1. Hsiang, Y. H., Hertzberg, R., Hecht, S. and Liu, L. F., Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I. *J. Biol. Chem*., 1985, **260**, 14873–14878.
2. Thomas, C. J., Rahier, N. J. and Hecht, S. M., Camptothecin: current perspectives. *Bioorg. Med. Chem*., 2004, **12**, 1585–1604.
3. Pantazis, P., Han, Z., Chatterjee, D. and Wyche, J., Water-insoluble camptothecin analogues as potential antiviral drugs. *J. Biomed. Sci*., 1999, **6**, 1–7.
4. Sadaie, M. R., Mayner, R. and Doniger, J., A novel approach to develop anti-HIV drugs: adapting non-nucleoside anticancer chemotherapeutics. *Antiviral Res*., 2004, **61**, 1–18.
5. Bodley, A. L. and Shapiro, T. A., Molecular and cytotoxic effects of camptothecin, a topoisomerase I inhibitor, on trypanosomes and *Leishmania*. *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 3726–3730.
6. Ramesha, B. T. *et al.*, New plant sources of the anti-cancer alkaloid, camptothecine from the Icacinaceae taxa, India. *Phytomedicine*, 2013, **20**, 521–527.
7. Suma, H. K., Kumar, V., Senthilkumar, U., Kumara, P. M., Ravikanth, G., Santhoshkumar, T. R. and Shaanker, R. U., *Pyrenacantha volubilis* Wight. (Icacinaceae), a rich source of camptothecine and its derivatives from the Coromandel Coast forests of India. *Fitoterapia*, 2014, **97**, 105–110.
8. Hemalatha, R. G. and Pradeep, T., Understanding the molecular signatures in leaves and flowers by desorption electrospray ionization mass spectrometry (DESI MS) imaging. *J. Agricul. Food. Chem*., 2013, **61**, 7477–7487.
9. Ifa, D. R., Srimany, A., Eberlin, L. S., Naik, H. R., Bhat, V., Cooks, R. G. and Pradeep, T., Tissue imprint imaging by desorption electrospray ionization mass spectrometry. *Anal. Methods*, 2011, **3**, 1910–1912.
10. Mohana Kumara, P., Srimany, A., Ravikanth, G., Uma Shaanker, R. and Pradeep, Y., Ambient ionization mass spectrometry imaging of rohitukine, a chromone anti-cancer alkaloid, during seed development in *Dysoxylum binectariferum* Hook.f (Meliaceae). *Phytochemistry*, 2015, **116**, 104–110.
11. Li, B., Bjarnholt, N., Hansen, S. H. and Janfelt, C., Characterization of barley leaf tissue using direct and indirect desorption electrospray ionization imaging mass spectrometry. *J. Mass Spectrom*., 2011, **46**, 1241–1246.
12. Thunig, J., Hansen, S. H. and Janfelt, C., Analysis of secondary plant metabolites by indirect desorption electrospray ionization imaging mass spectrometry. *Anal. Chem.*, 2011, **83**, 3256–3259.
13. Yan, X. F., Wang, Y., Yu, T., Zhang, Y. H. and Dai, S. J., Variation in camptothecin content in *Camptotheca acuminata* leaves. *Bot. Bull. Acad. Sin*., 2003, **44**, 99–105.
14. Cabral, E., Mirabelli, M., Perez, C. and Ifa, D., Blotting assisted by heating and solvent extraction for DESI-MS imaging. *J. Am. Soc. Mass Spectrom*., 2013, **24**, 956–965.
15. Srimany, A., Ifa, D. R., Naik, H. R., Bhat, V., Cooks, R. G. and Pradeep, T., Direct analysis of camptothecin from *Nothapodytes nimmoniana* by desorption electrospray ionization mass spectrometry (DESI-MS). *Analyst*, 2011, **136**, 3066–3068.
16. Borisjuk, L., Walenta, S., Rolletschek, H., Mueller-Klieser, W., Wobus, U. and Weber, H., Spatial analysis of plant metabolism: sucrose imaging within *Vicia faba* cotyledons reveals specific developmental patterns. *Plant J*., 2002, **29**, 521–530.
17. Alves, M., Sartoratto, A. and Trigo, J. R., Scopolamine in *Brugmansia suaveolens* (Solanaceae): defense, allocation, costs and induced response. *J. Chem. Ecol*., 2007, **33**, 297–309.
18. Fang, J., Reichelt, M., Hidalgo, W., Agnolet, S. and Schneider, B., Tissue-specific distribution of secondary metabolites in rapeseed (*Brassica napus* L.). *PLoS. ONE*, 2012, **7**, e48006.
19. Telek, L., Delpin, H. and Cabanillas, E., *Solanum mammosum* as a source of solasodine in the lowland tropics. *Econ. Bot*., 1977, **31**, 120–128.
20. Sharma, N. S., Varghese, S., Desai, J. and Chinoy, J. J., Biosynthesis of solasodine in developing berries of *Solanum khasianum* Clarke. *Indian. J. Exp. Biol.*, 1979, **17**, 224–255.
21. Kozukue, N. and Friedman, M., Tomatine, chlorophyll, *β*-carotene and lycopene content in tomatoes during growth and maturation. *J. Sci. Food. Agric*., 2003, **83**, 195–200.
22. Hebbar, R., Sashidhar, V. R., Uma Shaanker, R., Udaya Kumar, M. and Sudharshana, L., Dispersal mode of species influences the trypsin inhibitor levels in fruits. *Naturwissenschaften*, 1993, **80**, 519–521.

ACKNOWLEDGEMENTS. We thank the Department of Biotechnology (No. BT/PR8266/NDB/39/266/2013) and Department of Science and Technology, Government of India for financial support. P.M.K. thanks IIT Madras, for a postdoctoral fellowship and A.S. thanks the Council of Scientific and Industrial Research, Government of India for research fellowship.

Received 12 December 2015; accepted 27 September 2016

doi: 10.18520/cs/v112/i05/1034-1038

A cognitive method for building detection from high-resolution satellite images

Naveen Chandra* and Jayanta Kumar Ghosh

Department of Civil Engineering, Indian Institute of Technology, Roorkee 247 667, India

In recent years, high-resolution satellite (HRS) images have become an important source of data for extracting geo-spatial information. A deep understanding of human cognitive capabilities is required in order to automate the method of information retrieval from HRS images. The aim of this study is to emulate human cognitive processes by integrating cognitive task analysis for information extraction from HRS images. First, the preliminary knowledge about the cognitive processes which human beings acquire during the interpretation of satellite images is collected. Then, knowledge is represented in the form of rules which are based on the visual interpretation of the images by the human beings. During knowledge elicitation these rules are used to extract buildings from HRS images utilizing the mixture tuned matched filtering algorithm. Later, the method is tested using 14 HRS images of an urban area. The average of precision, recall and *F*-score is computed as 79.45%, 64.34% and 70.28% respectively.

Keywords: Building detection, cognitive processes, high-resolution satellite images, urban areas.

*For correspondence. (e-mail: naveenchandra0408@gmail.com)

BUILDING detection from urban areas has been an active field of research in computer vision[1]. It has also been a topic of interest in many applications such as change detection and urban monitoring. With the availability of very high resolution (VHR) images, different types of algorithms/methods have been proposed for building detection[2]. A review on object detection and its applications can be found in the literature[3–8]. The present approach of building detection is categorized into two parts (using three-dimensional images and monocular remotely sensed dataset) based on the data source (SAR, multispectral and LiDAR images). Initially, data-driven method was used for extracting low-level features from the monocular images[9]. Analysing the shadow information of buildings in monocular images is the prime concern for obtaining better accuracy. Shadow information is utilized for identifying the corner and edges of buildings[10]. Then, height of the buildings is obtained based on the shadow information. Shadows are used extensively for the verification of the building detection methods which have been proposed earlier[11–13]. In the past, few methods for building detection have been developed which are based on supervised classification algorithm[14]. Later, support vector machines (SVMs), graph theoretical tools and scale-invariant feature transform (SIFT) were used for detecting buildings from aerial images[15,16]. Recently, a new method for building detection has been developed by integrating shadow information with fuzzy logic and GrabCut partitioning algorithm[17]. Then a system was proposed for building detection using laser scanning data[18]. Some studies have used graphical models to improve the overall accuracy of the system[19,20]. A digital classification based method[21] was also used for detecting human settlements from SAR and optical images. Later, Markov random field and conditional random field models were used for object detection from an urban area[22]. Table 1 provides a brief review of the work done in the past in the context of building detection. In this communication, we describe a cognitive method for detecting buildings from high-resolution satellite (HRS) images. First, we briefly describe the cognitive approach used for building detection. Then we present the results obtained followed by conclusions of the study.

In this study, the process of building detection is combined with cognitive task analysis (CTA)[23]. CTA is used in psychological research where lot of decision making task are involved. CTA is a hierarchical method which defines the various psychological processes which are being used by an analyst for performing a complex task. It illustrates the different input parameters and cognitive capacities that are taken into account for obtaining the output of the task. CTA makes the process of information extraction easier as it uses the thought process of human beings for carrying out a task. It is defined as an extension of the traditional task analysis (TTA) approaches to provide information about thought processes and knowledge in order to determine performance of the observer. CTA generally uses different types of observations and interviews to determine knowledge which a subject uses while performing a complex task. Here, a complex task is defined as one which requires the use of automated (strategic and unconscious) and controlled (conceptual and conscious) knowledge to complete it. Much time is required for completing the task. CTA is the only approach which is used for describing knowledge necessary for evaluating the overall performance. CTA is used by an analyst to determine the precise and complete information of cognitive processes. The output is generally a description of procedural and conceptual knowledge which is used by the analyst while performing the task. This output is formatted so that it can be maintained as a record to be used later by a novice to complete the task. In the past, three techniques were defined for CTA, i.e. (i) process tracing, (ii) observation and interviews, and (iii) conceptual methods. Later, one more technique was introduced known as formal models (computational models). The key advantage of CTA is that it generates precise and detailed information on the basis of the nature of the subject performing the task. It is a rich source of information if implemented properly. CTA does not use hit or trail approach; rather it delivers a systematic approach to discover the cognitive processes of the subject. However, much time is required for analysing and verifying the data collected during CTA. The key steps for carrying out CTA are described here, which have been used for performing a complex task of building detection from a satellite image of an urban area. Figure 1 shows the overall architecture of the methodology used.

In the first phase, the analyst determines the flow of the task which is required for CTA. The analyst develops elementary understanding of the knowledge domain (image analysis) and identifies the method which will be used during knowledge elicitation. During this stage different techniques are used for collecting knowledge, such as observation, document analysis and interviews (structured and unstructured)[23]. In this study document analysis has been used for knowledge collection. The analyst begins by collecting reviews from the available resources, which explain the task related to the domain in which CTA is being carried out. These resources consist of various documents such as textbooks, research articles, reports, handbooks and glossaries. These documents are reviewed for obtaining the deep knowledge and for the confirmation of ideas that have been implemented in the past. The output of this phase is used by the analyst to identify the structures and the type of knowledge required for performing the complex task.

On the basis of the information acquired in the previous stage, the analyst deeply examines each and every task for determining the knowledge and sub-task which will be required for cognitive analysis of satellite images. In previous studies different methods such as semantic

Table 1. Description of previous work

Author	Data type	Image type	Article type	Year
Huertas and Nevatia	Airborne	Grey scale	Research	1988
Irvin and Mckeown	Airborne	Grey scale	Research	1989
Liow and Pavlidis	Airborne	Grey scale	Research	1990
Shufelt and Mckeown	Airborne	Grey scale	Research	1993
McGlone and Shufelt	Airborne	Grey scale	Research	1994
Weinder and Forstner	Airborne	Grey scale	Research	1995
Krishnamachari and Chellappa	Airborne	Grey scale	Research	1996
Baillard	Elevation	Grey scale	Research	1998
Zang	Spaceborne	Multispectral	Research	1999
Stassopoulou and Caelli	Airborne	Grey scale	Research	2000
Cord *et al.*	Elevation	Grey scale	Research	2001
Ruther *et al.*	Elevation	Grey scale	Research	2002
Lee *et al.*	Spaceborne	Grey scale	Research	2003
Benediktsson *et al.*	Spaceborne	Grey scale	Research	2004
Brenaner	–	–	Review	2005
Hongjian and Shiqiang	LIDAR	Grey scale	Research	2006
Sohn and Dowman	Spaceborne and airborne	Multispectral	Research	2007
Katartzis and Sahli	Airborne	Multispectral	Research	2008
Karantzalos and Paragios	Spaceborne and airborne	Grey scale	Research	2009
Haala and Kada	–	–	Review	2010
Cui *et al.*	Airborne	Multispectral	Research	2011
Tack *et al.*	Spaceborne	Multispectral	Research	2012
Senaras *et al.*	Spaceborne	Multispectral	Research	2013

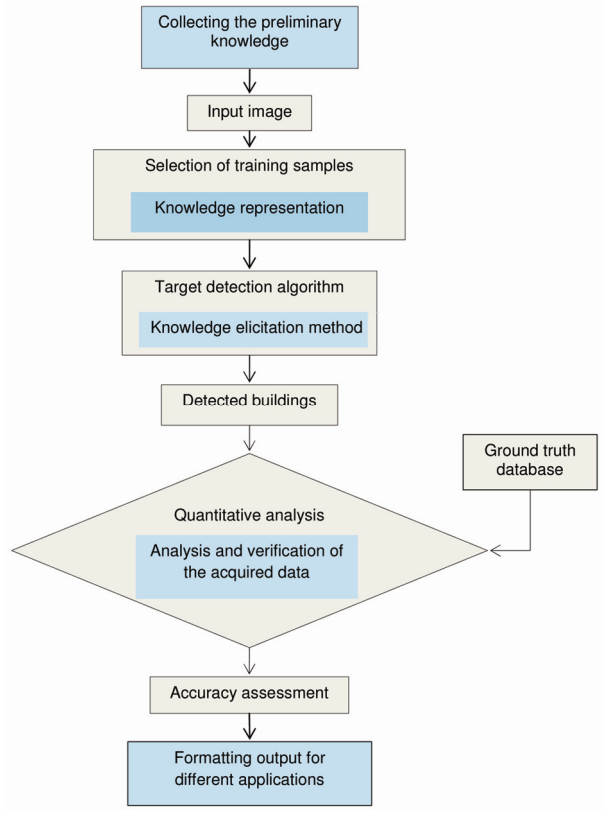

Figure 1. Overall architecture of the methodology used.

network, flow chart and concept maps have been used for knowledge representation[23]. In this study, rule-based method is taken into account for knowledge representation. First, a training dataset is prepared for the input image. The objects in the satellite image which are visually interpreted as buildings and non-buildings by human beings are labelled manually. Later, these interpretation results are used as rules for object detection from satellite images.

In the third phase, the analyst uses different methods for combining the knowledge acquired during the second phase. Several knowledge elicitation techniques (simulation method, prototyping method and observation method)[23] have been introduced in past in order to determine the information necessary for solving a complex problem. Elicitation is defined as a process of collecting knowledge or information from a variety of available resources. Due to the availability of different knowledge elicitation methods, it is difficult to select an appropriate method. In previous literature[24], three types of knowledge elicitation methods are discussed.

(i) Direct- and indirect-based: One of the simplest approaches is to directly collect information about a particular task from the subject expert. In this approach, information is gathered by asking questions from the domain expert. However, in the indirect approach the output of the knowledge elicitation needs to be analysed and verified rigorously so that the required information can be extracted.

(ii) Interaction based: In this approach the obtained knowledge is grouped together based on the interaction (interviewing, teachback, critiquing, sorting and case studying) with the subject expert.

(iii) Based on the type of knowledge acquired: Knowledge elicitation techniques can be grouped on the basis of the type of information (such as relationships, procedures

and evaluation) which has been acquired. Another key approach which falls in this category is known as classification. This method is used to classify different objects and entities in a particular domain. In order to detect the buildings from HRS images, a strong target detection method is needed. Therefore, mixture tuned matched filtering (MTMF)[25], which is a classification method and also a powerful target detection method is used here. MTMF algorithm detects the desired target from the HRS image based on the rules defined in the previous stage. These rules are defined on the basis of the cognitive interpretation of human beings. The MTMF algorithm is divided into three different stages:

(a) First, the forward minimum noise fraction (MNF) transformation of the reflection data is computed[25]. The MNF transformation is basically principal component transformation (PCT)[26] in which the information is segregated into two parts, where the first part is determined with compared feature image and the larger eigenvalues, and the second part is fixed with comparative eigenvalues[27,28]. The key advantage of MNF is that it can easily identify the relationship among several bands of the input image. It has the capability to collect data in smaller parts. The feature space isolates the spectral information of different components and frail dataset is updated during denoising phase. Due to this the seperability of features gets maximized[26].

(b) Matched filtering for determination of abundance estimation.

(c) Determination of false-positive pixels using mixture tuning (MT)[25]. The matched filter vector (MFV)[25] is determined using eq. (1) given below.

$$\mathrm{MFV} = \frac{(\mathrm{CV}_{\mathrm{mnf}}^{-1} \times \mathrm{TS}_{\mathrm{mnf}})}{(\mathrm{CV}_{\mathrm{mnf}}^{-1} \times \mathrm{TS}_{\mathrm{mnf}} \times \mathrm{CF}_{\mathrm{mnf}}^{\mathrm{T}}}, \quad (1)$$

where $\mathrm{CV}_{\mathrm{mnf}}^{-1}$ represents the diagonal inverse of covariance matrix for the MNF dataset and $\mathrm{TS}_{\mathrm{mnf}}$ represents the target spectrum that is converted to MNF space.

Further, matched filtering image ($\mathrm{MF_I}$)[25] is determined using eq. (2).

$$\mathrm{MF_I} = (\mathrm{MFV} \times \mathrm{DS}_{\mathrm{mnf}}), \quad (2)$$

where $\mathrm{MF_I}$ represents the output image obtained after matched filtering and $\mathrm{DS}_{\mathrm{mnf}}$ represents the MNF data.

Finally, MT is estimated using eq. (3).

$$\mathrm{MT}_i = \left(\sum \left(\left(\frac{D_{\mathrm{mnfi}} - \mathrm{dmv}_i}{\mathrm{IVeval}_i} \right)^2 \right) \right)^{1/2}, \quad (3)$$

where MT_i represents the mixture tuned value for pixel i, D_{mnfi} represents the MNF spectrum for a pixel i, dmv_i represents the mean value for pixel i and IVeval_i represents the value of the interpolated vector of eigenvalues for pixel i.

MTMF is a well-known spectral unmixing algorithm based on the concept of signal processing which uses linear mixing and statistical matched filtering model[29]. It is able to detect different classes of land cover using their spectral properties[29]. The process of unmixing is carried out by identifying abundance of the endmember. Further, the response of endmembers is maximized, while the response of unidentified background is minimized[30]. Whenever unmixing is applied for target detection, it is not necessary that an accurate value of target abundance will be extracted. The results of the image classification will be more productive if the estimated value of the abundance fraction for the desired target pixel vector is adequate to discriminate that pixel from neighbouring pixels[31].

It has been observed that the outcome of CTA is dependent on the procedure used during knowledge elicitation. Therefore, a qualitative and quantitative analysis is necessary to validate the output obtained after knowledge elicitation. In this phase, the output image obtained after knowledge elicitation is compared with the corresponding ground truth of the input dataset for calculating the overall accuracy.

In this last phase all the output of CTA (statistical and theoretical results) is organized in the form a report. This output is used in different automated applications. The CTA used must be compared with other available methods in order to reach a decision. The cognitive method implemented here can be used in various decision-making tasks in remote-sensing domain.

The benchmark dataset used here includes 14 test images[32] obtained from two satellites, namely Quick Bird and IKONOS-2 having a resolution of 0.60 m and 1 m respectively. All the test images consist of four multispectral bands (blue, green, red and near infrared) with a radiometric resolution of 11 bits per band. The ground-truth data for each test image are produced manually by an expert human operator[32]. The key point regarding this dataset is that buildings that are partially visible are also included in the ground truth dataset.

To evaluate the overall performance of the method, three standard quality measures (precision, recall and F-score) given in eqs (4)–(6) below are used[33,34].

$$\mathrm{Precision} = \frac{\| \mathrm{TP} \|}{(\| \mathrm{TP} \| + \| \mathrm{FP} \|)}, \quad (4)$$

$$\mathrm{Recall} = \frac{\| \mathrm{TP} \|}{(\| \mathrm{TP} \| + \| \mathrm{FN} \|)}, \quad (5)$$

$$F\text{-score} = \frac{(2 * \mathrm{Precision} * \mathrm{Recall})}{(\mathrm{Precision} + \mathrm{Recall})}. \quad (6)$$

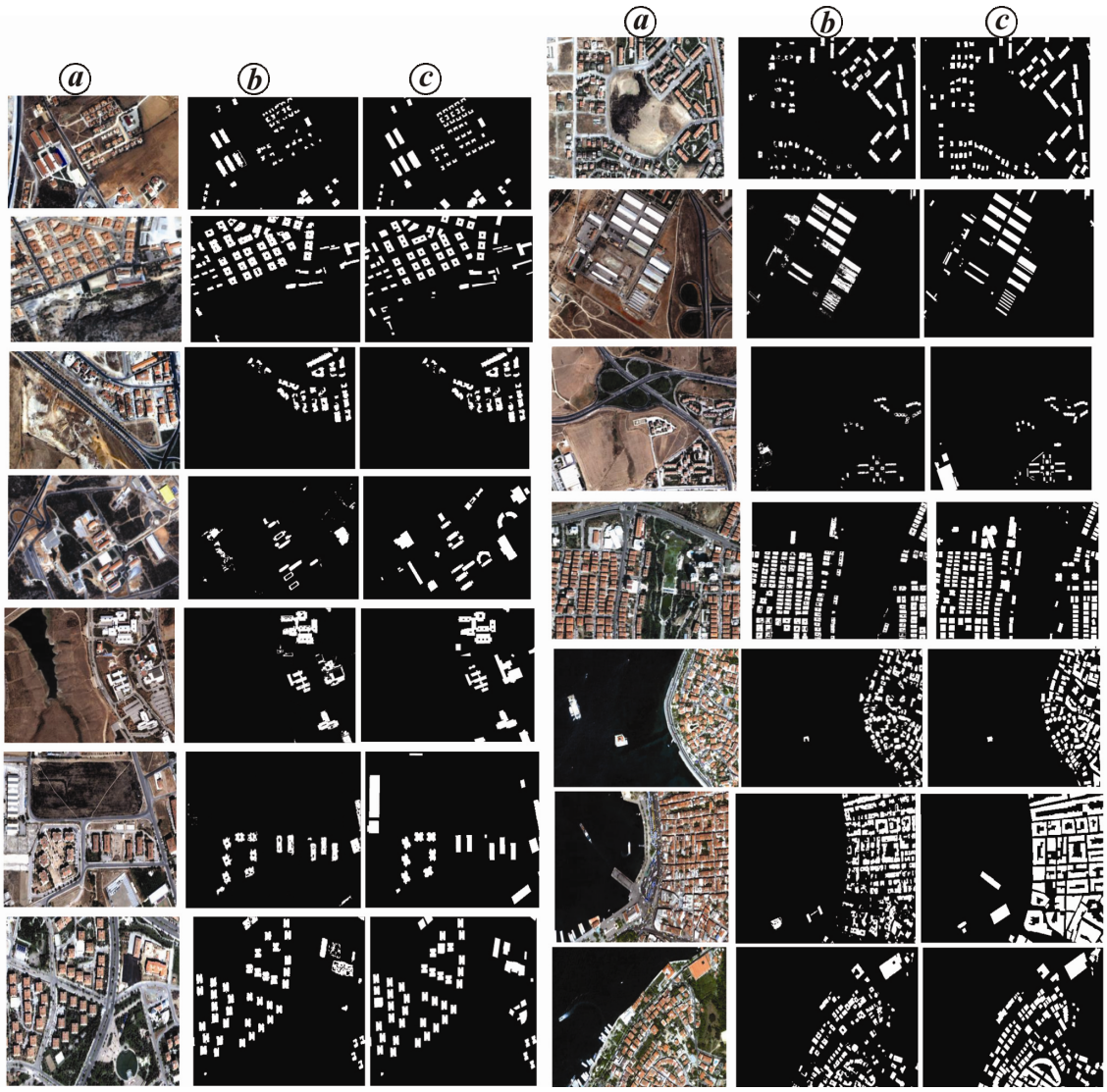

Figure 2. Output of the building detection method: ***a***, test images, ***b***, detected buildings; ***c***, corresponding ground truth.

During the assessment all the pixels of the image were classified into three different classes, namely true positive (TP), false positive (FP) and false negative (FN)[35]. TP indicates a pixel that is labelled as a building by the proposed method and also represents a building in the ground truth dataset. FP signifies a pixel that does not represent any of the pixels labelled as buildings in the ground truth dataset, while FN represents a pixel that is labelled as a building in the ground truth dataset but it is not available in the proposed method. In eqs (4) and (5) ||.|| denotes the number of pixels assigned to each class and *F*-score is the combination of precision and recall into a single score.

The qualitative assessment of this approach is performed through visual inspection of the results illustrated in Figure 2. On the basis of visual analysis, it can be seen that all the true pixels of buildings are clearly detected in test patch 7. Further, performance of the method is below average for test patch 4 because several false pixels are detected along with few true pixels of the buildings. The proposed cognitive method could not clearly detect the boundaries of the buildings; however, still it performs well for all test images.

Tables 2 and 3 show the quantitative and statistical results of the method respectively, while Figure 3 shows the comparative results. In the entire dataset, maximum precision (93.46%) is produced by test patch 14 and minimum precision (41.19%) by test patch 4, whereas maximum recall (86.38%) is produced by test patch 2 and minimum recall (18.99%) by test patch 4. Further, maximum *F*-score (85.27%) is produced by test patch 7 and minimum *F*-score (25.99%) by test patch 4. The overall average of precision, recall and *F*-score is computed as 79.45%, 64.34% and 70.28% respectively, which shows promising results for such a complex dataset. This building detection benchmark dataset consists of buildings of different shapes and sizes, and the method has resulted in a fair performance.

Table 2. Quantitative performance of the method

Dataset	MTMF			MPP[32]		
Test image ID (size)	Precision (%)	Recall (%)	*F*-score (%)	Precision (%)	Recall (%)	*F*-score (%)
1 (560*367)	73.50	70.42	71.93	35.1	48.8	40.8
2 (554*483)	79.46	86.38	82.77	41.0	75.7	53.2
3 (468*304)	84.95	73.13	78.60	40.3	63.7	49.4
4 (896*600)	41.19	18.99	25.99	39.2	36.1	37.6
5 (1213*958)	79.61	74.76	77.11	46.2	82.2	59.2
6 (922*634)	77.91	51.50	62.01	41.9	59.2	49.1
7 (928*639)	90.53	80.60	85.27	56.1	69.8	62.2
8 (1009*695)	85.94	67.81	75.80	49.9	49.7	49.8
9 (1615*1209)	72.60	77.65	75.04	37.8	67.7	48.5
10 (1656*1240)	68.89	35.63	46.97	32.3	43.7	37.1
11 (1222*915)	85.36	75.50	80.13	63.9	53.7	58.4
12 (1311*848)	88.39	67.25	76.39	70.8	56.7	62.9
13 (1193*771)	90.59	55.93	69.16	67.9	60.2	63.9
14 (1193*772)	93.46	65.21	76.82	76.7	66.5	71.2
Average	79.45	64.34	70.28	52.7	59.9	56.1
Maximum	93.46	86.38	85.27	76.7	82.2	71.2
Minimum	41.19	18.99	25.99	32.3	36.1	37.1

MTMF, Mixture tuned matched filtering algorithm; MPP, Marked point process model.

Table 3. Statistical results of the target area

Image ID	Target count	Total area (pixels)	Average area (pixels)
1	365	18,659.000	51.120548
2	128	36,918.000	288.42188
3	257	12,157.000	47.303501
4	884	144,181.00	163.10068
5	457	178,066.00	389.64114
6	904	93,636.000	103.57964
7	197	63,572.000	322.70050
8	739	85,107.000	115.16508
9	1251	278,329.00	222.48521
10	4391	657,679.00	149.77887
11	669	219,124.00	327.53961
12	168	95,550.000	568.75000
13	229	145,320.00	634.58514
14	152	91,316.000	600.76318

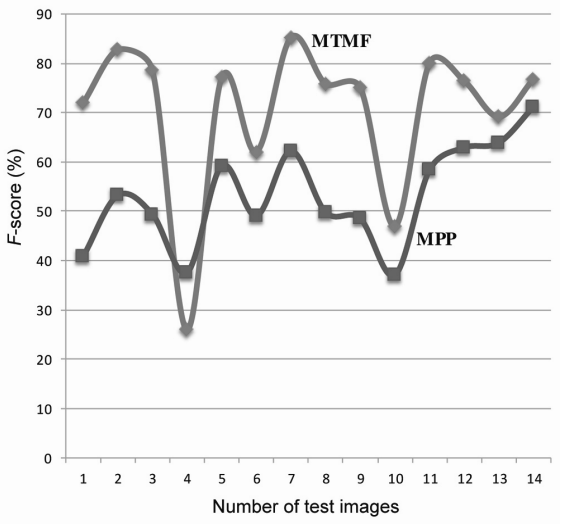

Figure 3. Comparitive results of the *F*-score.

The computational complexity was also taken into consideration, and it was noticed that the time taken for processing of the data was directly dependent on the size of the input image. Lager the size of input image more will be the computational time, while smaller the size of input image, lesser will be the computational time. Maximum time was taken by test patch 10, whereas minimum time was taken by test patch 3 due to their different sizes.

Lastly on the basis of qualitative and quantitative results, it can be inferred that the cognitive approach for detecting man-made objects from HRS images performs well.

CTA is one of the key contributions in the field cognitive psychology. When an analyst performs a given task, CTA is capable of generating descriptive information which is directly dependent on the performance of the analyst. CTA is a relevant source of information which is the outcome of the cognitive processes of an analyst. Some of the automated methods for building detection have limitations due to differences in size, shape, colour and density (which is relatively high in urban areas compared to rural areas) of buildings. Therefore, in this study a method is implemented which is capable of detecting buildings, irrespective of their shape and size, from HRS images. This method has a limitation that it is not capable of performing discrete separation between non-building and building regions which have similar spectral values. In future this cognitive approach for building detection needs to be tested using a larger geographical region and more number of testing images for improving the accuracy.

1. Jin, X. and Davis, C. H., Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. *EURASIP J. Adv. Signal Process.*, 2005, **14**, 1–11.

2. Xiao, Y., Lim, S. K., Tan, T. S. and Tay, S. C., Feature extraction using very high resolution satellite imagery. In Proceedings of Geoscience and Remote Sensing Symposium IGARSS'04, Anchorage, Alaska, 2004.
3. Baltsavias, E. P., Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. *ISPRS J. Photogramm. Remote Sensing*, 2004, **58**, 129–151.
4. Mayer, H., Automatic object extraction from aerial imagery – a survey focusing on buildings. *Comput. Vision Image Understand.*, 1999, **74**, 138–149.
5. Unsalan, C. and Boyer, K. L., A system to detect houses and residential street networks in multispectral satellite images. *Comput. Vision Image Understand.*, 2005, **98**, 423–461.
6. Brenner, C., Building reconstruction from images and laser scanning. *Int. J. Appl. Earth Obs. Geoinf.*, 2005, **6**, 187–198.
7. Haala, N. and Kada, M., An update on automatic 3D building reconstruction. *ISPRS J. Photogramm. Remote Sensing*, 2010, **65**, 570–580.
8. Ozgen, C., Approaches for automatic urban building extraction and updating from high resolution satellite imagery. Doctoral dissertation, Middle East Technical University, Ankara, Turkey, 2009.
9. Huertas, A. and Nevatia, R., Detecting buildings in aerial images. *Comput. Vision, Graphics, Image Process.*, 1988, **41**, 131–152.
10. Irvin, R. B. and McKeown, D. M., Methods for exploiting the relationship between buildings and their shadows in aerial imagery. In Proceeding of OE/LASE, International Society for Optics and Photonics, Los Angeles, USA, 1989.
11. Liow, Y. T. and Pavlidis, T., Use of shadows for extracting buildings in aerial images. *Comput. Vision, Graphics, Image Process.*, 1990, **49**, 242–277.
12. McGlone, J. C. and Shufelt, J. A., Projective and object space geometry for monocular building extraction. In Proceedings of CVPR, Computer Vision and Pattern Recognition, IEEE Computer Society Conference, Seattle, WA, 1994.
13. Lin, C. and Nevatia, R., Building detection and description from a single intensity image. *Comput. Vision Image Understand.*, 1998, **72**, 101–121.
14. Lee, D. S., Shan, J. and Bethel, J. S., Class-guided building extraction from Ikonos imagery. *Photogramm. Eng. Remote Sensing*, 2003, **69**, 143–150.
15. Inglada, J., Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. *ISPRS J. Photogramm. Remote Sensing*, 2007, **62**, 236–248.
16. Sirmacek, B. and Unsalan, C., Urban-area and building detection using SIFT keypoints and graph theory. *IEEE Trans. Geoscience Remote Sensing*, 2009, **47**, 1156–1167.
17. Ok, A. O., Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts. *ISPRS J. Photogramm. Remote Sensing*, 2013, **86**, 21–40.
18. Matikainen, L., Kaartinen, H. and Hyyppa, J., Classification tree based building detection from laser scanner and aerial image data. In Proceedings of International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2007.
19. Kumar, S. and Hebert, M., Man-made structure detection in natural images using a causal multiscale random field. In Proceedings of IEEE Computer Society Conference, Computer Vision and Pattern Recognition, 2003.
20. Korc, F. and Forstner, W., Interpreting terrestrial images of urban scenes using discriminative random fields. In Proceedings of Congress of the International Society for Photogrammetry and Remote Sensing, Beijing, China, 2008.
21. Srivastava, H. S., Patel, P., Sharma, Y. and Navalgund, R. R., Comparative evaluation of potential of optical and SAR data for the detection of human settlements using digital classification. *Int. J. Geoinformat.*, 2006, **2**, 21–28.
22. Zhong, P. and Wang, R., Object detection based on combination of conditional random field and markov random field. In Proceedings of IEEE International Conference on Pattern Recognition, Hong Kong, 2006.
23. Clark, R. E., Feldon, D., van Merrienboer, J. J., Yates, K. and Early, S., Cognitive task analysis. In *Handbook of Research on Educational Communications and Technology*, 2008, **3**, 577–593.
24. Burge, J. E., Knowledge elicitation for design task sequencing knowledge, Doctoral dissertation, Worcester Polytechnic Institute, Massachusetts, USA, 1998.
25. Goodarzi Mehr, S., Ahadnejad, V., Abbaspour, R. A. and Hamzeh, M., Using the mixture-tuned matched filtering method for lithological mapping with Landsat TM5 images. *Int. J. Remote Sensing*, 2013, **34**, 8803–8816.
26. Boardman, J. W. and Kruse, F. A., Automated spectral analysis: a geological example using AVIRIS data, north Grapevine Mountains, Nevada. In Proceedings of the Thematic Conference on Geologic Remote Sensing, Environmental Research Institute of Michigan, 1994.
27. Wang, J. and Zhang, J., A new idea in the clay alteration information extraction in vegetation coverage. *Commun. Inf. Sci. Manage. Eng.*, 2012, **2**, 21–26.
28. Green, A. A., Berman, M., Switzer, P. and Craig, M. D., A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. Geosci. Remote Sensing*, 1988, **26**, 65–74.
29. Harris, A. and Bryant, R. G., A multi-scale remote sensing approach for monitoring northern peatland hydrology: present possibilities and future challenges. *J. Environ. Manage.*, 2009, **90**, 2178–2188.
30. Williams, A. P. and Hunt, E. R., Estimation of leafy spurge cover from hyperspectral imagery using mixture tuned matched filtering. *Remote Sensing Environ.*, 2002, **82**, 446–456.
31. Chang, C. I. and Heinz, D. C., Constrained subpixel target detection for remotely sensed imagery. *IEEE Trans. Geosci. Remote Sensing*, 2000, **38**, 1144–1159.
32. Manno-Kovacs, A. and Ok, A. O., Building detection from monocular VHR images by integrated urban area knowledge. *IEEE Geosci. Remote Sensing Lett.*, 2015, **12**, 2140–2144.
33. Aksoy, S., Yalniz, I. Z. and Tasdemir, K., Automatic detection and segmentation of orchards using very high resolution imagery. *IEEE Trans. Geosci. Remote Sensing*, 2012, **50**, 3117–3131.
34. Ok, A. O., Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts. *ISPRS J. Photogramm. Remote Sensing*, 2013, **86**, 21–40.
35. Ghaffarian, S., Automatic building detection based on supervised classification using high resolution Google Earth images. *Int. Arch. Photogramm, Remote Sensing Spatial Inf. Sci.*, 2014, **40**, 101.

ACKNOWLEDGEMENTS. We thank Dr Ali Ozgun Ok for providing the dataset and Ashu Sharma for help and discussions.

Received 13 July 2016; revised accepted 29 September 2016

doi: 10.18520/cs/v112/i05/1038-1044