Proceedings of Student-Faculty Research Day, CSIS, Pace University, May 4th, 2012

An Evaluation of the Effect of Human Interaction on the Accuracy of the Interactive Visual System

Kathryn Durfee, Neville Kapoor, Matthew Muccioli, Richard Smart, David Wilkins, and Amir Schur
Seidenberg School of CSIS, Pace University, White Plains, NY 10606, USA
{nk73448w, kd00283n, mm07510w, rs29636p, dw95537, amir-i-schur}@pace.edu

Abstract

This paper presents the findings of an experiment conducted on the IVS flower recognition application. The objective of this experiment was to evaluate the impact of human-computer interaction on the accuracy of flower identification and provide suggestions for future work to improve the system. The application currently runs on two platforms, a desktop computer and a handheld PDA. In this experiment the desktop application was used. The experiment was done in three phases: manual identification, automatic identification, and man-machine (interactive) identification. In each phase, two parameters were evaluated: the average accuracy of the identification decision and the average time to reach the decision. The results of the experiment reveal that the combination of man-machine identification can offer significant increases in accuracy versus strictly human or computer identification, while also keeping the time required for identification to a minimum. The accuracy provided by the interactive identification, however, left significant room for improvement. Hence, in addition to the analysis of experimental results, this paper seeks to provide some recommendations for improvements to the IVS desktop application.

1. Introduction

Automatic flower recognition tools have been available for some time. For example, a study by Saitoh and Kane in 2000 demonstrated an effective automatic recognition system with high accuracy. In their study they took digital photographs of 16 wildflowers with one frontal flower image and one leaf image taken for each [6]. These images were then sent to a PC where features were extracted and fed into a "neural network" that identified the likely flower species. Flowers and leaves were essentially extracted from their original background and placed on a black sheet where certain key features were then extracted. There were a total of seventeen features of the leaf and the flower, 8 flower features and 9 leaf features. The system proved 95% accuracy with the use of all seventeen features but was less accurate using only the 8 flower features (84%) and even less accurate using only the 9 leaf features (75%). Although the study proved automatic flower recognition is feasible in terms of accuracy, it failed to incorporate the concept of time. At no point do the authors mention how long it took to extract the features from the photographs or for their system to return results. Further, issues with efficiency can be raised as well. For one, seventeen features from the flower and leaf may be time consuming and an over-collection of data. There was also no comparison to manual (human) flower recognition to truly compare the overall efficiency of their system.

There have also been several studies on pattern recognition involving human-computer interaction. Some of these studies have utilized flowers in order to test pattern ideologies and computer software. In 2004 Zou and Nagy evaluated the Computer Assisted Visual Interactive Recognition (CAVIAR) system, which is an interactive flower recognition tool [12]. In order to conduct their experiment they utilized a three-pronged testing format. In the first experiment, human subjects were given a brief introduction to the "rose curve" and then asked to manually classify flowers based on photographs. In the second experiment, computers classified flowers utilizing the "rose curve" without assistance. In the final experiment, the CAVIAR system was implemented where humans could make adjustments to the computer parameters (of the rose curve) in order to assist in locating the correct classification. Zou and Nagy found that human and machine interaction "significantly increased the accuracy compared to the unaided machine" and reduced the recognition time in comparison to the unaided human. In other words, human and machine interaction was more accurate and took less time than other methods. As a result, human and machine interaction proved extremely efficient in this instance. This study proves important not only because of the conclusions but because of the three-pronged experiment that served as a great method for comparison. Perhaps the biggest limitation of the study was its narrow focus and reliance on the "rose curve".

Flower recognition is just one small area of study within pattern recognition. Pattern recognition is being researched and conducted on multiple levels with many diverse applications; one such application is facial recognition. Facial recognition is being used in airports to detect potential threats, police stations to identify criminals, for access control, and in the design of human computer interfaces [5].

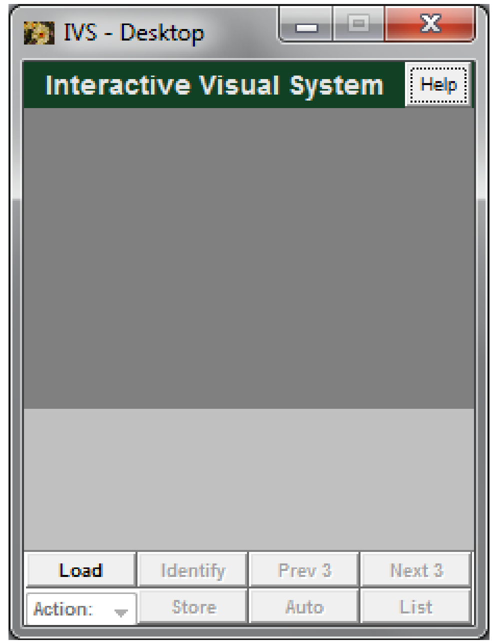
Google Inc. is currently using pattern recognition in creating vehicles that drive themselves. Such vehicles are programmed to detect the images of the street ahead and compare the patterns of the street with the information stored in its database to drive the path accordingly. If the vehicle detects the shape of a STOP sign and the color of the sign being red, it will slow and stop at the sign accordingly [8]. This is a real world application of pattern recognition technology but it is still in a research and development phase with many more advancements to go.

Some pattern recognition systems are similar to the flower recognition system used in the experiment discussed in this report. These systems use metrics to detect certain features in an object, compare them to ones in its database, and use a k-nearest neighbor algorithm to attempt to identify the object [7]. In such systems, understanding how human interaction with the machine can improve its accuracy can be invaluable. The experiment addressed in this report was conducted to evaluate such human interaction on a specific flower recognition system.

2. IVS Software

The CAVIAR system was reengineered to run on a desktop and personal digital assistant (PDA) and the application named the Interactive Visual System (IVS) [3][4]. The current study uses IVS for additional experimentation. The IVS is a java-based software tool designed for speedy and accurate identification of flowers.

Figure 1. IVS Graphical User Interface.



“IVS draws on innate perceptual ability to group "similar" regions, perceive approximate symmetries, outline objects, and recognize "significant" differences. It exploits computer capability of storing image-label pairs, quantifying features, and computing distances in an abstract feature space of shapes and colors. The IVS architecture was developed specifically for isolated object recognition in the field, where the time available for classifying each image is comparable to that of image acquisition” [3].

Users can take pictures of flowers they wish to correctly identify. The pictures are loaded into the IVS system and saved under the 'newImages' directory. The IVS can then be used to identify the flower. The IVS software uses characteristics of the flower petals and stamen when attempting flower identification.

Figure 2. Flower to be identified loaded in IVS.

Identification can be attempted automatically (no human interaction with the software) or the user can interactively assist IVS by selecting specific features of the flower to compare against the database. These features are found in a drop-down menu on the IVS GUI. They include the following: Petal count, Petal color 1, Petal color 2, Stamen color. Additionally, IVS can be used to crop the picture so that only the flower to be identified is processed by the software, and not the background. Further, IVS allows for the user to draw an outline around the petal shape to help narrow the range of flower petal choices.

Once one, some, or all of the features have been selected, the user then presses the Identify button. This directs IVS to compare the properties of the flower being identified with those of the flowers in its database. IVS then returns the top three closest matches for the user to choose from.

Figure 3.Top three choices selected by IVS

If a known species of a flower is identified this additional sample picture can be added to the database. New species can also be added to the IVS database by using the Store button. For this to be successful, all the features of the new flower have to be identified manually and then added to IVS. After this is done, IVS will allow the user to give the new species of flower a name and also to store it in the database.

Upon acquiring the software for this experiment, the software was preloaded with a flower database consisting of twenty five (25) different species of flowers. IVS allows for the user to add to the flower database by storing newly identified flowers. The experiment discussed in this report was conducted on an expanded inventory of 131 species of flowers.

3. Methodologies

This section explains the method of expanding the flower image inventory for both establishing references and for testing, and the method of conducting the experiments.

Expanding Inventory and Reference Data

In order to build the inventory of reference data in the IVS system and conduct a more thorough experiment, two months were spent gathering flower images. While it would have been preferred to work with actual flower samples, our urban location, the winter season, a limited time frame and budget, and our lack of a floral expert for identification necessitated our collection through other means. Flower images were gathered primarily via online databases provided by botanical gardens, agricultural educational institutions, and seed distributors; sites that could be trusted to have accurately identified images [2, 9, 10, 11]. Our image collection resulted in a total of **106 new species** to add to the existing reference set.

Prior to loading new image data, one image per species was pulled aside to be used as test data in the final experiment. It was important that these test images were not stored as reference data or the accuracy of the system would be biased. For all other (non-test) images collected, feature data was extracted and stored into the data.txt file, making that image part of the reference data for the experiment. In brief, the storage process requires the user to load each image and add feature data for all 6 features (petal color1, petal color2, stamen color, petal count, crop area, and petal outline) using the *Action* menu. After all feature data are selected, the *Store* tool is used, which adds the data input to the reference file (data.txt) and creates a new species folder within the images folder. Subjects were requested to not save images of such poor quality that they could not accurately extract at least 4 of the features, and this occurred in less than 3% of the images gathered.

Although the workload of storing new image data was divided amongst team members, the final data.txt file and all images were shared to ensure that all systems used in the final experiment were identical. Team members each sent their final species image folders and their data.txt files to the coordinator of the experiment, who compiled the data and transposed it to each subject's computer at the start of the experiment. The coordinator of the experiment reviewed each species folder sent by the team and removed any that seemed highly inaccurate (i.e. the flowers contained did not resemble one another and could not be believed to be of the same species). This resulted in the removal of roughly 12 species from our potential reference set. The final inventory used for the experiment combined both the initial species (25) with our new species (106) for a **total of 131 species and 535 images.**

Experiment Methodology

It should be noted that prior to conducting the experiment on our expanded inventory, a preliminary test was completed using 20 test images and the original reference data provided. This test used a methodology similar to that of the final experiment described below and allowed for adjustments and improvements to be made to the process prior to the final experiment. Any changes to the methodology from the preliminary test are noted as such in this section.

The final experiment was set-up in three phases: manual identification, auto identification, and interactive identification. The interactive identification stage was further comprised of three separate tests, each evaluating the use of a different set of features. Interactive A evaluated the use of just the petal count feature. Interactive B added petal and stamen colors to the petal count feature (4 features) and Interactive C added crop area (5 features).

For the final experiment, the sample size was increased to 30 test images (from 20 in the preliminary test). The test files were renamed Images 1-30 and placed into a file to be distributed to each of the subjects of the experiment. A key was made allowing the coordinator to quickly evaluate the subjects' accuracy later.

In the preliminary test, different test images were used for the manual and auto sections. For the final experiment, however, more control was added to the experiment and the same set of test images was used for all phases of the testing. It was initially thought that testers should be divided into two groups—one to start with the manual phase and the other with the auto phase—in order to reduce biases. Starting subjects with the auto phase and then having them proceed to manual would certainly skew results, as testers may recall species' names as provided by the IVS, thus inflating their accuracy on the manual phase. However, starting subjects with the *manual* phase would not result in inflated (or deflated) accuracy on the auto phase, provided the subjects were not told whether or not their manual answers were correct until after the experiment was completed. For this reason, all subjects in our final experiment began with the manual phase. This also allowed us to maximize efficiencies throughout the experiment.

In the manual phase, subjects were presented with the image of the flower and asked to immediately start a stopwatch. All subjects then used the same guidebook to attempt to accurately identify the flower [1]. Once the subject believed that he had accurately identified the flower, the stopwatch was stopped and the flower name recorded. Subjects were asked to record the species name to ensure the experiment's coordinator could come to a conclusive decision on the subject's accuracy. Following the preliminary testing, it was discovered that subjects would often take in excess of 10 minutes to identify a flower in the guidebook and still guess incorrectly. For this reason, the team and the client agreed to set a limit of 5 minutes for the manual section. If the subject's stopwatch reached 5 minutes, they were instructed to record their best guess and > 5minutes as the time. The coordinator of the experiment reviewed the answers provided by the subjects against the key and identified them as being correct (100% accurate) or incorrect (0% accurate). For reasons previously discussed, the accuracy of the subject's guess was not revealed until the conclusion of the entire experiment.

In the auto phase, subjects loaded each image to the IVS software, started their stopwatch and selected the *Auto* and *Identify* buttons to reveal the software's top matches. Upon identifying the position of the flower believed to be the correct match, the subject stopped the timer and recorded the position (e.g. 2nd place). The coordinator of the experiment stood by to inform the tester as to whether he was correct or incorrect and to record the position of the correct match as indicated in the key. It should be noted here that because all subjects were operating on identical systems, the auto phase returned the correct image in the same position for each subject. During the interactive phases, user input varies and so the position of the correct image is different for each subject. Having each of the subjects replicate the auto phase verified that the application can be used in a consistent manner to produce identical results.

In the interactive phases, testers provided human input through the selection of feature data. Again, the stopwatch was started at the time the image was loaded to the IVS screen. Subjects used the action menu to select and input the feature data being evaluated and then selected the *Identify* button. Subjects then reviewed the flowers returned and stopped the timer when they felt they had a correct match. The subject would then record his guess (e.g. 1st place, 2nd place) and the experiment's coordinator would note the position of the correct match.

4. Results

Before discussing the final experiment data results it is important to do a brief review of the preliminary experiment. In the preliminary experiment the IVS auto identification feature had an average accuracy of 27.50%. In other words, the IVS placed the correct flower in the top three positions nearly thirty percent of the time. Even more impressive, though, was the IVS accuracy increases during the interactive phases. The average accuracy during the first interactive phase was 41.25%. This number increased to 62.50% and then 66.25% in the final interactive phases, which utilized the most human interaction. From this data it appeared the IVS accuracy increased as the amount of human interaction with the system increased. It is also important to note that in the preliminary experiment manual human identification was around 50% accurate with an average time of three minutes and thirty seconds for identification. Given that none of the IVS interactive phases took longer than 52 seconds, the IVS was proving to be a quick and efficient flower identification system.

For the final data experiment the changes discussed in the methodology section were made. In terms of accuracy, the IVS auto identification feature was 13.33% accurate in placing the correct flower in the top three positions. The system showed only a slight increase in accuracy for the first interactive phase with a 15.56% accuracy. There was a large jump in accuracy for the second interactive phase with an average of 51.11% accuracy. The final interactive phase, however, had a decrease in accuracy to 38.89%. This result was somewhat unexpected due to the results of the preliminary data, which had a steady increase of accuracy throughout the interactive phases. That being said, the final interactive phases again proved more accurate than manual human identification, which had an average accuracy of only 32.22%. The change in accuracy results from the preliminary experiment to the final experiment may be a result of the larger image database as well as the quality of images entered into the database.

Table 1. Accuracy and Time results.
For the tests involving IVS the percentages indicate the percentage of times the correct flower was in the top three choices.

Test Type	Accuracy Tester 1	Tester 2	Tester 3	Average	Time Average
Manual Stage	40.00%	36.67%	20.00%	**32.22%**	**2:53**
Auto Phase	13.33%	13.33%	13.33%	**13.33%**	**55.9**
Interactive A	13.33%	16.67%	16.67%	**15.56%**	**43.8**
Interactive B	63.33%	60.00%	30.00%	**51.11%**	**40.5**
Interactive C	40.00%	36.67%	40.00%	**38.89%**	**44.4**

Although the accuracy findings for the top three positions were lower than expected the IVS system was still returning correct results. That is, the correct image was always an option within the IVS results and was typically within the top twenty images. For instance, the data was expanded to show the IVS accuracy within the top five, top nine and top nineteen placements. The top five data, even though only two positions were added, had an overall increase in accuracy with the final IVS phase showing a 53.33 % accuracy. The accuracy for the top nine positions (see table 2) also had a substantial increase. The first interactive phase was 30%, the second was 60% and the final phase was 68.89% accurate. This data proved more consistent with the original preliminary results and again showed the phases with the most human interaction were the most accurate. The top nineteen accuracy results (table 3) are important mainly because of the high percentages. The final two interactive phases had a 68.89% and 81.11% accuracy. Thus, with the help of human interaction the IVS was returning the correct flower in the top twenty positions 80% of the time. Therefore, the IVS may not have been returning the correct flower in the top few positions but it was still returning the correct results within reasonable position for the user. That is, the user could still locate the correct flower in the IVS results with a few clicks of the mouse.

Table 2. Interactive Accuracy Results.
For the tests involving IVS the percentages indicate the percentage of times the correct flower was in the top nine choices.

Top 9 Positions

Test Type	Accuracy			
	Tester 1	Tester 2	Tester 3	Average
Interactive A*	23.33%	36.67%	30.00%	**30.00%**
Interactive B*	70.00%	66.67%	43.33%	**60.00%**
Interactive C*	70.00%	66.67%	70.00%	**68.89%**

Table 3. Interactive Accuracy Results.
For the tests involving IVS the percentages indicate the percentage of times the correct flower was in the top nineteen choices.

Top 19 Positions

Test Type	Accuracy			
	Tester 1	Tester 2	Tester 3	Average
Interactive A*	56.67%	46.67%	46.67%	**50.00%**
Interactive B*	80.00%	70.00%	56.67%	**68.89%**
Interactive C*	83.33%	73.33%	86.67%	**81.11%**

In regards to time, the final experiment data was similar to the preliminary findings. Table 1 shows the manual identification stage had an average of two minutes and fifty-three seconds to locate the correct flower. The IVS phases were much lower with the auto phase having an average of 55.9 seconds. It is important to note that the IVS system itself takes an average of 2.033 seconds to return its results. The rest of the time is spent by the user attempting to locate the correct flower in the results. The first interactive phase had an average of 43.8 seconds for the user to locate the correct flower. The final two phases had 40.5 seconds and 44.4 seconds. Within the time measurement there are a lot of small factors to take into account. For one, there may be a correlation between accuracy and time. The auto phase took nearly 56 seconds to locate the correct flower but this is likely because of the poor accuracy in this phase. Also, the final interactive phase required the user to enter the most flower data, which is likely why it took longer than other interactive phases.

5. Conclusions

To conclude, the IVS proved to be a fairly efficient mechanism for flower identification in two important ways. For one, the IVS interactive testing exhibited an effective level of accuracy. In other words, although the IVS did not exhibit high-level accuracy it still returned the correct results within a reasonable distance. The correct flower was not guaranteed to be in the top three positions but was very likely to be within the top twenty positions. If nothing else, this allows a human user to navigate the results to locate the flower without an extensive search. In addition the IVS interactive testing demonstrated a significantly higher accuracy than the manual human identification. The manual testing proved not only inaccurate but very time consuming. This leads into the second point for IVS efficiency, which is time. The IVS proved to be extremely quick in returning results and ultimately for users to locate flowers. On average the IVS took about two seconds to generate results with additional time needed to enter interactive features and scroll through results. Even with such additions, no interactive testing phase averaged more than forty-five seconds to locate the correct flower. As a result, the IVS flower location times were a substantial improvement from manual identification times.

Whether efficient or not, the IVS demonstrated the importance of human interaction in pattern recognition. The auto identification testing, which relied solely on the IVS, exhibited extremely poor accuracy that seemed to worsen with a large flower database. This inaccuracy began to effect the user flower location times because they were sometimes forced to scroll through one hundred images before finding the correct flower. However, as the interactive testing phases began the user was asked to start entering flower features into the IVS. The first interactive testing phase had the user enter petal count and resulted in a large spike in accuracy from the auto identification phase. As the interactive phases continued the users entered more flower features such as color and the IVS accuracy appeared to increase. The data for the top three and top five positions does show a decrease in accuracy for the final interactive phase. However, with the data expanded to reveal the top nine and nineteen positions, steady increases can be seen as more human interaction is added.

The decrease in accuracy that was seen in the interactive phase of the final experiment, in comparison to the results of the preliminary testing, may be a result of poor quality images. As previously mentioned, images were gathered from online databases and some of these images were not always ideal images to use for a reference set. Poor lighting in images can skew the RGB color reading. Images where the flower is too distant make capturing accurate petal counts, stamen colors, or petal outlines difficult. Many images of flowers tend to have multiple flowers in the shot, making it difficult for the IVS to accurately extract a single flower sample from its background. When these types of poor quality images are used, the data entered into the reference set is less reliable and may adversely affect the IVS's ability to accurately identify flowers.

Additional experiments could be conducted on the IVS in the future, but it is recommended that a more controlled collection of quality images for the reference set and a few additional revisions to the IVS application. Some of these suggestions are outlined in the subsequent section.

Overall, the outcome of this experiment demonstrated that man-machine interaction can help significantly improve accuracy in the identification of flowers. This result is encouraging for other possible identification systems, of a wide-variety, which could benefit from additional research and adoption of man-machine interaction.

6. Recommendations for Future Work

Functionality of IVS

The IVS software is an effective tool for identifying flowers but improvements can be made to make it more user-friendly, which could in turn improve accuracy. A larger screen for the user to work from would be helpful. When looking at images, small flowers are made even smaller because of the size of the IVS' Graphical User Interface (GUI). The GUI should be made full screen, allowing for easier viewing of the flower images at full size. The returned results should also include a brief description of the flower as well instead of just a flower name. Revising the application to allow for full screen expansion would provide room for these additions. More prompts for the user as to which buttons to select and when and what actions are being taken could also be useful. The IVS should offer a way of zooming in on the images as well. This would be particularly helpful when extracting features from an image with the intent of adding to them to the reference set. If the user can zoom in, he/she may be able to more accurately outline petals, select stamen colors, count petals, etc. The IVS could also have a better light and color sensing algorithm to differentiate between shades, shadows, and petal colors when in less than optimal lighting.

During testing, the IVS system froze up and was also slow in returning results of the identified flowers. This problem seemed to occur after loading our new species and expanding the set of reference images. This may be a small glitch or it may be related to the system scanning several flower images to present results. If the latter is the case, this may be an issue. This experiment worked from a reference set of 130+ species of flowers which resulted in over 530 images on the system. If this set was increased, say to 10000+, the system would probably not be able to handle so much reference data and either crash or take even longer to return results.

Building a Better Reference Set

The accuracy of the system would benefit from better quality reference images. All of the images that currently comprise the reference set are of varying quality. Many have complicated backgrounds, poor lighting, awkward angles, or show bunches of the same flower versus just one sample. Using these varying flower images as the reference set are likely to be negatively impacting the system's ability to accurately identify flowers. Subjects of this experiment often noted that the better the image quality, the better the results obtained were. It is suggested that a skilled botanist assist with collecting actual samples of each flower type and photographing at least one image per species on a solid black background in good lighting. These images would be stored in the reference set and hopefully improve the accuracy of the system. This approach was something taken by Saitoh and Kane in the previously mentioned article.

Dissemination/Usage of IVS

In addition to researching Smartphone applications, a web application version should be considered. A web application would enable the owner/administrator of the application to continue to update the reference set and push these updates out to the users. The IVS desktop application could also connect to the web for updates as well. Users should be able to download to their system's reference set, more accurate and visible images as they become available. Images should be updated as needed and somehow made to be distributed to individual end users' IVS systems.

10. References

[1] D.M. Bradenburg, *Field Guide to Wildflowers of North America.* New York, NY: Sterling Pub Co., 2010.

[2] Connecticut Botanical Society, Gallery of Connecticut Wildflowers, http://www.ct-botanical-society.org/galleries/galleryindex.html, accessed March 2012

[3] A. Evans, J. Sikorski, P. Thomas, J. Zou, G. Nagy, S.-H. Cha,and C.C. Tappert, "Interactive Visual System," Pace CSIS Tech. Rep. no. 196, 2003.

[4] A. Evans, J. Sikorski, P. Thomas, S-H Cha, C. Tappert, G. Zou, A. Gattani, and G. Nagy, "Computer Assisted Visual Interactive Recognition (CAVIAR) Technology," 2005 IEEE International Conference on Electro-Information Technology, Lincoln, NE, May 2005 (Proceedings on CD-ROM only).

[5] X.Lu. (2003, May). *Image Analysis for Facial Recognition.* Available: https://www.msu.edu/~lvxiaogu/publications/ImAna4FacRcg_Lu.pdf

[6] T. Saitoh,. & T. Kane, "Automatic recognition of Wild Flowers," Proceedings of 15th International Conference on Pattern Recognition (ICPR '00), vol. 2, 2000. doi: 1051-4651/00

[7] A. Schur, private communications, February-April 2012.

[8] T. Vanderbilt. (2012, January 20). *Let the Robot Drive: The Autonomous Car of the Future is Here*. [Wired Magazine.] Available: http://www.wired.com/magazine/2012/01/ff_autonomouscars/all/1. Accessed April 2012.

[9] Wildflowers in Bloom, Horticulture Program at Texas A&M, http://aggie-horticulture.tamu.edu/wildseed/wildflowers.html, accessed March 2012.

[10] WildflowerInformation.org, http://www.wildflowerinformation.org/BotanicalList.asp, accessed March 2012.

[11] Wildflowers: Wildflower Finder, http:// http://wildflowerfinder.org.uk/, accessed March 2012.

[12] J. Zou & G. Nagy,. "Evaluation of model-based interactive flower recognition," Proceedings of the 17th International Conference on Pattern Recognition, volume 2, 2004 . doi: 1051