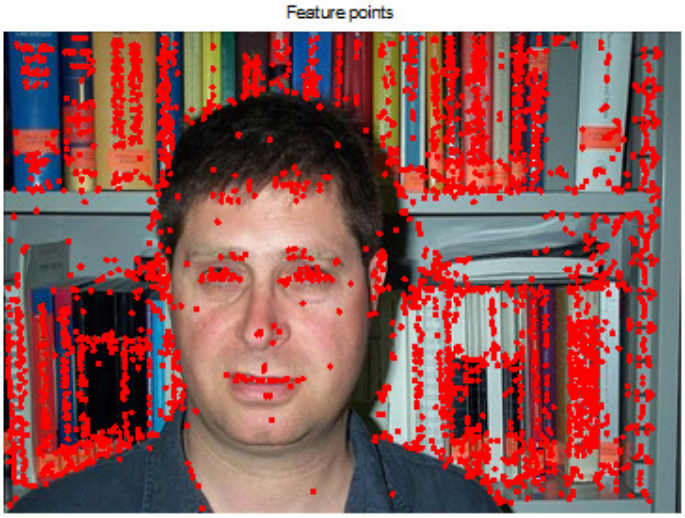
$$|y_i - \overline{y}| < \frac{2}{5} * (I_x) \tag{3b}$$

Keypoints that does not satisfy both of Equation 2, but satisfies one of Equation 3 are categorised based on their texture descriptors using KNN. Furthermore, the texture description of the "likely" foreground training samples are compared with those of the background training samples so as to achieve a set of training samples that is exclusive to the foreground. Assuming $R = \{r_1, r_2, r \dots \dots \dots r_L\}$is a set of texture descriptors for points that satisfied Equation 2, and $B = \{b_1, b_2, b_3 \dots \dots \dots b_L\}$ is the set of texture descriptors for points that satisfied at least one of Equation 3, then a sample is confirmed to belong to the foreground training set if it satisfies the Equation 4.

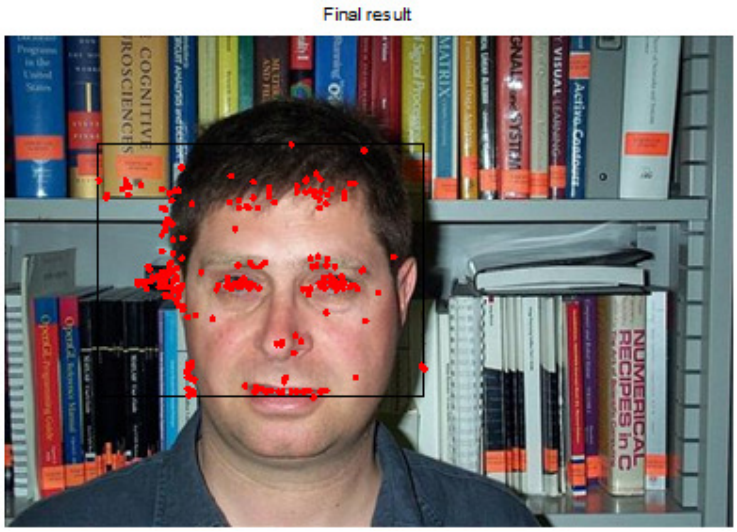
$$\frac{\sum_{j=1}^{L} |r_i - r_j|}{n(R)} < \frac{\sum_{j=1}^{L} |r_i - b_j|}{n(B)} \tag{4}$$

Equation 4 indicates that a legitimate foreground training candidate sample should record an average Euclidean distance from every other point within the "likely" foreground training sample group which is less that the average Euclidean distance recorded from the point to every point in the background training samples.

Preliminary experiments conducted as part of this study shows that although Equation 4 holds in most case, the reverse can also be the case, thus making the Equation a means of separating the "likely" foreground samples into two groups. In such case, the group which is least similar to the background training samples in term of texture description is then chosen to be the foreground training samples. In the implementation of the KNN, each feature point to be categorised is allocated the highest occurring label from the closest 5 neighbours, thus the points labelled as the foreground are grouped together to form the desired Region. In this way, the points located within the region of interest are effectively identified as shown in Figure 2B. The ROIs detected on more sample images from Caltech-101 are displayed in Figure 3.



a. The combined FAST and SURF feature points on a sample image



b. An illustration of the result of using the proposed algorithm on a sample image

Figure 2. Illustration of a.) The keypoints identified using SURF and FAST, and b.) The Region of Interest points identified using the proposed ROI detection algorithm

Figure 3. Results of application of the proposed ROI detection framework on sample images

5. EXPERIMENTS AND RESULTS

As discussed in Section I, the determination of ROI during unsupervised image categorisation is important because it ensures that most of the attention is paid to the images' foreground, thereby limiting the effect of images' background on the classification accuracy. This section examines the effect of this algorithm on the completely unsupervised combination of PLSA and K-means.

The experiments in this study used the 3 image datasets constituted from 12 Caltech-101 in [15]. While the number of images is fixed at 500, the categories are varied (4, 8 and 12categories). These classes are: Airplanes, Motorbikes, Face, Watch, Car, Backpack (Caltech-256), Ketch, Bonsai, Butterfly, Crab, Revolver, and Sunflower. In all the experiments in this section, the Histogram of Oriented Gradients (HOG) [16] feature extraction is chosen as the image feature extraction algorithm [11].

The proposed ROI detection algorithm is implemented on the image collections and the detected ROI images are converted to the various forms PLSA models, and then quantised into semantic groups using the k-means algorithm. The PLSA/K-means classification is implemented with 25 latent topics and 25 clusters, thus allowing a one to one mapping between each of latent topics and each of the semantic groups identified during K-means clustering.In evaluating the quality of the centroids presented at the completion of each categorisation process, each of the centroids is visualised and labelled, and the label appointed to the centroid is automatically applied to all the images in the centroid's cluster. With all the available images labelled, accuracy of the unsupervised categorisation is determined through a comparison between the new labelling and the ground truth. By varying the visual codebook size under the 3 chosen image collections, this section determines the appropriate visual codebook sizes for the Bag-of-visual word modelling which precedes the PLSA classification. The graphical demonstration of their average performances over 5 runs is shown in Figure 4.

Using the data presented graphically in Figure 4, this paper identifies that the most appropriate visual codebook sizes for the implementation in the categorisation of 4, 8 and 12 categories image collections are 200, 500, and 900. Table 1 provides a comparison of the performance of PLSA classification under two scenario; 1) without using ROI, and 2) using ROI.

An overall assessment of Table 1 reveals a general increment in categorisation accuracy when ROI is included during PLSA/K-means categorisation. This increment can be attributed to the ability of the proposed ROI algorithm to minimise the amount of image background included during image modelling.

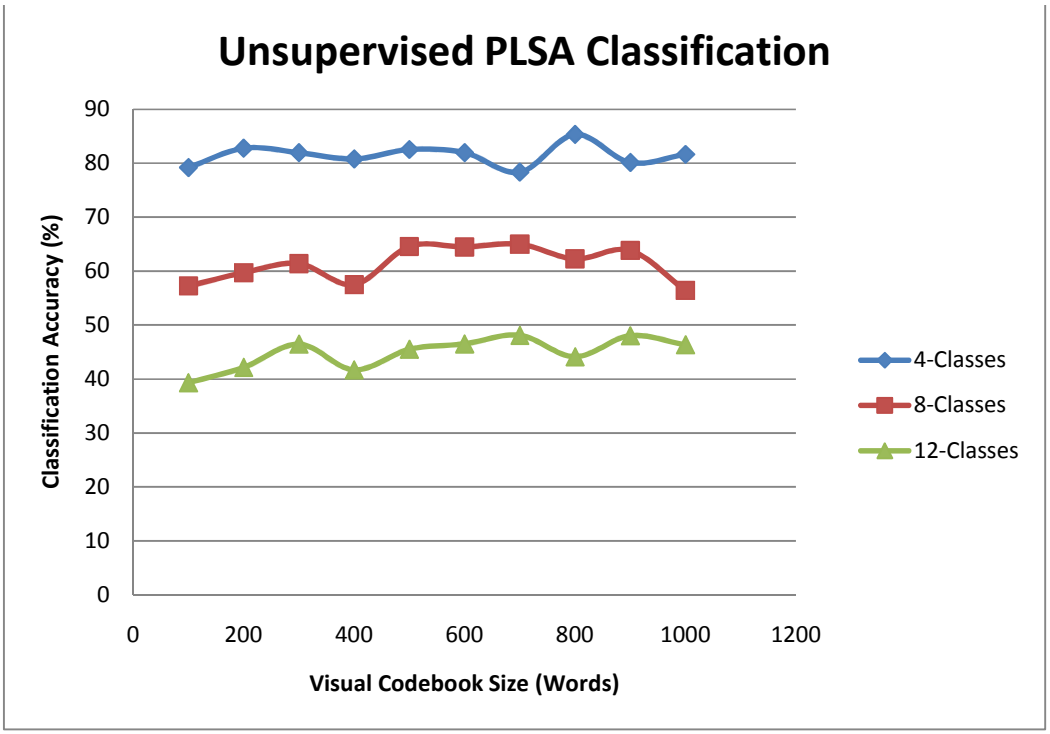


Figure 4.The unsupervised PLSA categorisations implemented with the inclusion of Region of Interest determination

Table 1. A comparison of the effects of ROI during Unsupervised image categorisation via PLSA

Number of categories	Visual Codebook sizes	Accuracies	
		Without using ROI detection	Using ROI detection
4	200	80.80%	82.77%
8	500	58.07%	64.56%
12	900	45.55%	48.07%

6. FUTURE WORKS

This paper has successfully demonstrated the use of FAST and SURF to be useful tools in the determination of an image's ROI, and the possibility of improving the accuracy of an image classification by limiting an image's modelling to the ROI of the image. However, it is important to note that the categorisation accuracies recorded with the use of ROI is lower than the accuracies obtainable under a completely supervised implementation of PLSA categorisation. The recorded performance can be further improved through the combination of the proposed ROI determination algorithm with the spatial pyramid algorithm proposed by Lazebnik et al. [17] or with semantic segmentation. These combinations will be investigated in future works.

To further improve the categorisation accuracy, there is the need to identify a more effective feature extraction algorithm especially that is able to accommodate the diverse nature of the nature dataset. A possible solution to this challenge is the combination of the HOG descriptor with another image feature extraction algorithm (such as shape or corner description algorithms). This will also be examined in future works.

7. CONCLUSION

For minimising the effect of image backgrounds on classification accuracies, this paper proposes the use determination of the ROI of each using SURF and FAST, and demonstrated the ability of the proposed algorithm to limit image modelling to relevant region within the image.

Using 3 image collections constituted from Caltech-101, this paper successfully demonstrates the effectiveness of the categorisation model in improving the unsupervised PLSA categorisation, and identified the inclusion of spatial pyramid and semantic segmentation alongside ROI determination as two approaches that may be employed in the search for higher accuracy during unsupervised PLSA categorisation.

ACKNOWLEDGEMENT

The authors of this work wish to thank the research students of their University for their support.

REFERENCES

[1] C.M.Bishop and J.M.Winn, "Non-linear Bayesian Image Modelling," in 6th European Conference on Computer Vision, ECCV 2000, Antibes, 2000.

[2] T.Tuytelaars and K.Mikolajczyk, "Local Invariant Feature Detectors: A Survey," Foundations and Trends in Computer Graphics and Vision, vol. 3, no. 3, p. 177–280, 2008.

[3] P.apsalas, K.Rapantzikos, A.Sofou and Y.Avrithis, "Regions of interest for accurate object detection," in International Workshop onContent-Based Multimedia Indexing, CBMI., London, 2008.

[4] Y. Huang, Q.Liu, F.Lv, Y.Gong and D.Metaxas, "Unsupervised Image Categorization by Hypergraph Partition," IEEE Transactions On Pattern Analysis And Machine Intelligence, vol. 33, no. 6, June 2011.

[5] D.-C.Lee and T.Schenk, "Image Segmentation From Texture Measurement," ISPRS Journal of Photogrammetry and Remote Sensing, vol. XXIX, no. 3, pp. 195-199, 1992.

[6] M.Guerrero, "A Comparative Study of Three Image Matcing Algorithms: SIFT, SURF, and FAST," Utah state University, Utah, 2011.

[7] R.Sukthankar and Y. Ke, "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors," in Computer Vision and Pattern Recognition, 2004. CVPR 2004, Pittsburg, 2004.

[8] E.Rublee, V.Rabaud, K. Konolige and G.Bradski, "ORB: an efficient alternative to SIFT or SURF," IEEE International Conference on Computer Vision (ICCV),, Barcelona, 2011.

[9] H.Bay, T.Tuytelaars and L. V. Gool, "SURF: Speeded Up Robust Features," in Computer Vision-ECCV , Zurich, Springer Berlin Heidelberg, 2006, pp. 404-417.
[10] N.Khan, B.McCane and G. Wyvill, "SIFT and SURF Performance Evaluation against Various Image Deformations on Benchmark Dataset," in International Conference on Digital Image Computing: Techniques and Applications, Noosa, 2011.
[11] L.J. &. O.Gwun, "A Comparison of SIFT, PCA-SIFT and SURF," International Journal of Image Processing, vol. 3, no. 4, pp. 143-152, 2008.
[12] C.Liu, J.Yang and H. Huang, "P-SURF: A Robust Local Image Descriptor," Journal Of Information Science And Engineering, vol. 27, pp. 2001-2015, January 2011.
[13] M.Trajkovic and M. Hedley, "Fast corner detection," Image and Vision Computing, vol. 6, no. 2, pp. 75-87, 1998.
[14] H.A. Elnemr, "Statistical Analysis of Law's Mask Texture Features for Cancer and Water Lung Detection," International Journal of Computer Science, vol. X, no. 6, pp. 196-202, 2013.
[15] A.Olaode, N.Golshah and C.Todd, "Unsupervised Image Classification by Probabilistic Latent Semantic Analysis for the Annotation of Images," in 2014 lnternational Conference on Digital lmage Computing: Techniques & Applications (DICTA), Wollongong, 2014.
[16] N.Dalal and B.Triggs, "Histograms of Oriented Gradients for Human Detection," INRIA, Montbonnot, 2004.
[17] S.Lazebnik, C.Schmid and J.Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories," in Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference, Illinois, 2006.

AUTHORS

Abass A. Olaode obtained a Master of Engineering in Telecommunication from the University of Wollongong, Australia in 2012, and is currently a research student at the same institution. He is currently conducting a research into the application of unsupervised machine learning in the elimination of semantic gap from image retrieval.

Golshah Naghdy is an Associate Professor at the School of Electrical, Computer and Telecommunication Engineering, University of Wollongong. She was a Senior Lecturer at Portsmouth University before joining Wollongong University in 1989. Her research interests include biological and machine vision in particular a generic vision system based on "wavelet neurons" and its application in the development of artificial retina implants, medical image processing, content based image retrieval, and robotics.