



# 利用词袋模型检测建筑物顶面损毁区域

涂继辉<sup>1,2</sup> 眭海刚<sup>2</sup> 冯文卿<sup>2</sup> 孙开敏<sup>2</sup>

1 长江大学电子信息学院,湖北 荆州,434023

2 武汉大学测绘遥感信息工程国家重点实验室,湖北 武汉,430079

**摘要:**针对航空影像中已分割出的建筑物顶面,提出了一种利用视觉词袋模型检测建筑物顶面损毁区域的方法。该方法首先利用简单线性迭代聚类方法对建筑物顶面进行超像素分割,然后对超像素区域利用颜色和梯度方向直方图特征构建视觉词袋模型,最后使用支持向量机(support vector machine, SVM)对超像素区域中的损毁区域进行检测。实验结果表明,该方法能有效判定建筑物顶面损毁区域,对提高建筑物整体损毁检测精度具有重要意义。

**关键词:**建筑物顶面损毁检测;视觉词袋模型;超像素分割;简单线性迭代聚类(SLIC);SVM

中图法分类号:P237; TP753 文献标志码:A

航空遥感具有实时性强、机动快速、影像分辨率高、经济便捷等特点,且能够在高危地区作业,非常适合于各种自然灾害的检测与应急救援,因此如何利用航空影像对建筑物进行全方位的高精度定量损毁评估是目前研究的热点。建筑物的顶面是在航空影像中观测到的建筑物最显著的特征,其损毁检测研究对提高建筑物整体损毁精度评估具有重要意义<sup>[1-2]</sup>。

国内外学者利用航空影像对建筑物顶面损毁检测展开一些研究工作,主要方法体现在两个方面:(1)基于单时相的建筑物顶面损毁检测<sup>[3]</sup>。由于灾后航空影像易于获取,因此这类方法更加符合实际生产需求。此类方法主要利用建筑物顶面的面积信息进行损毁检测,但是灾后很多建筑物顶面损毁的特征是顶面发生了破损或者裂缝,而整体面积未有变化,因此该方法在高精度减灾评估中具有一定的局限性。(2)基于多时相的建筑物顶面损毁检测<sup>[4]</sup>。此类方法主要利用顶面面积变化检测的方法进行建筑物顶面损毁检测,但灾前的航空影像一般很难获取,特别是临近受灾前时间段的倾斜影像更加难以获取,而且灾前灾后多时相的影像如何进行高精度配准也是目前的难点。

基于以上研究,根据建筑物顶面纹理特征,本

文提出了一种基于视觉词袋模型的航空影像建筑物顶面损毁区域检测方法。该方法利用超像素分割使得损毁检测从传统的以像素为单位变为有语义信息的像素块,有效克服了语义鸿沟的问题,同时利用视觉词袋模型对超像素分割像素块赋予语义信息,有利于支持向量机(support vector machine, SVM)分类器的训练和识别。利用本文方法对北川地震遗址的航空影像中的大量建筑物顶面进行了损毁检测,检测效果较好,为提高灾害中建筑物整体损毁精细评估提供了重要手段。

## 1 基于词袋模型的建筑物损毁检测方法

在航空影像中建筑物顶面已经提取的前提条件下,本文方法的整体流程如图1所示,分为3步:(1)对建筑物顶面进行超像素分割;(2)提取分割区域的特征,建立视觉词袋模型;(3)对分割区域进行SVM分类,获取损毁区域。

### 1.1 超像素分割

当建筑物顶面完好时,它的局部区域特征是颜色和纹理结构具有一致,而当建筑物顶面有损毁时,其局部区域特征是颜色出现了非一致性,且纹理结构混乱。因此,利用超像素的方法对建筑

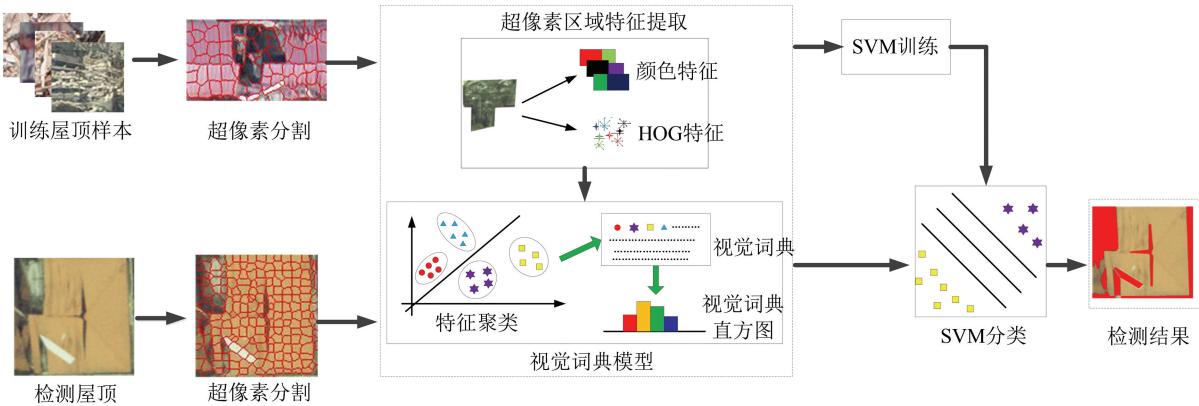


图1 研究方法流程图

Fig.1 Flowchart of the Proposed Method

物顶面进行分割,可以将建筑物顶面的损毁和非损毁区域像素进行分类,有助于下一步基于局部区域的语义分析。目前,超像素分割的方法很多<sup>[5]</sup>,典型的超像素方法有简单线性迭代聚类(simple linear iterative clustering, SLIC)<sup>[6]</sup>、Watersheds<sup>[7]</sup>、NormalizedCuts<sup>[8]</sup>、MeanShift<sup>[9]</sup>、QuickShift<sup>[10]</sup>和TurboPixel<sup>[11]</sup>等,其中SLIC是一种基于k-means思想的超像素分割算法,具有使用简单、处理速度快和产生的超像素最规整的特点。所以,本文采用SLIC算法进行超像素分割,又由于当前建筑屋顶面多为彩色,因此本文采

用的SLIC利用像素的颜色相似度和平面空间信息对像素进行聚类。具体算法描述如下。

1) 初始化聚类中心。对于有N个像素的图像,如需要产生的超像素区域的数目为K,则每个超像素的平均面积为N/K,每个种子点的平均距离近似为S=√(N/K),在图像内以该距离均匀分配聚类中心。

2) 移动聚类中心。为了避免种子点处于图像边缘位置而对聚类造成干扰,种子点被移动到其3×3区域内梯度值最小的位置。

3) 计算像素点(x<sub>1</sub>,y<sub>1</sub>)和(x<sub>2</sub>,y<sub>2</sub>)的相似度d。

$$d_c = \sqrt{(l(x_1, y_1) - l(x_2, y_2))^2 + (a(x_1, y_1) - a(x_2, y_2))^2 + (b(x_1, y_1) - b(x_2, y_2))^2} \quad (1)$$

$$d_s = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2)$$

$$d((x_1, y_1), (x_2, y_2)) = \sqrt{d_c^2((x_1, y_1), (x_2, y_2)) + m^2(\frac{d_s^2((x_1, y_1), (x_2, y_2))}{S})} \quad (3)$$

式中,l<sub>i</sub>、a<sub>i</sub>、b<sub>i</sub>为像素点i在CIE Lab空间中L、a、b通道的数值;(x,y)为像素点坐标值;m为平衡参数。

4)按最小相似度对图像的像素点进行聚类,不断迭代,直到新种子点和原来的种子点之间的距离足够小时迭代终止。

5)把面积较小的区域与其距离最近的区域合并,以保证区域的连通性。

为了提高算法的运行速度,对每个种子点聚类时,只在以种子点为中心的2S×2S区域内搜索像素点。

## 1.2 视觉词袋模型的建立

单词包模型(bag-of-word, BoW)最早应用于文本检索领域,是一种利用基于语义的方法对自然语言进行检索和处理的算法<sup>[12]</sup>,许多学者后来将该方法引入到计算机视觉和多媒体领域,成为一种有效的基于语义特征提取和描述的算法,

在图像分类、目标识别和图像检索中得到了广泛应用<sup>[13]</sup>。本文利用视觉词袋模型对超像素区域进行语义特征的描述,即在建筑物顶面超像素区域内提取特征构建损毁和非损毁的语义词袋模型,为下一步的SVM分类训练提供正负样本。视觉词袋模型的实现流程分为特征的描述、视觉词典的生成、视觉词汇的构建和SVM训练4个部分,具体流程如下。

1) 特征描述。根据建筑物顶面损毁特征,本文利用颜色特征和(histogram of oriented gradient, HOG)纹理特征对分割区域进行特征描述。由于HSV颜色空间较RGB颜色空间更加符合人眼视觉,对亮度更加敏感,有利于颜色特征的区分,所以本文利用H、S和V对分割区域进行非均匀量化,即将色调量H分为8个级别,饱和度S和亮度V各分为3个级别,然后将3个颜色分量合成一个值L:L=9H+3S+V(L∈[0,71])。

由于超像素分割区域在形态上有一定的差异,本文根据超像素区域的长轴长度以及质心位置来确定一个矩形区域,用于计算 HOG 的特征。根据本文实验统计,超像素区域长轴为 30 像素,所以提取以质心为中心  $30 \times 30$  像素的区域用于计算 HOG 特征,HOG 的细胞大小为  $6 \times 6$  像素,梯度方向抽取 9 个方向块,所以可以产生 225 个 HOG 的特征。为了防止较小的特征分量在计算中被忽略,对两类不同的特征量进行归一化处理,使得所有分量在特征描述时具有相同的权重。

2) 视觉词典生成。设上一步得到的每个超像素分割区域的特征向量为  $\mathbf{F} = [f_1 \ f_2 \cdots f_N]$ , 其中  $N$  表示每个区域的特征数量。对特征向量  $\mathbf{F}$  进行  $K$ -均值聚类,生成一个  $K$  维向量的视觉词典  $\mathbf{W} = [w_1 \ w_2 \cdots w_K]$ 。

3) 基于视觉词典对图像的表达。在视觉词袋模型中,图像是由不同频率的视觉词典组成的,本文对超像素分割区域进行视觉词典的直方图统计,得到统计向量  $\mathbf{N} = [n_1 \ n_2 \ n_3 \cdots n_i \cdots n_K]$ 。由于不同的视觉单词对图像的表达作用不同,直接使用最初的视觉词典对图像进行表达会对后续的识别造成不佳的影响。为了区分不同的视觉单词对图像表达的贡献,本文利用了文本信息检索中的 tf-idf<sup>[14]</sup>(词频-反转文件频率)的加权方法来描述视觉词典中单词的权重。视觉词典中单词的权重向量为  $\mathbf{T} = [t_1 \ t_2 \ t_3 \cdots t_i \cdots t_K]$ , 权重  $t_i$  的表达式为:

$$t_i = \frac{n_{id}}{n_d} \lg \frac{N}{n_i} \quad (4)$$

其中,  $n_{id}$  表示第  $i$  个视觉单词在分割区域中出现的频率;  $n_d$  表示所有视觉单词在分割区域中出现的总频率;  $N$  表示分割区域的总数;  $n_i$  表示第  $i$  个视觉单词在所有分割区域中出现过的次数。那么损毁和非损毁区域的加权特征表达为:

$$f_i = n_i \times t_i \quad (5)$$

4) 基于 SVM 的分类训练。利用式(5)的视觉词袋模型分别对建筑物顶面的损毁区域和非损毁区域训练 SVM 分类器,得到损毁区域的分类器。

### 1.3 建筑物顶面损毁判定

基于视觉词袋模型的建筑物顶面损毁判定流程如下所示,分类器采用 SVM, 图像分割采用 SLIC 的超像素方法。

1) 利用 SLIC 的方法对建筑物顶面进行分割,得到不同的分割区域  $R = \{R_1, R_2, R_3 \dots R_n\}$ 。

2) 取出某一个分割区域  $R_i$ , 提取  $R_i$  区域的颜色特征和 HOG 特征, 将特征归一化。

3) 统计分割区域  $R_i$  的视觉词典直方图, 根据式(5)得到特征表达  $f_i$ 。

4) 将特征表达  $f_i$  放入经过训练的 SVM 分类器中, 判断区域  $R_i$  是否为损毁区域。

## 2 建筑物顶面损毁检测实验与分析

本文利用武汉大学测绘遥感信息工程国家重点实验室于 2014 年 4 月在北川地震遗址进行灾害考察时航拍的遥感影像进行实验, 图幅大小为  $8260 \times 6166$  像素, 影像分辨率为 0.2 m。本文主要利用了航空影像中的下视影像, 用北川地震遗址数据中 50 个建筑物顶面作为训练样本集、50 个建筑物顶面作为测试数据集。为了定量评价建筑物顶面损毁检测的性能, 本文采用准确率  $P$  和召回率  $R$ , 分别定义为:

$$P = \frac{T_p}{T_p + F_p} \quad (6)$$

$$R = \frac{T_p}{T_p + F_n} \quad (7)$$

其中,  $T_p$  表示正确检测到的损毁区域;  $F_p$  表示非损毁却被检测为损毁的区域;  $F_n$  表示损毁被检测为非损毁的区域。由于词袋模型中特征描述的选取是最后目标识别的关键, 为了验证本文算法中特征选取的有效性, 本文利用 Surf 特征<sup>[15]</sup>、Gabor 特征<sup>[16]</sup>和本文提出的方法分别进行比较。

### 2.1 实验结果

本文实验中,首先选用航空影像中的 50 个建筑物顶面作为样本训练数据, 经过 SLIC 分割后, 通过人工分类方式共获得 8921 个未损毁区域和 4367 个损毁区域, 如图 2(a) 所示为部分损毁区域的样本, 图 2(b) 为部分未损毁区域的样本; 然后利用 1.2 节的流程对样本进行特征提取、视觉词典生成以及 SVM 分类器训练; 最后选取

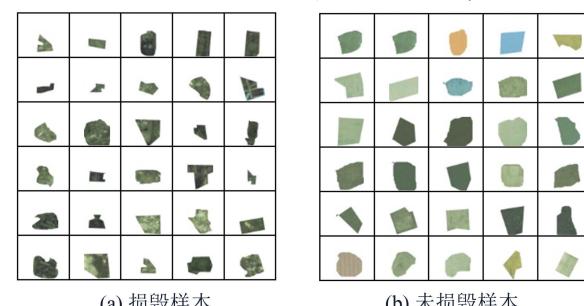


图 2 训练样本

Fig.2 Training Samples

50个建筑物顶面作为测试数据,利用1.3节的流程对测试数据进行损毁检测,列出了部分如图3(e)所示的检测结果。本文还分别选用了Surf和Gabor两种特征作为词袋模型的特征描述,检测结果分别如图3(c)、3(d)所示。表1给出了Surf、Gabor和本文算法对50个建筑物顶面检测后的平均检测精度。本文算法之所以能取得较好的效果是由于建筑物顶面未损毁时候纹理较为均匀且方向一致性较强;而HOG的算法主要通过计算和统计图像局部区域的梯度方向直方图来构成特征,因此对局部区域的方向性变化具有一定的敏感度。同时,由于建筑物顶面损毁区域的颜色较其他区域亮度更暗,因此颜色特征的选取也

有助于对损毁区域的检测。从图3中可以发现,本文算法对部分建筑物顶面小的白色损毁区域未能检测出来,主要是由这部分区域纹理均匀且颜色较亮,和未损毁区域特征非常相似,因此本文算法未能正确检测。另外两种特征未能取得较好的检测结果是因为损毁区域的亮度较暗,Surf特征点难以提取,Gabor作为特征对光照变化不敏感,且能容忍一定程度的图像旋转和变形。

表1 损毁检测精度评价

Tab.1 Performance of Variant Methods

方法	准确率	召回率
Surf 特征	0.73	0.71
Gabor 特征	0.81	0.79
HOG+Color 特征	0.92	0.88

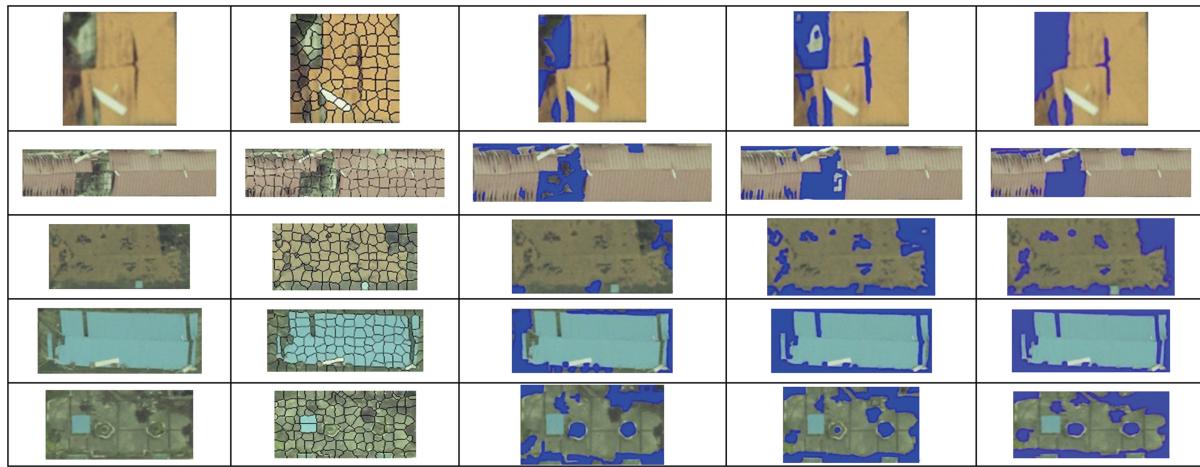


图3 损毁检测结果

Fig.3 Results of Damage Detection by Different Methods

## 2.2 参数选择

在视觉词袋模型中,视觉词典的长度K是一个非常关键的参数,它的选取决定了最后检测效果。如图4(a)所示,当K的范围在20到45之间

时,检测平均精度变化不大,当K取45时,精度达到最优。图4(b)是选择K=45作为视觉单词的长度时得到的检测精度。

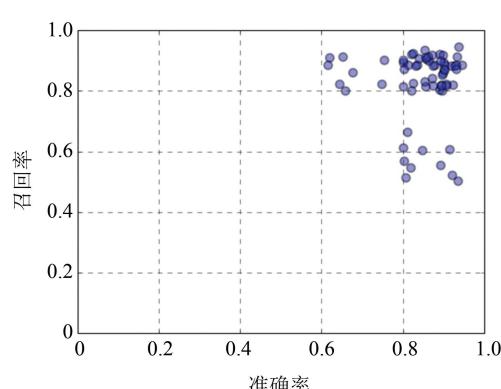
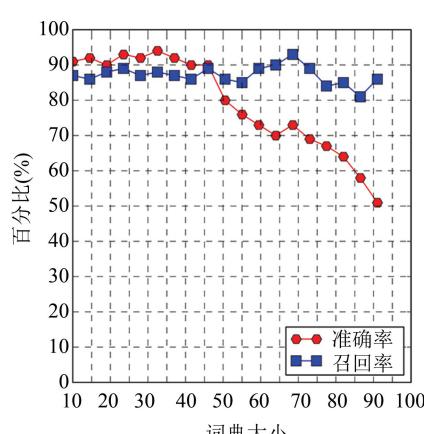


图4 视觉词典长度与精度关系图

Fig.4 Relationship Between Precision, Recall and Codebook Length

### 3 结语

建筑物顶面损毁检测对于建筑物整体损毁评估具有重要意义,本文提出了一种利用视觉词袋模型检测航空影像中建筑物顶面损毁区域的方法。该方法首先利用 SLIC 方法对建筑物顶面进行超像素分割,使得损毁检测变为基于像素块的语义信息检测;然后在超像素分割区域构建视觉词袋模型,将像素块赋予语义信息;最后利用 SVM 对视觉词袋模型进行分类训练和损毁识别,从而达到对建筑物顶面损毁的判定。实验结果表明,该方法能够有效检测建筑物顶面损毁区域,为提高建筑物整体损毁评估的精度和自动化程度提供了解决方法。

### 参 考 文 献

- [1] Dong Laigen, Shan Jie. A Comprehensive Review of Earthquake-Induced Building Damage Detection with Remote Sensing Techniques[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2013, 84: 85-99
- [2] Li Deren, Liu Likun, Shao Zhenfeng. An Integration of Aerial Oblique Photogrammetry and Mobile Mapping System for Urban Geographical Conditions Monitoring[J]. *Geomatics and Information Science of Wuhan University*, 2015, 40(4): 427-435(李德仁, 刘立坤, 邵振峰.集成倾斜航空摄影测量和地面移动测量技术的城市环境监测[J].武汉大学学报·信息科学版, 2015, 40(4): 427-435)
- [3] Li Manchun, Cheng Liang, Gong Jianya, et al. Post-earthquake Assessment of Building Damage Degree Using LIDAR data and Imagery[J]. *Science in China Series E: Technological Sciences*, 2008, 52: 133-143
- [4] Sui Haigang, Tu Jihui, Song Zhina, et al. A Novel 3D Building Damage Detection Method Using Multiple Overlapping UAV Images[C]. ISPRS Technical Commission VII Symposium, Istanbul, Turkey, 2014
- [5] Wang Chunyao, Chen Junzhou, Li Wei. Review on Superpixel Segmentation Algorithms[J]. *Application Research of Computers*, 2014, 31(1): 6-12(王春瑶, 陈俊周, 李炜. 超像素分割算法研究综述[J]. 计算机应用研究, 2014, 31(1): 6-12)
- [6] Achanta R, Shaji A, Smith K, et al. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 2274-2282
- [7] Vincent L, Soille P. Watersheds in Digital Spaces: an Efficient Algorithm Based on Immersion Simulations[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, 13(6): 583-598
- [8] Shi Jianbo, Malik J. Normalized Cuts and Image Segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 888-905
- [9] Comaniciu D, Meer P. Mean Shift: A Robust Approach Toward Feature Analysis[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(5): 603-619
- [10] Vedaldi A, Soatto S. Quick Shift and Kernel Methods for Mode Seeking[C]. International Conference on Computer Vision, Marseille, France, 2008
- [11] Levinstein A, Stere A, Kutulakos K N, et al. Turbopixels: Fast Superpixels Using Geometric Flows [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(12): 2290-2297
- [12] Lodhi H, Saunders C, Shawe-Taylor J, et al. Text Classification Using String Kernels[J]. *The Journal of Machine Learning Research*, 2002, 41(2): 419-444
- [13] Sivic J, Zisserman A. Video Google: A Text Retrieval Approach to Object Matching in Videos[C]. International Conference on Computer Vision, Nice, France, 2003
- [14] Jégou H, Douze M, Schmid C. Packing Bag-of-Features[C]. International Conference on Computer Vision, Kyoto, Japan, 2009
- [15] Bay H, Tuytelaars T, Gool L V. Surf: Speeded Up Robust Features[C]. International Conference on Computer Vision, Marseille, France, 2008
- [16] Liu C, Wechsler H. Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition[C]. International Conference on Image Processing, Marseille, France, 2002

# Detection of Damaged Areas Based on Visual Bag-of-Words Model from Aerial Remote Sensing Images

TU Jihui<sup>1,2</sup> SUI Haigang<sup>2</sup> FENG Wenqing<sup>2</sup> SUN Kaimin<sup>2</sup>

1 Electronics and Information School, Yangtze University, Jingzhou 434023, China

2 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing,  
Wuhan University, Wuhan 430079, China

**Abstract:** An approach for damaged rooftops areas detection is proposed based on visual bag-of-words model. First, the building rooftop is segmented into different superpixel areas using simple linear iterative clustering(SLIC) method, then features of color and histograms of oriented gradients are extracted from each superpixel area and the visual bag-of-words (BoW) model is employed to build the semantic feature vectors of damaged and non-damaged area. Finally, damaged and non-damaged parts of rooftop superpixel areas are discriminated using SVM. Experimental results show that the proposed method can be feasible and effective for detection of damaged rooftop areas, which is an important significance for improving the accuracy of overall building damaged detection.

**Key words:** detection of damaged rooftop areas; visual bag-of-words model; superpixel segmentation; SLIC; SVM

**First author:** TU Jihui, PhD, associate professor, specializes in remote sensing image processing and information extraction, and 3D reconstruction. E-mail: green666@126.com

**Corresponding author:** SUI Haigang, PhD, professor. E-mail: haigang\_sui@263.net

**Foundation support:** The National Key Research and Development Program of China, No.2016YFB0502603; the National Natural Science Foundation of China, No. 41471354.