

RAGWEED DETECTION BASED ON SURF FEATURES

Ádám Schiffer, Zoltán Sari, Péter Müller, Ildikó Jancsár, Géza Varady, Zsolt Ercsey

Original scientific paper

The paper describes a parameter study corresponding to automatic detection of ragweed based on SURF features. The basic idea behind the method is to build a feature database from very simple ragweed samples containing characteristic features of the leaves of the plant, and compare the feature database to features extracted from natural images which contain or lack ragweed. The results of the study clearly show that the approach is promising and has value as a standalone method, or as a potential training basis for a classification expert system.

Keywords: feature detection; parameter study; ragweed

Otkrivanje ambrozije na osnovu SURF značajki

Izvorni znanstveni članak

U radu se opisuje analiza parametara koji odgovaraju automatskom otkrivanju ambrozije na temelju SURF značajki. Osnovna ideja je sastaviti bazu podataka značajki iz vrlo jednostavnih uzoraka s karakterističnim obilježjima listova biljke i usporediti bazu podataka tih značajki sa značajkama dobivenim iz običnih slika sa ili bez ambrozije. Rezultati istraživanja jasno pokazuju da takav pristup ima svrhu te vrijednost kao posebna metoda ili kao moguća baza za učenje ekspertnog sustava klasifikacije.

Ključneriječi: ambrozija (*limundžika*); analiza parametara; otkrivanje značajki

1 Introduction

Common ragweed (*Ambrosia artemisiifolia* L) is considered a troublesome weed among European countries, especially in Hungary. Although it can only be found in Hungary since the First World War, now this invasive species is the most widely spread weed causing significant economic damage from the agricultural point of view. Besides that, it is of importance from human biological point of view: its pollen is recognized as a significant cause of allergic symptoms.

In Hungary, the 350 weed species are regularly monitored and surveyed. Since the first national weed survey in 1950, ragweed jumped from the 21st rank to number 1 position regarding its penetration and coverage. It can be found on 5 out of 7 million cultivated hectares, while 5,3% of the arable land is estimated to be infected; as a result 120-200 billion HUF is an average loss of the Hungarian economy [10]. Besides its economic importance, according to statistics of the National Institute of Environmental Health it is a highly important issue from the human biological point of view also, namely approximately 20% of the overall population suffer from allergic disease because of the ragweed pollens [11].

Besides Hungary, which seems to be the most exposed country in Europe, it is also important elsewhere. For example in Switzerland up to 12% of the population suffers from allergies to the ragweed pollen [7], while 20% of the population in the region of Rhone in France suffers from ragweed pollen allergy [5]. On the other hand, although it is considered to be a native species, 10% of the United States population is also estimated as ragweed sensitive [8].

Even though its importance, the automated detection and recognition of the ragweed is currently still an open issue.

There are methods available to determine larger areas infected by ragweed, basically with the help of satellite

pictures. Auda et al. first determined that the spectral view of the common ragweed may be separated from the gramineae at 660 nm in the red and infrared (>750 nm) range [4]. Unfortunately, usually there are no multispectral pictures with such information, therefore they have extended their research and suggested supervised classification of satellite picture pixels supported by on-site surveys [6]. Unfortunately the method is not robust thus cannot be generally used.

Other available methods focus on various pollen information. In this regard, in Hungary a mapping method in the visualization of the pollen load, namely the Ragweed Pollen Alarm System (PPRR) was developed and it is run by the Department of Aerobiological Monitoring, National Institute of Environmental Health. Counting the ragweed pollen is still done manually but recent advances correspond to the automated detection in the future [9]. As a first step, they developed a pollen recognition algorithm based on colour segmentation after colouring the pollens with magenta.

2 Ragweed detection

The main objective of the research presented in this paper was the development and investigation of a methodology for processing digital images taken from the nature fields in order to determine the presence of ragweed. The proposed method has the potential of determining if a certain field segment requires herbicides, thus it can help to make a further step towards precision agriculture. However, it is very difficult to process pure image inputs with high variability to obtain the suggested ragweed distribution, since there are many similar plants growing together and the colour variation of the crops, weeds and soils is much greater than our database.

The digital images used in the system development were collected in several fields around Pécs and in laboratories over the period of May to August of 2014. The most common images were *Ambrosia artemisiifolia*

L. (ragweed), *Artemisia vulgaris*, *Anthriscuscerefolium*, *Reseda lutea*, *Daucuscarota*, *Rudbeckiahirta*.

The aim of the ambrosia leaves feature detection is to process and examine the ambrosia images to extract characteristic features corresponding to leaf shapes. In this manner we are able to detect objects similar to the ambrosia leaves based on their features in different images. This detection is also possible when the image contains the object with different transformations like scaling and rotation, or when parts of the objects are occluded.

2.1 The SURF algorithm

The Scale-Invariant Feature Transform (SIFT) is a modern approach to detect feature points [1]. The SURF algorithm is based on similar principles but it utilizes a different scheme and provides faster and possibly better results [2]. To detect the feature points on the ambrosia image, SIFT uses cascading filters called Difference of Gaussians (DoG). It is calculated on progressively downscaled images. Both SIFT and SURF techniques examine the image at different scales called *scale space* using the Gaussian kernels. The scale space is divided into levels and octaves. The octave corresponds to doubling the deviation σ of the Gaussian function and the octave is divided into uniformly spaced levels. One potential interpretation is to use large amounts of Gaussian blurring that averages out nearly all useful information in images. Both approaches build pyramids with different levels of octaves. The relation between levels and octaves is illustrated in Fig. 1.

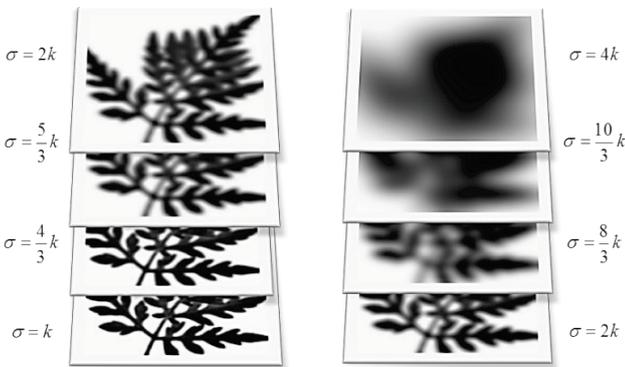


Figure 1 The relation between levels and octaves shown in an ambrosia image

The points of interest are those which have the extreme among 8 neighbours in the actual level and 2×9 neighbours in the level below and above.

2.1.2 Calculating Hessian points of interest

Both approaches are based on blob detector to find points of interest. The determinant of a Hessian matrix is an expression of the local change around the given point.

$$H(z, \sigma) = \begin{bmatrix} C_{XX}(z, \sigma) & C_{XY}(z, \sigma) \\ C_{XY}(z, \sigma) & C_{YY}(z, \sigma) \end{bmatrix}, \tag{1}$$

where

$$\begin{aligned} C_{XX}(z, \sigma) &= A_g * \frac{\partial^2}{\partial x^2} G(\sigma, x, y), \\ C_{YY}(z, \sigma) &= A_g * \frac{\partial^2}{\partial y^2} G(\sigma, x, y), \\ C_{XY}(z, \sigma) &= A_g * \frac{\partial^2}{\partial xy^2} G(\sigma, x, y). \end{aligned} \tag{2}$$

$C_{XX}(z, \sigma)$ is the convolution of the image A_g with the second order derivate of the Gaussian function $G(\sigma)$, the symbol $*$ denotes convolution. Since convolution has very high computational cost, it is approximated and speeded-up using integral images and approximated kernels. Fig. 2 shows the convolved images applying deviation $\sigma = 6$.

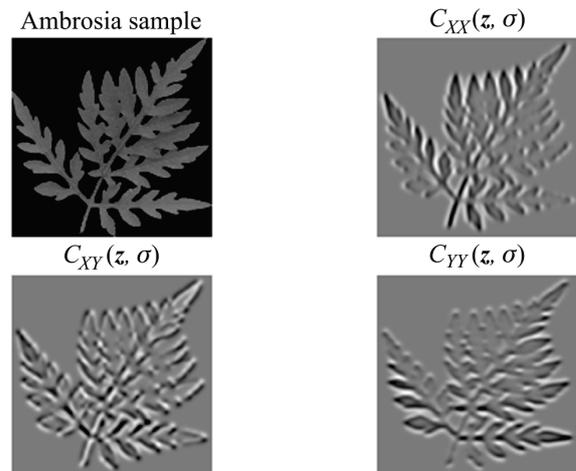


Figure 2 Original and convolved images with the second derivate of the Gaussian function $G(6)$

The SURF algorithm approximates the second order Gaussian kernels with rectangular boxes to speed up the process with the coarser kernels. In this way it is possible to calculate the approximated convolution effectively utilizing the integral image. To detect features across the scale space, several levels and octaves must be analysed. The SIFT algorithm scales the image down for each octave and uses progressively larger Gaussian kernels. The local maxima of the filter response image can be found in the regions where both $C_{XX}(z, \sigma)$ and $C_{YY}(z, \sigma)$ are strictly positive and $C_{XY}(z, \sigma)$ is strictly negative. It can occur in the image where large intensity gradient variations in multiple directions are found. Visually these are different sized, blob-like structures corresponding to corners of the ambrosia leaves [1, 2].

2.1.3 Descriptors

The role of the descriptors is to produce a unique description of a feature, and calculated from the area surrounding a point of interest. All of SIFT and SURF algorithms have a scheme to determine the orientation to achieve the rotational invariance. The SURF method makes the distinction between bright blobs on dark background versus dark blobs on bright background. This property is represented by the sign of the Laplacian as the trace of the Hessian matrix:

$$\nabla^2 L = C_{XX}(z, \sigma) + C_{YY}(z, \sigma). \tag{3}$$

It is necessary to compare the full descriptor vectors with the same sign. It decreases the computational time as well. In the applied algorithm the ambrosia grayscale pattern and its negative are used together to build the representative pattern feature database. To illustrate the method Fig. 3 shows the ten strongest feature points found in different scale levels (3, 6, 9, 12). It can be observed that to find the ambrosia leaves the feature points must be examined at low scale levels.

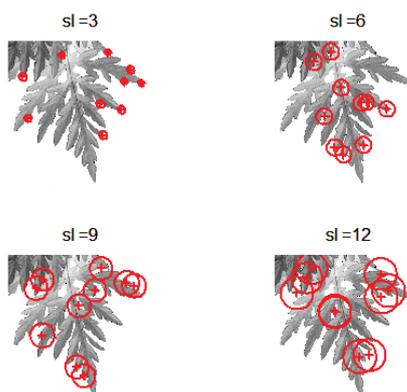


Figure 3 Ten significant feature points found at different scale levels

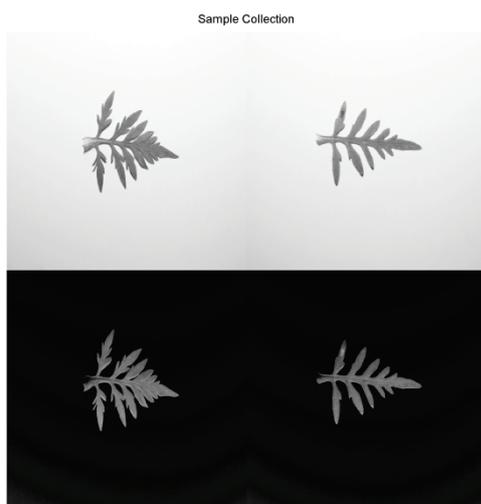


Figure 4 The reference images for the classification

3 Diagnostic test

The first step of the tests is to build the representative sample image database that contains two ragweed leaves and their negatives serving as the set of sample images. This reference set can be seen in Fig. 4. After this, the feature points must be detected in the pattern. Using feature vectors instead of images makes our ambrosia pattern recognition process more robust, and this approach also reduces the amount of the data that must be stored and analysed. All feature points from the sample image are stored in a matrix, from which with the tree of the k -nearest neighbour (k -NN) algorithm for the matching phase can be built. This tree stores information about the data used, distance metric, parameters, and the maximal number of data points in each leaf node. This

tree based representation allows fast searching for the nearest neighbours.

The examined probe image that contains the pattern to be identified must be analysed, so the feature points must be found as well. Four sample images from the total set of 28 can be seen in Fig. 5. It can be seen that the images are taken from the field directly. The ambrosia images (a) and (b) contain crops, weeds and soil, as well as the non-ambrosia images (c) and (d).

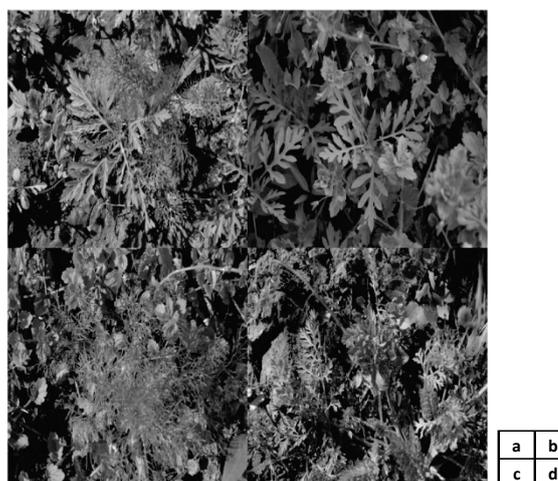


Figure 5 Four images from the sample database

The object recognition is performed by matching each feature to the database of features extracted from the sample set of images. The matching method for each feature is to find the nearest neighbour in the database of features from the sample set. The nearest neighbour is defined as having a minimum distance metric for the descriptor vector. The noise on the target images, like background clutter renders the algorithm more difficult. To select the good matches a further match thresholding (T_M) is introduced for the distance.

The developed algorithm was tested on an image database containing 14 ambrosia and 14 non-ambrosia weeds. From the results the sensitivity and specificity of the method can be calculated. First step is to calculate the true-positive (TP), false-positive (FP), true-negative (TN) and false-negative (FN) values:

Table 1 Diagnostic test for the ambrosia presence

	Ambrosia present	Ambrosia absent	
Ambrosia test positive	TP=13	FP=0	Total test positive 13
Ambrosia test negative	FN=1	TN=14	Total test negative 15
	Total ambrosia 14	Total non-ambrosia 14	Number of total images 28

The sensitivity is the probability of being test positive when the ambrosia disease is present:

$$Sensitivity = \frac{TP}{TP + FN}. \tag{4}$$

The specificity is the probability of being test negative when the ambrosia disease is absent:

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

Table 2 Calculation of specificity and sensitivity

	Ambrosia present	Ambrosia absent
Ambrosia test positive	13 (TP)	0 (FP)
Ambrosia test negative	1 (FN)	14 (TN)
	Sensitivity = 92%	Specificity = 100%

3.1 The method of parameter study

The fundamental concept of the method described is to build a feature database from simple but characteristic samples of ragweed based on the SURF algorithm outlined above. This database serves as the basis of comparison with actual probe images which can contain ragweed. The probe images have been selected randomly from a natural collection of images without any kind of pre-processing. The only constraint was the resolution of the images, namely all the probe images have the same resolution. Even though different screen resolutions make different image resolutions [3], however, this is neglected in the current research where the image process is based on the original images and therefore screen resolution problems have no impact. The sample database consists of two images and their negative versions (see Fig. 4). The image database consists of 28 images from which 14 have ragweed, and the other 14 do not have ragweed on it. For the sake of simplicity the images containing ragweed were numbered by 1-14.

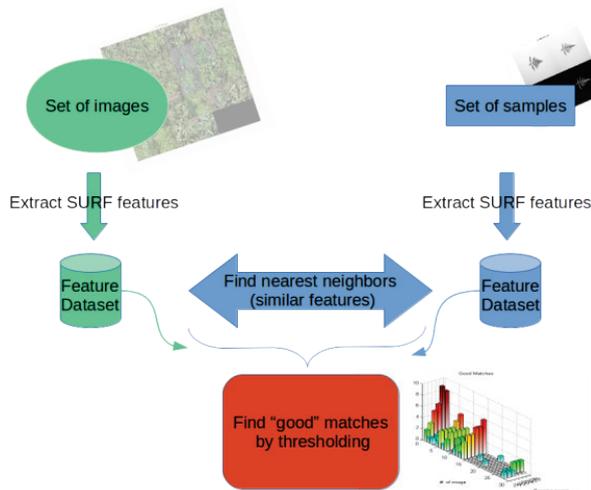


Figure 6 Method of feature comparison

3.2 Description of feature comparison method

In order to compare the features of probe images and sample database we followed the approach described here. First of all the feature vectors have to be extracted from sample images for all relevant parameter values in the parameter set, and similarly we need the features to be extracted from the probe images for all relevant parameter values as well. Then we have to search for the nearest

neighbours of sample features in the images database in order to decide how many ragweed-like features can be found in the probe images. The brief procedure can be followed in Fig. 6.

After finding the corresponding matching features a further match thresholding step should be considered in order to make the matches more robust. This thresholding, which is basically based on the distance metric of matches helps to find the best correspondences from the good candidates among all potential matches. The thresholding ensures that the algorithm only considers the best matches between the features of the sample set and probe images.

3.3 Parameters considered

During the parameter study we have investigated most relevant parameters of SURF feature detection, like octaves, scale levels, metric threshold of feature extraction, and the distance metric of the match thresholding described above. These parameters form a four dimensional parameter space and the parameter sweep carried out holds relevant information about the effect of the parameters and the overall viability of the proposed approach as well.

4 Results of the parameter study

This section contains numerical results of the parameter study. In each figure the running parameter was the index of scale level applied during feature extraction (indices running from 1-6 correspond to actual scale levels applied from 3-8). The results suggest, similarly to the examination of strongest features on the leaves, that the smaller scale levels contain more relevant features. Fig. 7 presents the result of a typical feature matching procedure which shows that the number of acceptable feature correspondences is considerably higher for the probe images containing ragweed (images indexed 1-14), but with the application of match thresholding the difference becomes more apparent.

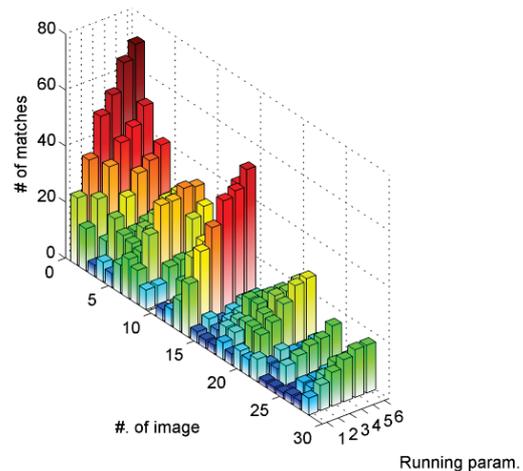


Figure 7 Result of feature matching over image index

Fig. 8 shows the result of the same procedure after match thresholding, showing that the probe images containing ragweed are practically separated by the much higher values of numbers of acceptable matches. It can

also be seen that there can be some false positives and false negatives as well for certain values of the running parameter (scale level), but the number of these false results is quite low leaving the method with reasonably high sensitivity and specificity as it is shown in Fig. 9 and Fig. 10 respectively.

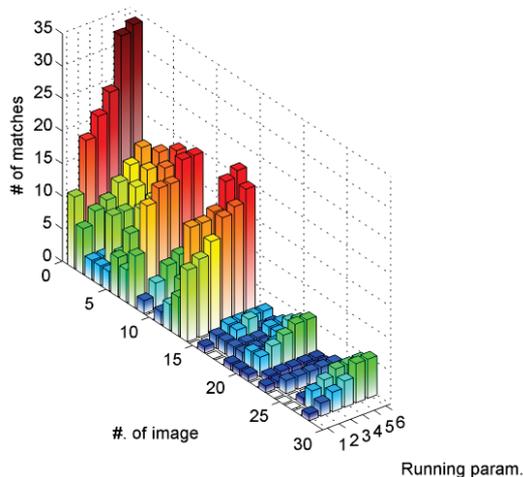


Figure 8 Result of feature matching over image index with match thresholding

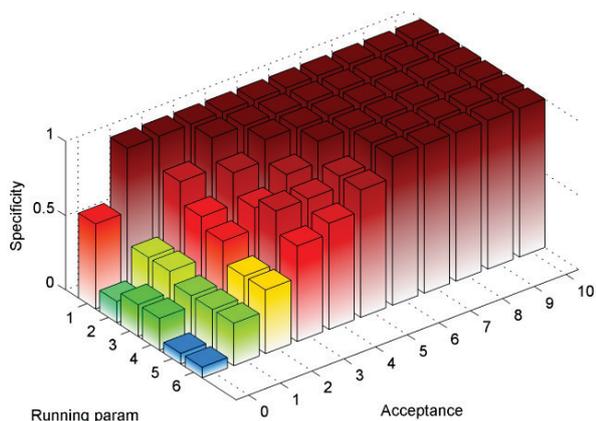


Figure 9 Specificity of the separation by acceptance level

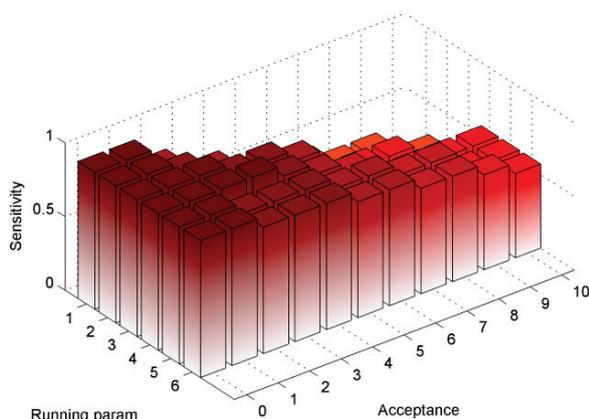


Figure 10 Sensitivity of the separation by acceptance level

The specificity and sensitivity of the method was calculated as described above. From the figures it can be seen that the method has specificity close to 100% and a corresponding sensitivity above 90%, which is acceptable. It can be also apparent from the figure that selecting a higher acceptance level will result in a drop of sensitivity,

which is normal, since increasing the level of acceptance will result in increase of false negative results.

The overall accuracy of the model can be also determined by the formula below

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

The parameter study carried out has shown that the overall accuracy of the method can reach a value well above 90% (see Fig. 11).

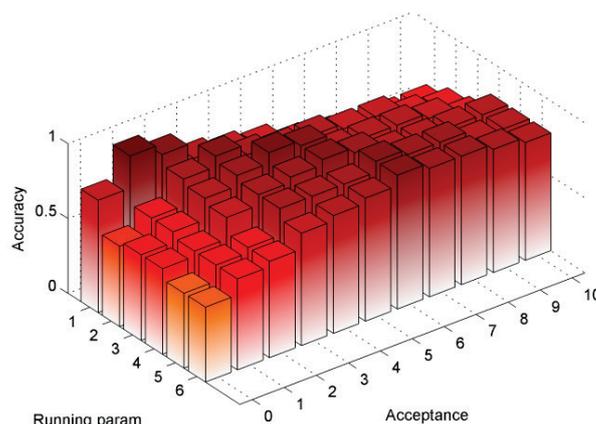


Figure 11 Accuracy of the separation by acceptance level

Separating the images based on the parameter values corresponding to the highest accuracy (96%) results in the following separation, considering a match positive if the number of matching features is more than one, as it is shown in Fig. 12.

The figure shows that for the selected set of parameters the method will result in one false negative and there are no false positives among the results, which corresponds to 100% specificity and 92% sensitivity, which are quite high values. Certainly these high values will not hold for all the parameter sets, they correspond to the highest accuracy in the parameter space.

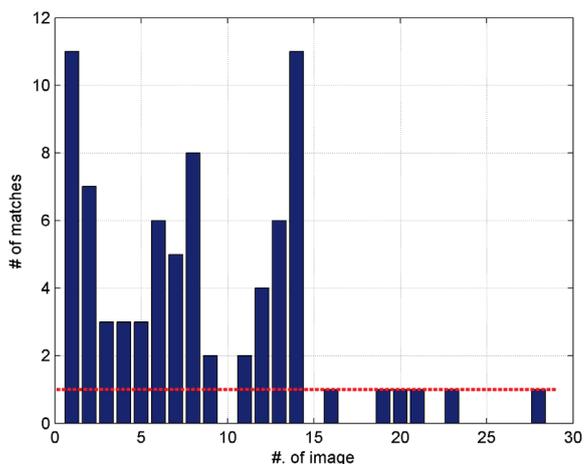


Figure 12 Result of the separation

5 Conclusion

The method described above has the ability to successfully separate images containing ragweed. This phase of the research was primarily focused on the study

of viability of SURF features method for separation of natural images of plants containing ragweed from probe images. Since the method clearly has merits it can serve as the basis of an expert system developed for classification.

Acknowledgements

The project has been supported by the European Union, co-financed by the European Social Fund. Comprehensive Development for Implementing Smart Specialization Strategies at the University of Pécs, EFOP-3.6.1.-16-2016-00004. The authors acknowledge the support of T. Morschhauser (University of Pécs, Faculty of Sciences, Institute of Biology).

6 References

- [1] Lowe, D. G. Distinctive image features from scale invariant keypoints. // *International Journal of Computer Vision*. 60, 2(2004), pp. 91-110.
<https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- [2] Bay, H.; Tuytelaars, T.; Gool, L. V. SURF: Speeded Up Robust Features. // *Computer Vision – ECCV 2006, Lecture Notes in Computer Science*, 3951, 2006, pp. 404-417.
- [3] Malič, B. Consideration of screen and image resolution. // *Tehnički vjesnik*. 17, 3(2010), pp. 367-370.
- [4] Auda, Y.; Blasco, F.; Gastellu-Etchegorry, J. P.; Déchamp, C. Essai préliminaire de détection des champs d'ambrosie par télédétection spatiale. // *Revue française d'allergologie et d'immunologie Clinique*. 42, 5(2002), pp. 533-538.
[https://doi.org/10.1016/S0335-7457\(02\)00178-8](https://doi.org/10.1016/S0335-7457(02)00178-8)
- [5] Auda, Y.; De' Champ, C.; Dedieu, G.; Blasco, F.; Duisit, D.; Pontier, J.-L. Detection of invasive plants using remote sensing: a case study of ragweed in the Rhône-Alps region. // *International Journal of Remote Sensing*. 29, 4(2008), pp. 1109-1124. <https://doi.org/10.1080/01431160701355231>
- [6] Auda, Y.; Déchamp, C.; Méon, H.; Duisit, D.; Pontier, J.-L.; Rakotoniaina, S. Detection of large ragweed populations using remote sensing data. // In: Clot B., Comtois P. & Escamilla-Garcia B. *Aerobiological Monographs Vol. 1, Towards a comprehensive vision*, 2011. ISBN 978-2-8399-0466-7.
- [7] Taramarcz, P.; Lambelet, C.; Clot, B.; Keimer, C.; Hauser, C. Ragweed (Ambrosia) progression and its health risks: will Switzerland resist this invasion? // *Swiss Med WKLY*. 135, (2005), pp. 538-548.
- [8] Gergen, P. J.; Turkeltaub, P. C.; Kovar, M. D. The prevalence of allergic skin test reactivity to eight common aeroallergens in the US population: results from the second National Health and Nutrition Examination survey. // *Journal of Allergy Clinical Immunology*. 80, (1987), pp. 669-679. [https://doi.org/10.1016/0091-6749\(87\)90286-7](https://doi.org/10.1016/0091-6749(87)90286-7)
- [9] Csépe, Z.; Magyar, D.; Mányoki, G.; Bobvos, J.; Elekes, P.; Páldy, A. Recent developments in pollen information in Hungary. // *Egészségtudomány*, LVII., 2013. 4.
- [10] Novák, R.; Dancza, I.; Szentey, L.; Karamán, J. (edt.) *Az ötödik országos gyomfelvételezés Magyarország szántóföldjein*. // Vidékfejlesztési Minisztérium, Budapest, 2012 (ISBN: 978-963-08-1580-2)
- [11] Mányoki, G.; Apatini, D.; Novák, E.; Magyar, D.; Bobvos, J.; Bobvos, G.; Málnási, T.; Elekes, P.; Páldy, A. *Parlagfű helyzetkép és megoldási javaslatok az Aerobiológiai Hálózat mérései alapján és az OKI-AMO feldolgozásában*. // National Institute of Environmental Health. Budapest. 2011.

Authors' addresses

Ádám Schiffer, Assistant Prof., Ph.D.
University of Pécs,
H-7624, Pécs, Rókus 2., Hungary
adam.schiffer@mik.pte.hu

Zoltán Sári, Assistant Prof., Ph.D.
University of Pécs,
H-7624, Pécs, Rókus 2., Hungary
sari.zoltan@mik.pte.hu

Péter Müller, Research Assistant
University of Pécs
H-7624, Pécs, Rókus 2., Hungary
muller.peter@mik.pte.hu

Ildikó Jancsár, Associate Prof., Ph.D.
University of Pécs,
H-7624, Pécs, Rókus 2., Hungary
ildiko.jancskar@mik.pte.hu

Géza Várady, Associate Prof., Ph.D.
University of Pécs,
H-7624, Pécs, Rókus 2., Hungary
geza.varady@mik.pte.hu

Zsolt Ercsey, Assistant Prof., Ph.D.
University of Pécs
H-7630, Pécs, Boszorkány 2., Hungary
ercsey@mik.pte.hu