# Image Representation for Image Mining: A Study Focusing on Mining Satellite Images for Census Data Collection

Frans Coenen and Kwankamon Dittakan

**Abstract** This paper firstly presents a taxonomy for mage representation in the context of image mining. The main premise being that the actual mining algorithms that may be used are well understood, it is the preprocessing of the image data that remains a challenge. The requirement for the output from this preprocessing is some image representation that us both sufficiently expressive while at the same time being compatible with the mining process to be applied. Three categories of representation are considered: (i) statistics-based, (ii) tree-based and (iii) point series based. The second contribution of this paper is an analysis of the proposed representations categories with respect to a novel image mining application, the collection of individual household census data from satellite imagery, more specifically Google earth satellite imagery. The representations are considered both in terms of generating census prediction models and in terms of applying such models for larger scale census prediction.

## 1 Introduction

Image mining is an important element of the canon of data mining. Decision making is routinely supported by visual information and visualisations of data. At the same time our ability to collect visual information (image data) is increasing rapidly, partly because of technological advancements and partly (and associated with the first) the increasingly reduced cost of collecting such data. For example the collection of retina images are now routine for anyone visiting an optician, whilst the cost

Frans Coenen
Dept. Computer Science, The University of Liverpool, Liverpool, L69 3BX, UK
e-mail: coenen@liverpool.ac.uk

Kwankamon Dittakan
Faculty of Technology and Environment, Prince of Songkla University (PSU), Thailand
e-mail: kwankamon.d@phuket.psu.ac.th

of Magnetic Resonance Imaging (MRI) scans has reduced considerably. The computing power available to process images is also rapidly increasing. Consequently the demand for utilising image data for the purpose of extracting information (image mining) is increasing. It should also be noted here that the images we wish to process, although typically 2D in nature, can also be in a 3D format; our ability to collect 3D (volumetric) data has also been advancing such that 3D data is now also readily available.

The challenge of image mining is not so much the algorithms used to extract knowledge from image data, these tend to be well understood, but the end to end process from the initial image representation to the final knowledge interpretation. Although our ability to process large quantities of data is increasing, typically we are still not able to represent image data in its entirety (pixel by pixel), nor in most cases would this be useful; although the use of techniques such as the Convolutional Neural Networks (CNNs) [23] is a significant step in this direction. The manner in which we represent the image data we wish to mine is of great significance (the "rubbish in rubbish out" aphorism is applicable here). This paper seeks firstly to provide an overview of image representation for image mining by considering an "image representation for image mining" taxonomy. Secondly this paper seeks to present this taxonomy in the context of a particular application domain, the mining of satellite imagery to collect census data.

Broadly, image representation for image mining can be viewed according to whether we wish to consider an entire image or simply one or more elements within an image. The later requires the representation process to be preceeded by a segmentation process so as to isolate the elements of interest (segmentation is outside the scope of this paper). In both cases similar techniques can be used for representation purposes, the distinction is the amount of storage that might be required. The most common representation used for data mining in all its forms is the feature vector representation where we conceptualise the domain of interest in terms of an $n$ dimensional feature space where each dimension is an attribute contained in the domain of interest. Using the feature space concept each example (record or image) is defined in terms of a *feature vector* $V = \{v_1, v_2, ..., v_n\}$ where each element $v_i$ relates to an attribute value in dimension $i$. Thus, using the feature vector mechanism, prior to the application of any mining activity, it is necessary to first preprocess the data so that a collection of feature vectors, $\Phi = \{v_1, V_2, ..., V_m\}$, can be generated. The image mining domain is no exception, the challenge is identifying the image features to be included in the feature space. In this context three categories of representation are considered in this paper: (i) statistical, (ii) graph based and (iii) point series. A second challenge is that frequently the number of dimensions is large (the "curse of dimensionality" aphorism is also applicable here). Many of these dimensions (features) are likely to redundant or not useful. To reduce the set of dimensions a feature selection process is typically adopted. Although not a central theme of this paper a number of feature selection methods ($\chi^2$, Gain ratio and Information gain) were considered respect to the population estimation application domain.

There are a great variety of data mining techniques applicable to image data and data in general [12]. So as to limit the scope of the work presented in this paper the

focus is on prediction (classification) using supervised learning. A process whereby prelabelled training data is required from which a predictor can be "learnt". Thus the desired feature vectors used for training purposes need to include a class attribute value $c$ drawn from a set of such values $C$. Thus in this case the feature vectors are of the form $\{v_1, v_2, ..., v_n, c_i\}$ where $c_i \in C$. To obtain some degree of confidence in a generated predictor a further prelabelled test set is required to which the prediction model can be applied and the generated predictions (classes) compared to the known predictions. Once we are satisfied with the operation of our predictor it can go into usage and be used to label previously unseen data.

To illustrate the ideas concerning image representation presented in this paper a census collection application domain is considered. A census is a mechanism for acquiring and collecting information about a population; a mechanism widely used with respect to a variety of national, and local, government management and planning activities. The most important element of a census is population count. However, census collection, and the associated post processing, is expensive. The UK Office for National Statistics (UKONS) reported that the UK 2011 census cost some £480 million [33]. The US 2010 census was reported to have cost $13 billion [30]. The cost of census collection is also increasing, according to the Australian Bureau of Statistics the Australian 2006 census cost around AUD 300 million, whilst the 2011 census cost around AUD 440 million [35]. The cost with respect to rural areas is typically greater than that of urban areas because the communication and transport infrastructure in rural areas tends to be less well developed. There is also often a lack of good will on behalf of a population to participate in census collection, even if they are legally obliged to do so, because people are often suspicious of the motivation behind censuses.

A potential solution is the usage of technology, namely the internet. However, this requires a literate population and access to the necessary infrastructure. In many parts of the world people remain unconnected to the internet. It is also interesting to note that in the context of the UK 2011 census it was found that a frequently cited reason for households not to have internet access was because of a life style decision not to do so [29]. The solution advocated in this paper is to create a prelabelled training set of household images, extracted from Google Earth, with known family sizes and use this data to build a household size prediction model that can then be used for large scale census collection exercises. Of course it is acknowledged that this approach will not work well in cities where it will be difficult to distinguish buildings in terms of number of inhabitants, however, it was anticipated that the approach would work well in rural areas; areas where census data collection tends to be more of a challenge. In the context of this census collection application domain, and with reference to the proposed image representation for image mining taxonomy, the challenge is how best to represent the satellite household image data. A number of different representations (using the proposed taxonomy) are considered in this paper, evaluated and utilised with respect to a large-scale study featuring a rural area of Ethiopia.

The rest of this paper is structured as follows. In Section 2 a review is presented, founded on work presented in [7], regarding existing work on automating the census

collection process using satellite image data. The proposed image representation for image mining taxonomy is then presented in Section 3. The proposed solution to the automated extraction of census data from satellite imagery, using the proposed taxonomy presented in the previous section, is then presented in 4. Three different catagories of representation are considered and evaluated as presented in Section 5. The main findings are presented in the concluding section, Section 6.

## 2 Previous Work

In this section some discussion concerning previous work on population estimation is presented; the application domain focus for the discussion on image representation for data mining presented later in this paper. Population estimation has been a subject of researched amongst the Geographic Information Systems (GIS) and remote sensing communities for some time. From the literature we can broadly divide this research activity into two categories: (i) area interpolation and (ii) statistical modelling [42]. The work presented in this paper subscribes to the second. Using area interpolation the idea is to use existing census information concerning some geographic area and extrapolate this to obtain a population estimation for a wider or alternative geographic area [24]. Statistical modelling in turn is concerned with the relationship between population size or density and data obtained from GIS and/or satellite imagery.

Existing work on statistical modelling for the purpose of population estimation can be further categorised according to the nature of the data on which the population estimation generation is based, namely: (i) light intensity, (ii) land usage, (iii) dwelling unit count, (iv) image pixel characteristics and (v) physical or socio-economic characteristics. The central idea on which the first is based is that there is a functional relationship between population size and the amount of night time light emanating from an area. Examples can be found in [3, 6, 28, 36, 39], where the relationship between population density and light frequency were used to convert light frequency into a population density estimation. In [36] the reported evaluation was directed at Japan and China, whereas in [6] and [28] it was directed at China only; in [3] the evaluation was directed at a population estimation of the Brazilian Amazon, whilst in [39] the study was directed at the USA.

Work within the second sub-category is directed at the correlation between population density and different types of land usage. The idea is to determine population density according to land usage with respect to a set of one or more sample areas and apply this knowledge to additional areas (as in the case of area interpolation). Land usage categories are typically identified from satellite image data. In [22] it is suggested that population densities for different types of land usage can be determined from sample surveys or census statistics. Four different types of land usage were extracted from four different cities in California, USA, and population densities computed. In [27], six types of land usage were identified in the context of Landsat TM satellite images centred on Atlanta, USA. A regression model was then

applied to produce population densities for Atlanta. A common way of distinguishing land usage types is by using some form of texture analysis. There are a variety pf texture analysis techniques that can be adopted, but one involves the usage of Local Binary Patterns (LBPs) [25, 34], a techniques utilised with respect to the work presented in this paper (see Section 4).

The third sub-category of approach to population size estimation using statistical modelling is to estimate the total dwelling unit count in a defined region and multiply this by an average number of people expected to live in a dwelling unit. There are various ways of obtaining an estimate of the dwelling unit count, but one suggested approach is to estimate this by analysing remote sensing images; a idea also promoted in this paper. In the past, when there was no effective ways of automatically identifying residential buildings within remote sensing imagery, the dwelling units were manually identified from aerial photographs (a laborious and time consuming process). With the advancement of technology and the availability of satellite imagery more advanced feature extraction techniques have been developed for this purpose [13]. In [4] a dwelling unit count based approach is presented using IKONOS satellite images of the Al Shaabia district in Khartoum, Sudan. The dwelling unit count approach has some similarity with respect to the work represented in this paper.

In the fourth sub-category, the relationship between image pixel characteristics and population densities is used for the purpose of population estimation. The image pixel characteristics in question can be represented using a variety of mechanisms, but common examples include mechanisms based on the spectral reflectance values of image pixels and mechanisms based on image texture analysis. Examples of using pixel characteristics for population estimation are presented in [18] and [24]. In [18] a system was presented whereby texture analysis was applied to Google Earth satellite images, using block sizes of $64 \times 64$ and $32 \times 32$ pixels, to estimate population densities with respect to cities in Pakistan. In [24] a variety of features were used, including: spectra signatures, principle components, vegetation indices, fraction images, texture and temperature. These features were extracted from Landsat ETM+ satellite images and used to measure population density in the city of Indianapolis, Indiana, USA.

The final category of population estimation is founded on the usage of various kinds of physical and socioeconomic information which is then interpolated to give population estimations. For example, information about demography, topography and transportation networks have all been used to estimate population size. In [26] a mechanism was presented for estimating population size by determining the correlation between the population in urban areas and the distance to the nearest Central Business Distract (CBD), distance to major roads, slope and the age of the community.

What all the above approaches to population estimation modelling have in common is that they are focussed on regions or areas rather than specific households as in the case of the work presented in this paper. As far as the authors are aware the fundamental approach of estimating populations sizes at the household level, as presented in this paper, is entirely unique.

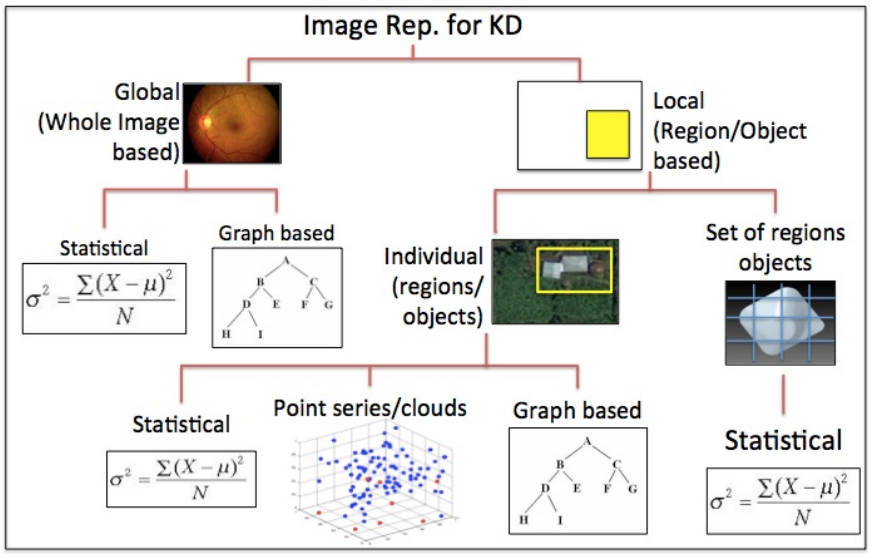



**Fig. 1** Taxonomy for Image Representation for Image Mining.

## 3 The Image Representation Taxonomy

The proposed image representation for image mining taxonomy is presented in Figure 1. From the figure, at a high level and as noted in the introduction to this paper, the image representations can be categorised according to whether we are interested in entire images or some region (object) within an image set. We use the terms *global* and *local* to differentiate between the two. An example where we might wish to consider entire images, as shown in the figure, is in the case of retina image analysis where we typically wish to classify images according to whether they feature some eye condition or not. The most common eye conditions that are considered in this respect are Age-related Macular Degeneration [2, 15, 45, **?**] and Diabetic Retinopathy [1, 37]. Frequently quoted examples where we might wish to consider objects within images are with respect to MRI brain scan data; in 2D the mid-sagital slice is often used. For example in [9] the object of interest was the corpus callosum, the part of the brain that connects the left and right hand sides of the brain, which was analysed in the context of the presence of epilepsy (or not). In [41] the ventricles are considered but in terms of 3D image analysis.

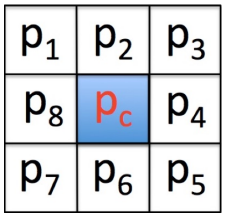
In Figure 1 a distinction is made between non-contiguous region/object applications (such as the census collection based on individual household sizes domain considered later in this paper) and contiguous region/object applications where we have multiple neighbouring regions/objects. In the figure the latter is illustrated with a sheet metal forming 3D image application taken from [21] (see also [11]) where a pre-specified shape is "pressed" out using a sheet metal forming machine. However, the process introduces distortions (referred to as *springback*). The idea presented

in [21] was that if these distortions can be predicted they can be compensated for. Sub-shapes in the manufactured shape were thus isolated and considered to be objects in a 3D image mining exercise. The feature vectors in this case comprised a shape description, not unlike the shape descriptions used in [41] to represent ventricles extracted from MRI scans, and a numeric distortion class label. Where we have multiple disconnected objects in an image these can be processed in the same way as if there were only one object.

Regardless of whether the nature of images are considered globally or locally they can be represented either in terms of a set statistics extracted from the image (object) or in terms of some graph/tree representation. In the case of single objects we can also consider the boundary of the object which can then be represented as a point series or curve. Statistical techniques are the simplest approach to image representation for image mining. The most obvious statistics that may be used are the first order statistical functions such as the mean, variance and standard deviation of the intensity or RGB or grayscale colour values. For example in [44] (see also [43]) eleven different first order statistical features were extracted from a breast biopsy image set in order to predict the presence of breast cancer (or otherwise). In the case of objects we can use morphometrics of various kinds describing the size and/or shape of an object (size can be expressed simply in terms of a pixel count). Such simple statistics often do not work well because they are not expressive enough; however they provide for a good bench mark representation and are considered later in this paper as a means of representing households. A more sophisticated category of statistic involves the usage of second order statistical functions applied to an intermediate representation, examples include: (i) co-occurrence matrices, (ii) gradient analysis, (iii) Hough transforms and (iv) Local Binary Patterns (LBPs). This last is used later in this paper and thus is considered in further detail later in this section. Another example where the LBP concept has been used as an image representation for image mining can be found in [8] where X-ray images of knee joints are encapsulated using LBPs for the purpose of predicting the existence of osteoarthritis (or otherwise).

A LBP is a texture representation method which is statistical in nature [25, 34]. Using the LBP approach a binary number is produced for each pixel, by thresholding its value with its neighbouring pixels. Thus, with reference to Figure 3 the grayscale value for the centre pixel $p_c$ is compared with that of the eight neighbours and a value of 1 recorded if the value for $p_i$ is lower than that for $p_c$, and a value of 0 otherwise. In this manner we get eight binary values making up an eight bit number. There are 256 different options, thus we can generate a 256 dimensional feature space with each dimension having values of between 0 and the maximum number pixels that can exist in any one image in the image set. The example in Figure 3 considers eight neighbouring pixels at a radius of one, thus $8 \times 1$ LBPs. There are other possibilities, for example $8 \times 2$ or $16 \times 2$.

A alternative popular method for representing images is to apply some form of hierarchical decomposition to the image (with respect to both the global and local situations) and to store the result in a quad-tree (for 2D image data) or oct-tree (for 3D image data). Hierarchical decomposition has a well established track record in




$$d = \begin{cases} 1, & \text{if } p_i \leq p_c \\ 0, & \text{otherwise} \end{cases}$$

**Fig. 2** Ilustration of the LBP Concept.

the context of image analysis [31, 38, 40, **?**]. An example decomposition is given in Figure 3, based on [10], where the Corpus Callosum featured in a 2D MRI brain scan image has been segmented and decomposed (down to a maximum decomposition level of 3) and rendered as a quad tree. Once we have a collection of tree represented images/objects we can apply a subgraph mining technique (a good review of such techniques is given in [19]) to the tree set and extract frequently occurring sub-trees where frequent is defined in terms of some frequency count threshold $\sigma$. A popular frequent subgraph mining algorithm used for this purpose is the gSpan algorithm [17] adapted for the purpose of frequent sub-tree mining rather than frequent sub-graph mining. The set of extracted frequent subgraphs can then be viewed as features in a $n$-dimensional binary-valued feature space where $n$ is the number of sub graphs and each dimension has two values: present and not present. Issues with hierarchical decomposition include: (i) the boundary problem where regions appear in different branches of the tree and (ii) when to stop the decomposition (using either a critical function to measure homogeneity or a pre-specified maximum level of decomposition).

Although quad trees are the most commonly encounter tree-based formalism other types of tree (and graph) format can be adopted. In Figure 4 (taken from [14]) an alternative decomposition is shown with respect to a retina image. In this case the decomposition alternates between "angular" and "circular" division. Angular division involves partitioning using a minor arc to divide a region into two, whilst circular decomposition involves dividing a region into two using a radius emanating from the centre of the image. At the top level the image is divided into four quadrants; at subsequent levels the decomposition is conducted in a binary manner as indicated by the example tree shown on the right of the image.

Another popular mechanism for representing images is as a point series (or curve). Although not indicated in Figure 1 this can also be applied globally. The simplest form of point series is a histogram, which can be directly translated into a feature vector representation. For example histograms of intensity values, orientation gradients or LBPs. Alternatively, given an object of interest contained within an image, we can represent the boundary in terms of a point series using, for example, the concept of chain coding. Given a collection of labelled point series, representing a set of images or objects within an image set, a new image/object can be classified directly using (say) the well-established KNN algorithm ($k = 1$ is often used).

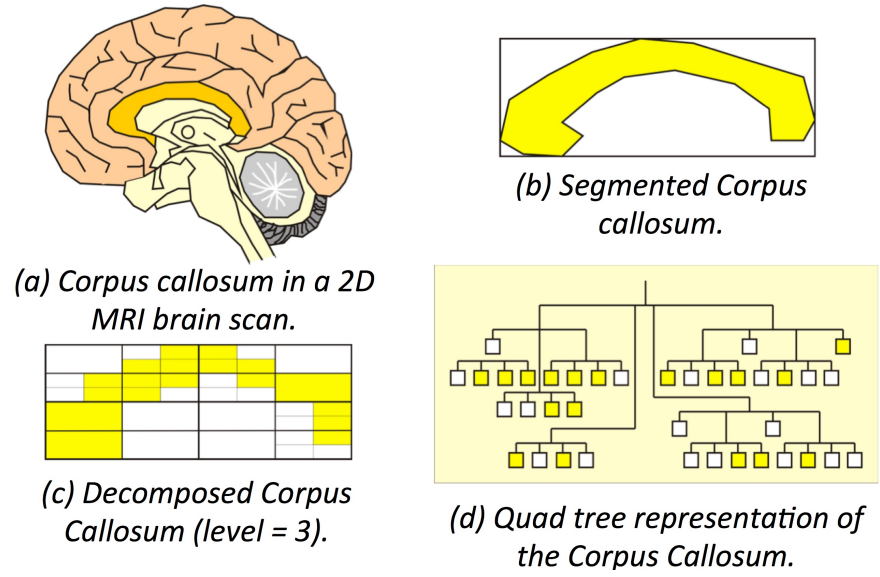



**Fig. 3** Hierarchical decomposition, an example using MRI brain scan data.

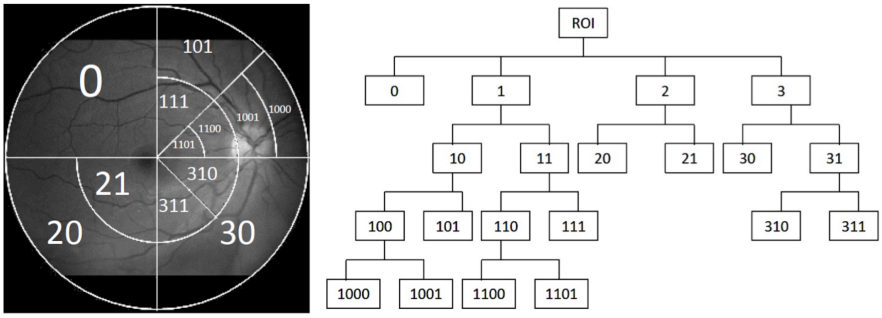



**Fig. 4** Circular whole image decomposition with respect to a retina image (max level of decomposition = 4) and Associated tree structure [14].

When using algorithms such as KNN we need an appropriate similarity measure; Euclidean distance is frequently used as a comparison measure but requires the point series to be of the same length. Alternatively, we can look to work on time series analysis [20], for example the use of Dynamic Time Warping (DTW) which produces a warping path distance defining the difference between two point series [5, 32]. DTW has the added advantage that the point series to be compared do not have to be of the same length. The later was used in [9] to define shapes in 2D MRI brain scan data and in [41] to define 3D MRI brain scan shapes.

**Fig. 5** Google Earth image featuring a village in the Horro district of Ethiopia [7].

## 4 Census Prediction Model Generation

In this section we return to the census collection estimation application domain. Recall that the idea is to build a household size predictor using prelabelled household images extracted from Google Earth and then to use this predictor to produce population estimations of large areas. Recall also that to act as a focus for this work a rural area of Ethiopia was selected. More specifically the district of Horro located 300 km to the northwest of Addis Ababa. An example Google earth image from this area is presented in Figure 5. Inspection of the image indicates a large number of households.

To collect the required training data an "on the ground" team visited sample households at two sites, Site A and Site B, within the district and collected family size information together with the latitude and longitude of each household so that the associated Google satellite images could be retrieved. At the time the data was collected Google Earth did not readily facilitate the automated extraction of satellite imagery, so instead the Google Static Map Service was used. This featured an API that allowed users to download satellite images (one image at a time) specified according to various parameter settings: (i) latitude and longitude of the centre of the area of interest. (ii) image size (in pixels) and (iii) zoom Level (level of detail). An image size of $1280 \times 1280$ pixels and a zoom level of 18 was used. Each household was surrounded with a $256 \times 256$ pixel bounding box defined so as to cover the largest anticipated household (by superimposing a box we do not have issues with irregular shaped household plots). In this manner data for 120 households was obtained, 70 households for Site A and 50 for Site B. The distinction between the two sites was that for Site A the available Google Earth images were obtained in the "wet season" and so were mostly green, while those obtained for Site B were obtained during the dry season so were mostly brown. Some statistics concerning this training data are given in Table 1 and Figure 6. Note from the table that the

population sizes have been grouped according to three class labels: (i) Small, (ii) Medium and (iii) Large. The reason for this was for prediction purposes categorical classification systems as well as regression models (which produced a real value) were considered; as described in further detail later in this section.

| Family | Min. | Max. | Ave. | Mode | Site | Site |
|---|---|---|---|---|---|---|
| Small | 2 | 5 | 4.04 | 5 | 38 | 19 |
| Medium | 6 | 8 | 7.00 | 6 | 32 | 21 |
| Large | 9 | 12 | 9.80 | 9 | 10 | 10 |
| All | 2 | 12 | 6.31 | 6 | 70 | 50 |

**Table 1** Statistics for training data (Sites A and B).

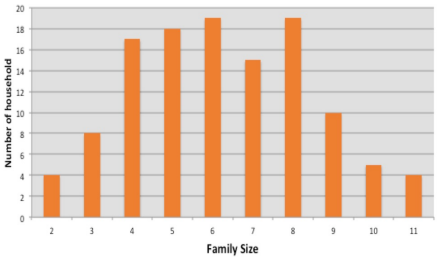



**Fig. 6** Histogram for training data household population sizes (Sites A and B) [7].

Given this training data the next stage was to represent the households using an appropriate mechanism compatible with prediction model generation. To this end each of the three categories of representation identified in the taxonomy presented in Section 3 was used and a comparison conducted. Details concerning each individual representation are given in Sub-sectons 4.1, 4.2 and 4.3 below; and the conducted comparative evaluation reported on in Sub-section 4.4.

## *4.1 Satistics Based Image Representation*

For the statistics based representation LBPs were used, generated as described in Section 3. LBPs with eight neighbours and a radius of one were used ($8 \times 1$ LBPs). Experiments were conducted (not reported here) using other LBP configurations but no advantage was found. An example of the process of converting a Google household image to an LBP image is given in Figure 7. The left hand image shows the raw Google image, the middle image the associated grayscale image to which the LBP mechanism was applied and the right hand image the resulting LBP rendition of the original Google image.

Once counts for each of the 256 possible LBPs had been obtained a 256 element feature vector could be generated, one for each household image. Feature selection was then applied so as to reduce the overall number of dimensions and retain dimensions which were good discriminators of household size. A subset of the LBPs was thus retained. A number of feature selection methods were considered but $\chi^2$ feature selection, with $k = 40$ (where $k$ is the number of dimensions to be retained), was found to produce the best result. Consideration was also given to augmenting the LBP representation with additional statistics (such as contrast, correlation, energy, homogeneity) but this was also found to have little effect.

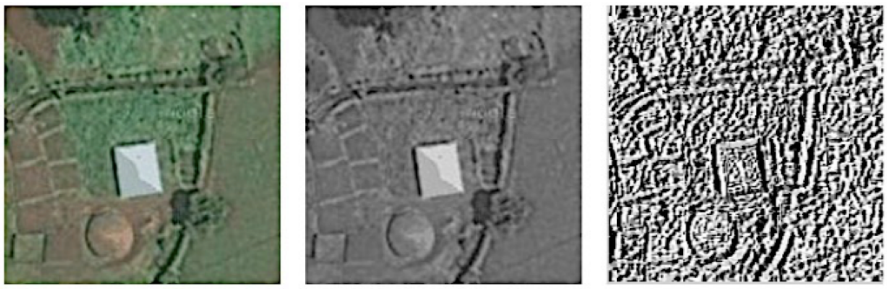


**Fig. 7** Process of converting a Google household image to an LBP image [7].

### *4.2 Tree Based Image Representation*

The tree-based image representation generation process is illustrated in Figure 8 (see also Figure 3) where we start with a colour Google image (not shown), convert this to a grayscale household image and the apply the decomposition. In this manner a quadtree was generated representing each household in the training data. The nodes were labelled with a greyscale encoding generated using a mean intensity of the greyscale colours in each region; for this purpose eight labels were derived, each describing a range of 32 consecutive intensity values. Frequent Sub-graph Mining (FSM) was then applied to the tree collection as discussed in Section 3. A variation of gSpan was used, but other FSM algorithms would be equally applicable. A $\sigma$ value of 10 was used for the FSM; in other words for a sub-graph to be considered frequent it had to appear in 10% of the tree represented images in the training set. Note that low $\sigma$ values are better (nothing will be missed), however many more Frequent Sub-Graphs (FSGs) will be identified than when a higher $\sigma$ value is used. A feature selection strategy was thus adopted so as to reduce the number of dimensions in a manner whereby only highly discriminative features were retained. In this case gain ratio feature selection, with $k = 55$, was found to produce the best result. Each record was then presented in a feature vector format ready for prediction model generation.

### *4.3 Point Series Based Image Representation*

For the point series based household mage representation seven different colour histograms were generate: three describing the three channels in the RGB image colour formalisation, three describing the three channels in the HSV image formalisation, and one using the greyscale formalisation. For each histogram 32 bins were used, thus feature vectors measuring $7 \times 32 = 224$ elements were generated. The authors again experimented with including statistical measures, but it was again found that this made no difference (and in some cases proved to be decremental). Feature se-

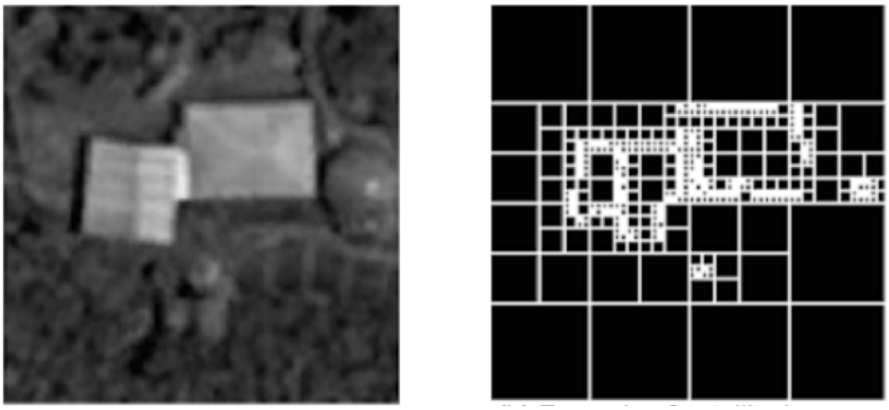


**Fig. 8** Hierarchical decomposition process for an example Google household image [7].

lection was also applied to reduced the size of the feature space. As before a number feature selection strategies were considered ($\chi^2$, Gain ratio and Information gain). Gain ration Feature Selection with $k = 25$ was found to give the best results.

## *4.4 Household Image Representation Evaluation*

This sub-section presents a comparison of the above suggested image representation techniques, one for each representation category included in the proposed image representation for image mining taxonomy. The comparison was conducted by generating and testing predictors using each representation. For this purpose two different categories of prediction model were considered: (i) classification models and (ii) regression models; the distinction being that the first is used to predict a class label while the second is used to predict a real value.

In the classification case three different classes were considered (as given in Table 1) and six classifier generation models: (i) the Bayesian Network (BN) model, (ii) the Neural Network (NN) model, (iii) Logistic Regression (LR), (iv) Sequential Minimal Optimisation (SMO), (v) Averaged One Dependence Estimation (AODE) and (vi) the well known C4.5 decision tree generation algorithm (C4.5). These were coupled with $\chi^2$ and Gain ratio feature selection (with different values of $k$). The metrics used for the evaluation were: (i) Accuracy (AC), (ii) Area Under receiver operating Characteristic (AUC), (iii) the F-measure (FM), (iv) Sensitivity (SN) and (v) Specificity (SP). The best results, generated using Ten Cross Validation (TCV), are given in Table 2 (derived from work included in [7]), with the very best results highlighted in bold font. For the statistics-based household image representation $\chi^2$ feature selection with $k = 40$ was used. For the tree-based image representation $\sigma = 10$ was used for FSG mining and gain ration feature selection with $k = 55$. For the point series representation gain ration feature selection was also used but

| Classification | Site A | | | | | Site B | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Model Generator | AC | AUC | FM | SN | SP | AC | AUC | FM | SN | SP |
| Stats. based (LR) | **0.771** | 0.859 | **0.778** | **0.771** | **0.885** | 0.680 | 0.756 | 0.679 | 0.680 | 0.803 |
| Stats. based (NN) | **0.771** | **0.881** | 0.759 | **0.771** | 0.852 | 0.720 | 0.824 | 0.718 | 0.720 | 0.825 |
| Tree Based (AODE) | 0.629 | 0.815 | 0.627 | 0.629 | 0.753 | 0.800 | 0.863 | 0.785 | 0.800 | 0.871 |
| Tree based (BN) | 0.600 | 0.808 | 0.596 | 0.600 | 0.734 | **0.800** | **0.879** | **0.792** | **0.800** | **0.876** |
| Tree based (NN) | 0.686 | 0.819 | 0.685 | 0.686 | 0.782 | 0.620 | 0.789 | 0.628 | 0.620 | 0.829 |
| Tree based (SMO) | 0.729 | 0.791 | 0.727 | 0.729 | 0.818 | 0.620 | 0.733 | 0.610 | 0.620 | 0.781 |
| Point based (BN) | 0.700 | 0.807 | 0.687 | 0.700 | 0.782 | 0.700 | 0.798 | 0.692 | 0.700 | 0.829 |
| Point Series (C4.5) | 0.671 | 0.724 | 0.668 | 0.671 | 0.760 | 0.500 | 0.598 | 0.499 | 0.500 | 0.718 |
| Point based (LR) | 0.657 | 0.822 | 0.662 | 0.657 | 0.806 | 0.640 | 0.821 | 0.633 | 0.640 | 0.798 |

**Table 2** Household Image Representation Evaluation using classification models and TCV.

with $k = 25$. From the table it is interesting to note that the "best" representation depends on whether we have wet (green) season or dry (brown) season image data. For the wet season the statistics based representation, using LBPs and coupled with either LR or NN, provided best results; while for the dry season the tree based representation (coupled with BN) provided the best results.

In the context of the linear regression models a number of models were considered: (i) Linear regression (Linear Reg.), (ii) Least Median Squared regression (LMedS), (iii) Isotonic Regression (IsoReg) and (iv) Support Vector Machine regression (SVMreg). Each was considered in isolation and when coupled with different feature selection strategies. Each was applied in the context of the satellite household image training data expressed using the different representations discussed above. The metrics used for comparison purposes in this case were: Correlation Coefficient (Coef), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Note that metrics used for evaluating classification models, where we wish to predict a categorical class, tend to be different to those that are typically used for evaluating regression models where we are predicting a real value. Best results, again generated using TCV, are presented in Table 3 (also derived from work included in [7]), with the very best results again highlighted in bold font. In this case the LBP statics-based image representation outperformed the other representations hence only results using the LBP representation are shown. The table also gives results with and without the application of a feature selection strategy. A number of such strategies were considered, but Correlation-based Feature Selection (CFS) was fund to produce the best results. From the table it can be seen that the best performing regression model was SVM regression.

## 5 Large Scale Study

Once a prediction model has been generated and tested so that an appropriate degree of confidence can be attached to the model it can be placed into service. This section considers firstly how the models, generated as described in the previous section, can

| Regression Method | Site A | | | Site B | | |
|---|---|---|---|---|---|---|
| | Coef | MAE | RMSE | Coef | MAE | RMSE |
| LinearReg | -0.080 | 2.167 | 2.570 | 0.274 | 1.981 | 2.407 |
| LMedS | -0.288 | 3.262 | 3.894 | 0.215 | 1.952 | 2.353 |
| IsoReg | -0.309 | 2.382 | 2.841 | 0.156 | 1.940 | 2.295 |
| SVMreg | -0.279 | 3.367 | 3.970 | 0.308 | 1.778 | 2.056 |
| LinearReg+CFS | 0.084 | 2.145 | 2.550 | 0.400 | 1.727 | 2.093 |
| LMedS+CFS | 0.252 | 1.988 | 2.373 | 0.428 | 1.687 | 2.038 |
| IsoReg+CFS | -0.202 | 2.287 | 2.706 | 0.109 | 1.912 | 2.282 |
| SVMreg+CFS | **0.307** | **1.957** | **2.330** | **0.587** | **0.143** | **1.802** |

**Table 3** Household Image Representation Evaluation using regression models and TCV.

be applied in the context of regional census collection. The section firstly presents the process whereby such a census might be conducted and secondly considers the effectiveness of the result by considering a particular benchmark region. The test area chosen for this purpose was an entire village and its surrounding lands within the Horro district. The reasons why this area was chosen was because this area was similar to the areas from which the prediction model training data was obtained and because the population size of this village was known; in 2011 the village was reported to comprise 459 households and a population of $3,223$ (thus ground truth data was available).

The rest of this section is organised as follows. Sub-section 5.1 describes the satellite data collection process using the Google Static Map Service API. Once the satellite image data has been collected the images need to be segmented to identify "household images"; the mechanism whereby this was conducted is presented in Sub-section 5.2. To ensure no data was missed an overlap was used when collecting the satellite image data, thus it was possible that specific households would appear in more than one image. It was thus necessary to first remove such duplicates before any further processing could be conducted. The duplicate household detection and pruning process is considered in Sub-section 5.3. Once the household images had been identified they could be represented using one of the image representations considered above, to which any of the prediction models also considered above could be applied. To evaluate the process the best performing prediction models (see Sub-section 4.4) were applied to the data. The results are presented and discussed in Sub-section 5.4.

## *5.1 Satellite Image Data Collection*

In total 600 Satellite images, covering the area of interest were collected using the Google Static Map Service API. An image size of $1280 \times 1280$ pixels and a zoom level of 18 was used; because these were the parameters used for the training set collection. Using the Google Static Map Service API images are downloaded in an

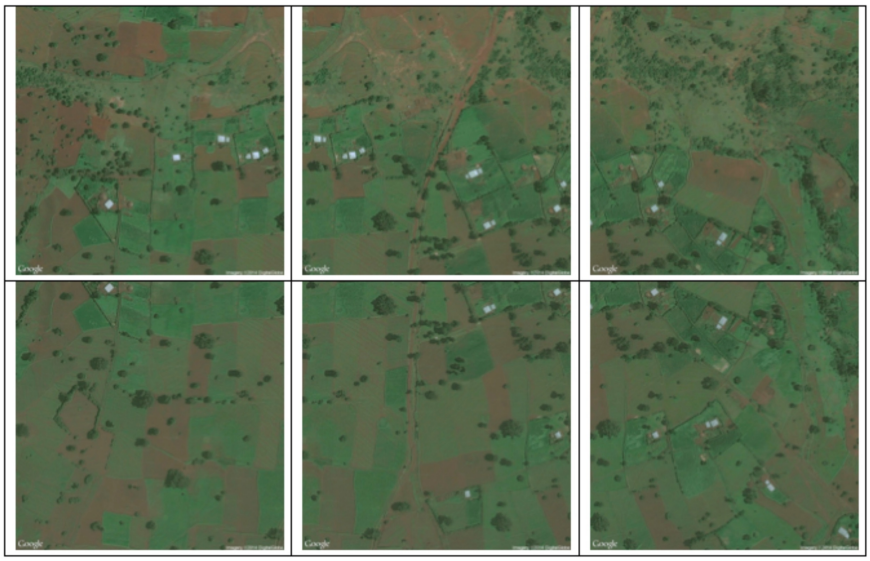


**Fig. 9** Fragment of collected satellite image patchwork [7].

iterative manner image by image. A 320 pixel overlap was used, designed so that every household will appear in its entirety in at least one collected satellite image. For this to operate correctly it was necessary to: (i) convert the top-left corner latitude and longitude of the current image into $x$ and $y$ pixel values, (ii) add the required offset to obtain the top-left $x$ and $y$ coordinates of the next image in the sequence, (iii) convert these new $x$ and $y$ coordinates back to a latitude and longitude and (iv) repeat; a time-consuming process. Note that Cartesian coordinates are planer values while latitude and longitude are geoidal values, thus conversion was also not straight forward; note that the Google Static Map Service uses the EGM96 spheroid (Earth Gravitational Model 1996). It took 356 seconds to collect the 600 required satellite images. Together these images formed a "patchwork" covering the area of interest. A fragment of this patchwork is shown in Figure 9.

## *5.2 Image Segmentation*

The downloaded satellite images could contain zero, one or more households. It was thus necessary to segment the images so as to identify households. As noted previously, the typical household comprised at least one building with a tin roof that was readily discernable (see Figure 5). Visual inspection of a sample of the images indicated that this was true in all the cases sampled. This feature could therefore be usefully employed to identify households in the collected satellite image data. Note that with respect to the training data (Section 4) segmentation was not required

| Channel | Min | Max |
|---|---|---|
| Hue | 0.35 | 0.65 |
| Saturation | 0.05 | 0.15 |
| Value | 0.80 | 1.00 |

**Table 4** Adopted HSV threshold values for household image segmentation.

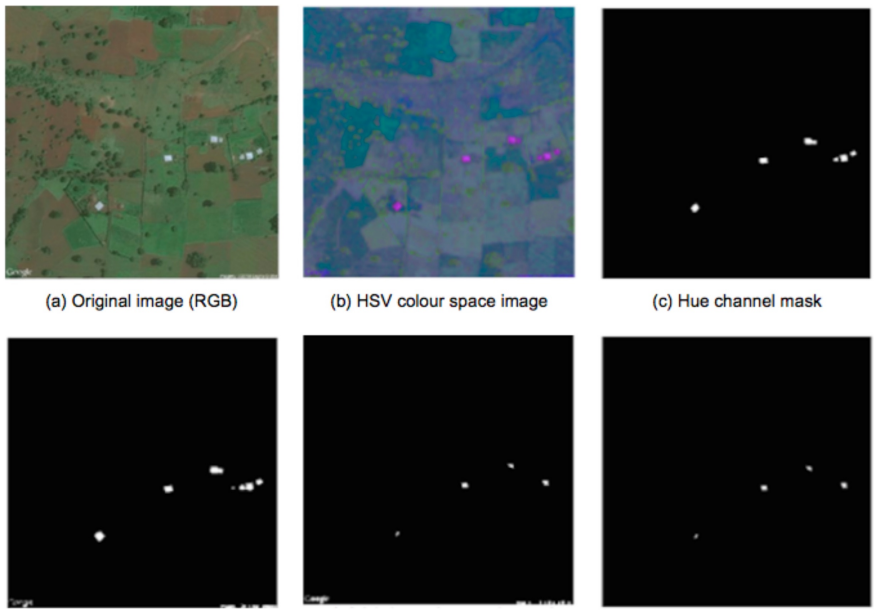



**Fig. 10** Illustration of The Satellite Image Segmentation Process [7].

because we knew where the households were because their latitude and longitude had been collected as part of the knowledge acquisition process.

The segmentation was conducted using a number of image masks. Experiments were conducted using a variety masking techniques (a significant challenge was the illumination of roads and water ways). The most appropriate mechanism was found to be when the HSV representation was used together with a set minimum and maximum thresholds. Given an image represented using one of the HSV channels, pixels with values below and above the threshold were set to black and the remaining pixels within the threshold range to white. Thus three masks were produced: (i) hue, (ii) saturation and (iii) value. By combining these masks pixels set to white in all three masks were identified as households. Extensive experimentation (not reported here) was conducted to determine the most appropriate threshold values, the selected values are given in Table 4. The entire segmentation process is illustrated in Figure 10 where we have an originnl image featuring four households, the image translated into the HSV colour space, three masks (hue, saturation and value) and the final result.

On completion of the segmentation process each household was represented as a "blob" of white pixels. The centroid of each blob was considered to be its location, described in terms of latitude and longitude coordinates, and this location was identified in the original image. Each location in the original image was then surrounded by a $w \times w$ bounding box ($w = 256$ was used as this was the same value used for the prediction model training as described in Section 4). In this manner a collection of household images was obtained. Note that the minimum bounding boxes will be smaller and/or non-symmetrical near the edges of each image.

### *5.3 Duplicate Detection and Pruning*

Using the above process 526 household images were identified. However, this included duplicate households; households that appear in more than one image. Inspection of the Figure 9 indicates a number of duplicate households, some appearing in two images and in some cases in four images. Such duplicate households had thus to be pruned before any further processing could be conducted. The duplicate detection and pruning process was as follows. The identified households were listed in order of latitude. This list was then processed and households with the same latitude and longitude label (within a level of tolerance) identified. If two households with the same centroid latitude and longitude both comprised $256 \times 256$ pixel boxes the later was pruned. If the boxes were unequal in size the household featuring the smaller sized box was pruned.

Using the above process a total of 526 households were detected including duplicates. Duplicate detection identified 100 duplicate households, thus 426 out of a known number of 459 households were identified. Suggested reasons for the discrepancy were as follows. There was a two year time difference between the ground truth survey and the satellite images; a period during which some households may have fallen into disuse (manual inspection of a proportion of the collected satellite images indicated that some buildings did indeed appear to be roofless, thus supporting this conjecture). Inspect of the satellite imagery indicated that a small number of buildings were very poorly defined and in some cases had not been segmented correctly. It was also possible that the duplicate household detection mechanism had detected some duplicates that were in fact not duplicates (although no evidence for this was found). The overall run time to segment and process the collected satellite image data was 1,370 seconds (22.8 minutes), about 2.28 seconds per satellite image and 2.6 seconds per household.

### *5.4 Population Estimation Results and Evaluation*

Once an appropriate set of household satellite images had been generated previously derived classification and/or regression models (of the form described in the

| Representation | Prediction Model | Feature Selec. Strat. | Population Estimation | Accuracy (%) | Total Run Time (Mins.) |
|---|---|---|---|---|---|
| Statistics-based (LBPs) | Neural Network classifier generated using Site A wet season data | $\chi^2$ | 2,545 | 78.96 | 29.49 |
| Graph-based ($\sigma = 10$) | Bayesian Network classifier generated using Site B dry season data | Gain Ratio ($k = 55$) | 2,495 | 77.41 | 35.42 |
| Statistics-based (LBPs) | SVM Linear regression generated using Site A wet season data | CFS | 2,548 | 79.06 | **29.48** |
| Statistics-based (LBPs) | SVM Linear regression generated using Site B dry season data | CFS | **2,760** | **85.63** | **29.48** |

**Table 5** Population Estimation Results.

| Family size | Average household size ($a$) | Predicted num. households ($b$) | Estimated population size ($a \times b$) |
|---|---|---|---|
| Small | 4.04 | 156 | 630 |
| Medium | 7.00 | 261 | 1827 |
| Large | 9.80 | 9 | 88 |
| Total | | 426 | 2545 |

**Table 6** Estimation of population size with respect to the Neural Network classification model generated using the Site A data set.

previous section) could be applied and an overall population estimation extracted from the image data. In the case of classification models a class label was produced for each household, to turn this into a population estimation each class needed to be translated into a number of persons and then summed to give a total number of persons. In the case of the classification models described in Section 4 the average number of persons associated with each class was used. In the case of the regression models a population size was derived directly.

The obtained results using the best performing classifiers/predictors identified in Sub-section 4.4 (see also [7]) are presented in Table 5 (best results highlighted in bold font): (i) Neural Network classifier generated using statistics-based (LBP) Site A wet season data , (ii) Bayesian Network classifier generated using Graph-based Site B dry season data, (iii) SVM Linear regression generated using statistics-based (LBP) Site A wet season data and (iv) SVM Linear regression generated using statistics-based (LBP) Site B dry season data. The calculation of the individual population sizes using Neural Network and Bayesian Network classification is given in Tables 6 and 7 respectively (derived from [7]). The best performing approach used a statistics-based representation coupled with a SVM Linear Regression model. This produced a population estimation of 2,760 compared to a "ground truth" of 3,223; thus an accuracy of 86%.

| Family size | Average household size ($a$) | Predicted num. households ($b$) | Estimated population size ($a \times b$) |
|---|---|---|---|
| Small | 4.04 | 226 | 7913 |
| Medium | 7.00 | 135 | 945 |
| Large | 9.80 | 65 | 637 |
| Total | | 426 | 2495 |

**Table 7** Estimation of population size with respect to the Bayesian Network classification model generated using of the Site B data set.

An accuracy of 86% might be argued to be unsatisfactory, however, we can point to a number of reasons for the difference between the predicted and "ground truth" population sizes. Firstly the data from which the classification (regression) models were generated might not reflect the data to which they were applied as closely as was anticipated. Measures for determining the similarity between satellite image data sets are a subject for future work. Secondly, as already noted, there was a two year time lag between the date of the census collection (2011) and the date of the satellite image acquisition (2013). Manual inspection of a number of images showed signs of derelict (abandoned) households. It may thus be the case that between 2011 and 2013 depopulation had taken place and that the produced population estimates were in fact a better reflection of population size than initially thought. There have been recent reports concerning the depopulation of rural Ethiopia [16]. Thirdly, and again as already noted previously, census collection is often viewed with suspicion. Local authorities may suspect that it is to be used for the levying of a local tax and thus there may be an incentive to under report population size. Alternatively it may be suspected that the census is to be used for allocating development funds in which case there may be an incentive to over report.

## 6 Conclusions

In this paper we have presented a taxonomy for image representation for image mining together with an illustration of the practical application of the taxonomy in the context of an automated census data collection application. The main premise is that although the data mining algorithms that we might wish to apply to image data are well understood the end to end Knowledge Discovery in Data (KDD) process is less well established. The main challenge is how to represent image data in such a way that the salient features are maintained while at the same time ensuring compatibility with the data mining algorithms to be applied. The proposed taxonomy, at a high level, differentiates between global representations and local representations; the first being directed at applications where we wish to consider image data in its entirety and the second where we wish to consider one or more objects within individual images. A distinction was also made between objects that are connected

(contiguous) and not connected. In the taxonomy, again at a high level and regardless of whether we are considering images in their entirety or at a local level, we can identified three categories of representation: (i) statistical, (ii) tree (or graph) -based and (iii) point series based. The particular nature of the individual representations that can be included in the categories depends on whether we are working at a global or local level. At the local level we can, for example, consider the nature of the boundary of the objects of interest which would not be applicable at the global level.

The taxonomy was applied to a census estimation application domain so as to illustrate the usage of the ideas presented by the taxonomy with respect to a novel application domain. The motivation for the application was the resource required to collect census data. The idea was to build predictors to predict individual household sizes from satellite images of households. It was noted that although the idea would not work well in urban areas it would work well in rural areas where the cost of census collection is the greatest. Three exemplar representations were considered, one from each category: (i) a quadtree representation from which frequently occurring sub-trees were extracted and used to generate feature vectors (one per household), (ii) a statistics-based representation founded on the use of LBPs and (iii) a point series representation founded on the use of collections of histograms. Training and test data was obtained from two sites (Site A and Site B) in a rural area of Ethiopia featuring households with a known location and "family size". The known locations were used to obtained Google Satellite images of the individual households. The distinction between the two sites was that for Site A the satellite imagery was obtained during the wet (green) season whilst that for Site B was obtained in the dry (brown) season. The collated individual household images were then represented, using the three selected exemplar representations, to produce three versions of the data. This data was then used to train predictors. Two categories of predictor model were considered. Classification models where a "family size" class label was predicted ("small". "medium" or "large"), and regression models where an actual household size was produced. A range of classification and regression models were considered coupled with different feature selection mechanisms. Of these two classification models and two regression models were found to give the best performance when evaluated using the training/test data and TCV, as follows:

1. Statistics-based using LBPs, Neural Network classification (generated using Site A data) and $\chi^2$ feature selection.
2. Tree-based using Bayesian Network classifier generated (generated using Site B data) and Gain Ratio feature selection.
3. Statistics-based using LBPs, SVM Linear regression (generated using Site A data) and CFS feature selection.
4. Statistics-based using LBPs, SVM Linear regression (generated using Site B data) and CFS feature selection.

The prediction models were then applied to a wider area (but in the same rural region of Ethiopia), an entire village, where the number of households and population size was known. The best performing approach was found to be the LBP Statistics-

based representation coupled with a SVM Linear regression model (generated using the Site B data) and CFS feature selection. An accuracy of 85.63% was recorded. Although (at face value) the population estimation produced was not as accurate as the ground truth census data (this was to be expected), the proposed method offered significant cost and time savings. A number of reasons as to why the prediction was not identical to the "ground truth" value can be identified:

1. The training data from which the prediction models were generated might not reflect the data to which they were applied as closely as was anticipated. Measures for determining the similarity between satellite image data sets are a subject for future work
2. There was a two year time lag between the date of the census collection (2011) and the date the satellite images were acquired (September 2013). Manual inspection of a number of images indicated signs of derelict (abandoned) households. It was thus conjectured that it might be the case that between 2011 and 2013 depopulation had occurred and that the produced population estimate was in fact a better reflection of population size than initially thought.
3. Census collection is often viewed with suspicion therefore there may have been incentives to over or under report and therefore the "ground truth" value night not have been entirely accurate (it should not be regarded as a "gold standard").

Whatever the case the results indicated that by using the proposed framework effective population estimates can be obtained, in rural areas, at a very low cost (almost zero).

With respect to image mining in general it can be observed that decisions are regularly made with the support of imagery of some sort (Satellite Image, MRI, OCT, and so on). It can also be observed that our ability to collect imagery of all kinds (both 2D and 3D) has enhanced rapidly over the last decade (we can do it cheaper and faster); we have seen a rapid growth in the global image sensor market. There is substantial benefit to be gained from applying image mining to this image date although it is essential that appropriate image representation is used. There is also a lot of scope for alternative representations, especially fuzzy and deep learning approaches and lots of scope for further application. A further issue to be addressed is explanation mechanisms to give reasons as to why particular predictions and/or classifications were arrived at with respect to previously unseen images; this is of particular relevance with respect to medical imaging applications.

## 7 Acknowledgements

The authors wold like to thank the following whose ideas helped formulate the contents of this paper: (i) Abdulrahman Albarrak from the Department of Computer Science at The University of Liverpool, (ii) Ashraf Elsayed from the Department of Computer Science at the University of Alexandria, (iii) Marta García-Fiñana from the Department of Biostatistics at the University of Liverpool, (iv) Hanafi Hijazi

from the School of Engineering and Information Technology at the University of Malaysia Sabah, (v) Vanessa Sluming from the School of Health Science at the University of Liverpool, (vi) Akadej Udomchaiporn from King Mongkut's Institute of Technology Ladkrabang and (vii) Yalin Zheng from the department of Eye and Vision Science at the Royal Liverpool University Hospital.

## References

1. Albarrak, A., Coenen, F., Zheng, Y.: Classification of Volumetric Retinal Images Using Overlapping Decomposition and Tree Analysis. Proc. 26th IEEE International Symposium on Computer-Based Medical Systems (CBMS'13), pp11-16 (2013).
2. Albarrak, A., Coenen, F. and Zheng Y.: Volumetric Image Classification using Homogeneous Decomposition and Dictionary Learning: A Study Using Retinal Optical Coherence Tomography for Detecting Age-related Macular Degeneration. Journal of Computerized Medical Imaging and Graphics, 55, pp113-123 (2016).
3. Amaral, S., Monteiro, A.V.M., Câmara, G., Quintanilha, J.A.: DMSP/OLS night time light imagery for urban population estimates in the Brazilian Amazon. International Journal of Remote Sensing, 27(5), pp855-870 (2006).
4. Al Salman, A.S., Ali, A.E.: Population estimation from high resolution satellite imagery: A case study from khartoum. Emirates Journal for Engineering Research, 16(1), pp63-69 (2011).
5. Berndt, D.j., Clifford, J.: Using dynamic time warping to find patterns in time series. Proc. AAAI Workshop on Knowledge Discovery in Databases, pp229-248 (1994).
6. Cheng, L., Zhou, Y., Wang, L., Wang, S., Du. C.: An estimate of the city population in China using DMSP night-time satellite imagery. In Proc. IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS), pp691-694 (2007).
7. Dittakan, K.: Population Estimation Mining From Satellite Imagery. PhD thesis, University of Liverpool (2015).
8. Dittakan K., Coenen, F.: Detection of Osteoarthritis Using Local Binary Patterns: A Study Directed at Human Joint Imagery. Proc. 14th Pacific Rim International Conference on Artificial Intelligence (PRICAI 2016), Springer LNCS 9810, pp 93-105 (2016).
9. Elsayed, A., Hijazi, M.H.A., Coenen, F., García-Fiñana, M., Sluming, V., Zheng, Y.: Classification of MRI Brain Scan Data Using Shape Criteria. Annals of the British Machine Vision Association (BMVA), Vol 2011, No 6, pp1-14 (2011).
10. Elsayed, A., Coenen, F., García-Fiñana, M., Sluming, V.: Region Of Interest Based Image Classification: A Study in MRI Brain Scan Categorization. In Karahoca, A. (Ed.), Data Mining Applications in Engineering and Medicine, InTech - Open Science, Slavka Krautzeka, Croatia, pp225-248 (2012).
11. El Salhi, S., Coenen, F., Dixon C., Khan, M.: Predicting Springback Using 3D Surface Representation Techniques: A Case Study in Sheet Metal Forming. Journal of Expert Systems With Applications, Volume 42, Issue 1, pp79-93 (2014).
12. Han, J., Kamber, M., Pei, J.: Data Mining. Concepts and Techniques, 3rd edition, Morgan Kaufmann (2011).
13. D. Haverkamp. D.: Automatic building extraction from IKONOS imagery. Proc. Annual Conference of the American Society for Photogrammetry and Remote Sensing (2004).
14. Hijazi, M.H.A., Coenen F., Zheng, Y.: Data Mining Techniques for The Screening of Age-Related Macular Degeneration. Journal of Knowledge Based Systems, 29, pp83-92 (2012).
15. Hijazi, M.H.A., Coenen F., Zheng, Y.: Data Mining for AMD Screening: A Classification Based Approach. International Journal of Simulation Systems, Science and Technology (IJSSST), 15(2), pp64-68 (2015).

16. Hamza, I.A., Iyela, A.: Land use pattern, climate change, and its implication for food security in Ethiopia: A review. Ethiopian Journal of Environmental Studies and Management, 5, pp26-31 (2012).
17. Huan, J., Wang, W., Prins, J." Efficient Mining of Frequent Subgraph in the Presence of Isomorphism. In Proceedings of the 2003 International Conference on Data Mining (ICDM'03), 549-561, (2003).
18. Javed, Y., Khan, M.M., Chanussot, J.: Population density estimation using textons. Proc. IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS'12), pp2206-2209 (2012).
19. Jiang, C., Coenen, F., Zito, M.: A Survey of Frequent Subgraph Mining Algorithms. Knowledge Engineering Review, Vol 28, Number 1, pp75-105 (2013).
20. Karter J.: Time Series Analysis with Matlab. CreateSpace Independent Publishing Platform (2016).
21. Khan, M., Coenen, F. Dixon, C., El Salhi, S., Penalva, M., Rivero, A. (2014). An Intelligent Process Model: Predicting Springback in Single Point Incremental Forming. International Journal of Advanced Manufacturing Technology, pp1-12, (2014).
22. Kraus, S.P., Senger, L.W., Ryerson, J.M.: Estimating population from photographi- cally determined residential land use types. Journal of Remote Sensing of Environment, 3(1), pp35-42 (1974).
23. Krizhevsky, A., Sutskever. I., Hinton, G.: ImageNet classification with deep convolutional neural networks. In: Proc. NIPS (2012).
24. Li, G., Weng.Q.: Using Landsat ETM+ imagery to measure population density in Indianapolis, Indiana, USA. Journal of Photogrammetric Engineering and Remote Sensing, 71(8), pp63-69 (2005).
25. Liang, P., Li, S.F., Qin, J.W.: Multi-resolution local binary patterns for image classification. In Proceedings of the Twentieth International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), pp164-169 (2010).
26. Liu X., Clarke, K.: Estimation of residential population using high resolution satellite imagery. Proc. Third International Symposium on Remote Sensing of Urban Area, pp153-160 (2002).
27. Lo. C.: Zone-based estimation of population and housing units from satellite-generated land use/land cover maps. Remotely Sensed Cities, pp157-180 (2003).
28. Ma, T., Zhou, C., Pei, T., Haynie, S., Fan, J.: Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: A comparative case study from Chinas cities. Journal of Remote Sensing of Environment, 124, pp99-107 (2012).
29. Madden, P., Goodman, J., Green, J., Jenkinson, C.: Growing pains: Population and sustainability in the UK. Technical report, Forum for the future (2010).
30. Mather, M., Pollard, K., Jacobsen, L.A.: Report on America: First results from the 2010 census. Technical report, Population Reference Bureau, Washington, DC, USA (2011).
31. Montanvert, A., Meer, P. Rosenfield, R.: Hierarchical image analysis using irregular tessellations IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(4), pp307-316 (1991).
32. Myers, C.S., Rabiner, L.R.: A comparative study of several dynamic time-warping algorithms for connected word recognition. The Bell System Technical Journal, 60(7), pp1389-1409 (1981) .
33. Office for National Statistics.: National population projections, 2010-based statistical bulletin. Technical report, Office for National Statistics (2011).
34. Pietikäinen, M.: Image analysis with local binary patterns. In Proceedings of Scandinavian conference on Image Analysis (SCIA), Springer-Verlag, pp115-118 (2005).
35. Pink. B.: Census of population and housing: Nature and content australia 2011. Technical report, Australian Bureau of Statistics (2008).
36. . Pozzi, F., Small, C., Yetman, G.: Modeling the distribution of human population with nighttime satellite imagery and gridded population of the world. Proc. Future Intelligent Earth Observing Satellites conference (2002).

37. Pratt, H., Coenen, F., Broadbent, D.M., Harding, S.P., Zheng, Y. : Convolutional Neural Networks for Diabetic Retinopathy. Proc. Medical Image Understanding and Analysis (MIUA'16), Procedia Computer Science, Volume 90, pp200-205 (2016).
38. Samet, H.: The quadtree and related hierarchical data structures. ACM Computing Surveys, 16(2), pp187-260 (1984).
39. Sutton, P.: Modeling population density with night-time satellit imagery and GIS. Computer environment and urban sustems, Elsevier, pp227-244 (1997).
40. Tadmor, E., Nezzar, S., Vese, L.: Multiscale hierarchical decomposition of images with applications to deblurring, denoising and segmentation. Communications in Mathematical Sciences, 6(2), pp281-307 (2008).
41. Udomchaiporn, A., Coenen, F., García-Fiñana, M. and Sluming, V.: 3-D Volume of Interest Based Image Classification. Proc. 14th Pacific Rim International Conference on Artificial Intelligence (PRICAI 2016), Springer LNCS 9810, pp 543-555, (2016).
42. Wu, S.S., Qiu, X., Wang, L.: Population estimation methods in GIS and remote sensing: A review. Journal of GIScience and Remote Sensing, 42(1), pp80-96 (2005).
43. Zhang, Y., Zhang, B., Coenen, F. and Lu, W.: Breast Cancer Diagnosis from Biopsy Images with Highly Reliable Random Subspace Classifier Ensembles. Machine Vision and Applications, Vol 24, pp1405-1420 (2013).
44. Zhang, Y., Zhang, B., Coenen, F., Xiao, J. and lu, W.: One-Class Kernel Subspace Ensemble for Medical Image Classification. EURASIP Journal on Advances in Signal Processing, Vol 2014:17, 13 pages (2014).
45. Zheng, Y., Hijazi, M.H.A., Coenen, F.: Automated "Disease/No Disease" Grading of Age-Related Macular Degeneration by an Image Mining Approach. Investigative Ophthalmology and Visual Science, 53(13), pp8310-8318 (2012).