Sketch Based Flower Detection and Tracking

D. S. Guru, Y. H. Sharath Kumar and M. T. Krishnaveni

Abstract In this paper, we present a system for detecting and tracking of a flower in a flower video based on a query sketch of the flower. The proposed system has two stages detection and tracking. In first stage a sketch of a flower of interest is given as an input. The edge orientation information of the given sketch is matched against that of an individual frame in search of a location of the flower of interest using fast directional chamfer matching. In second stage the detection coordinates have been used for tracking the sketch part in flower videos. For tracking we used joint color texture histogram to represent a target and then apply it to the mean shift framework. For experimentation we created our own dataset of 10 videos of different flowers and their sketches. To study the efficiency of the proposed method we have compared the obtained results provided by five human experts.

Keywords Chamfer matching · Flower detection · Flower tracking · Color texture histogram

1 Introduction

A sketch is a rapidly executed freehand drawing that may serve a number of purpose, it might record something that the artist sees, it might record or develop an idea for later use or it might be used as a quick way of graphically

D. S. Guru (✉) · Y. H. Sharath Kumar · M. T. Krishnaveni
Department of Studies in Computer Science, University of Mysore, Mysore 570006, India
e-mail: dsg@compsci.uni-mysore.ac.in

Y. H. Sharath Kumar
e-mail: sharathyhk@gmail.com

M. T. Krishnaveni
e-mail: mtkveni@gmail.com

P. P. Swamy and D. S. Guru (eds.), *Multimedia Processing, Communication and Computing Applications*, Lecture Notes in Electrical Engineering 213,
DOI: 10.1007/978-81-322-1143-3_25,

demonstrating an image, idea or principal. It is an excellent way to quickly explore concepts. The objective of the current research work is to detect flowers from videos through sketch of flowers. Flowers and the ability to identify them have been fascinating humans for hundreds of years. The taxonomy originally contained approximately 8,000 plants, but has since been extended to encompass more than 250,000 flower species around the world [1]. However, even when an image is sufficient, identifying a flower may still need a guidebook because with advances in digital and mobile technology it is easy to draw pictures of flowers, but it is still difficult to find out what they are. Once we know the name of a flower we can find more information about a flower on the web, but the link between obtaining an image of a flower and acquiring its name is missing. Therefore, our aim is to create an automatic guide that identifies a sketch of a flower.

2 Related Work

The Classification of flowers has majorly three stages viz., segmentation, feature extraction and classification. Before extraction of features from a flower image, the flower has to be segmented. The goal is to segment out the flower given only that the image is known to contain a flower, but no other information on the class or pose. In second step, different features are chosen to describe different properties of the flower. Some flowers are with very distinctive shapes, some have very distinctive color, some have very characteristic texture patterns, and some are characterized by a combination of these properties. Finally extracted features are used to classify the flower.

Segmentation subdivides an image into its constituent parts or objects. The level to which this subdivision is carried depends on the problem being solved. That is segmentation should stop when the objects of interest in an application have been isolated. In general, automatic segmentation is one of the most difficult tasks in image processing. Flowers in images are often surrounded by greenery in the background. Hence, the background regions in images of two different flowers can be very similar. In order to avoid matching the green background region, rather than the desired foreground region, the image is segmented. Pixel labeling method [2] uses only pixel appearance to assign a label to a pixel. The Contour-based methods which try to find the boundary of an object by locally minimizing energy function so that the segmentation boundaries align with strong gradients in the image. These include [3–6]. Graph-based pixel labeling methods a global energy function is defined depending on both appearance and image gradients [7–11]. Another classical category of segmentation algorithm is based on the similarity among the pixels within a region, namely region based segmentation. In region merging techniques, the goal is to merge regions that satisfy a certain homogeneity criterion. These includes [12–14].

Different features are chosen to describe different properties of a flower. Some flowers are with very distinctive shapes, some have very distinctive colors, some have very characteristic texture patterns, and some are characterized by a combination of these properties. Some flowers exist in a wide variety of colors, but many have a distinctive color. The color of a flower can help narrow down the possible species, but it doesn't enable us to determine the exact species of the flower. To handle this problem the color feature is described by taking the HSV values of the pixels [10]. The HSV values for each pixel in an image are clustered using k-means to have the color vocabulary. Yoshioka et al. [15] in their work performed quantitative evaluation of petal colors using principal component analysis. They set a region of interest in each petal as a region representing the petal color pattern and defined the maximum square on each petal as the region of interest. Texture of a flower has also been exploited for classification. Some flowers have characteristic patterns which are distinctive on their petals. Nilsback and Zisserman [16], describe the textures by convolving the images with MR8 filter bank. The filter bank contains filters at multiple orientations. Rotation invariance is achieved by choosing the maximum response over orientations. Guru et al. [17] developed a neural network based flower classification system using different combinations of texture models such as color texture models, gray level co occurrence matrix, gabor responses. Guru et al. [18] designed a flower classification system using combinations of gray level co occurrence matrix, gabor responses. The features are fed into K nearest neighbor for classification. Guru et al. [19] proposed a method to classify flowers using only whorl region of flowers. The whorl region is identified using noise obtained through Gabor filter responses. Different texture features are extracted on whorl part of flower and compared with entire flower. The shapes of individual petals, their configuration, and the overall shape of the flower can all be used to distinguish flowers. The difficulty of describing a shape is increased due to natural deformations of a flower. The petals are often very soft and flexible and hence can bend, curl, twist etc., which make the shape of a flower appear very different. The shape of a flower also changes with the age of the flower and petals might even fall off. Nilsback and Zisserman [10] describe the shape features using rotation invariant descriptors. The scale invariant feature transform (SIFT) descriptors are computed on a regular grid and optimize over three parameters: grid spacing M, radius R and number of clusters. Nilsback and Zisserman [16] describes the shape features using SIFT descriptors on the foreground region and on the foreground boundary.

After feature extraction, the challenge lies in determining suitable classifier. Nilsback and Zisserman [10, 16] used nearest neighbor classifier and support vector machine to classify the flowers. In other work Varma and Ray [20] used multiple kernel classifier to classify the flowers. However, as the number of classes increases classification becomes computationally expensive. To overcome this problem Das et al. [2] proposed an indexing method to index the patent images using the domain knowledge. The color of the flower is defined by the color names present in the flower region and their relative proportions. The database can be queried by example and by color names. Fukuda et al. [21] developed a flower image retrieval system by combining multiple classifiers using fuzzy c-means clustering algorithm. Cho and Chi [22] proposed a structure-based flower image recognition method. The genetic evolution algorithm with adaptive crossover and mutation operations was employed to tune the learning parameters of the Back-propagation through Structures algorithm [23]. Saitoh et al. [6] describe an automatic recognition system for wild flowers. The objective is to extract both flower and leaf from each image using a clustering method and then to recognize using a piecewise linear discriminate function. In [24] color features of flower are characterized using a histogram of a flower region and shape features are characterized by centroid-contour Distance and Angle Code Histogram for the purpose of flower retrieval.

From the literature survey it is understood that, though there are a few attempts towards development of flower classification systems, no work is found on sketch based flower detection and Tracking. Hence in this work, we design a system for detecting and tracking of a flower in a flower video based on a query sketch of the flower. The proposed system has two stages detection and tracking. In first stage a sketch of a flower of interest is given as an input. Fast directional chamfer matching is used match the flower sketch in respective videos frames and later best detection co-ordinates as been picked for tracking the sketch part in flower videos.

3 Proposed Method

The proposed method has detection and tracking phases. In detection phase, the sketch of user interest is given as input. The edge orientation information of the given sketch is matched against that of an individual frame in search of a location of the flower of interest using fast directional chamfer matching. The best matching score between human expert and proposed method is selected for tracking. In tracking phase the detection coordinates have been used for tracking the sketch part in flower videos. For tracking we used joint color texture histogram to represent a target and then apply it to the mean shift framework. The block diagram of the proposed method is given in Fig. 1.

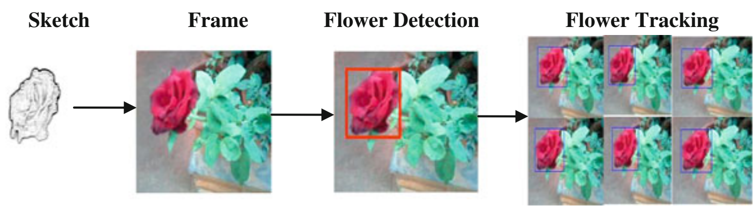

Fig. 1 Block diagram of the proposed work

3.1 Flower Detection

As flowers of different classes are more similar, developing a system to identify flowers based on sketches is a very challenging task. Additionally, flower videos captured in a real time, poses a number of challenges like variations in viewpoint, scale, illumination, partial occlusions, multiple instances etc. Also, the cluttered background makes the problem more difficult, as we need to identify the flowers from the background. Moreover, the greatest challenge lies in preserving the intra-class and inter-class variabilities.

All these challenges need a very sophisticated algorithm to identify flowers based on sketch. As we do not find any work on sketch of flowers specifically on videos so we shifted our focus towards object detection based sketches. By reviewing the research papers on sketch based object detection, we find efficient shape-matching algorithm called fast directional chamfer matching (FDCM) [25] which is used to reliably detect objects and estimate their poses. FDCM improves the accuracy of chamfer matching by including edge orientation. It also achieves massive improvements in matching speed using line-segment approximations of edges, a three-dimensional distance transform, and directional integral images.

The input sketch is matched with respective frames of flower videos using shape matching algorithm called fast directional chamfer matching which incorporates the edge orientation information of both query sketch and input frame. The RANSAC algorithm is used to compute the linear representation of an edge map. The algorithm initially hypothesizes a variety of lines by selecting a small subset of points and their directions. The support of a line is given by the set of points which satisfy the line equation within a small residual and form a continuous structure. The line segment with the largest support is retained and the procedure is iterated with the reduced set until the support becomes smaller than a few points.

Let $\mathrm{T} = \{t_i\}$ and $\mathrm{Q} = \{q_j\}$ be the sets of template and query edge map respectively. Let $\phi(t_i)$ denote the edge orientation of the edge point t_i. For a given location x of the template in the query image, directional chamfer matching aims to find the best $q_j \in Q$ for each $t_i \in T$ by minimizing the cost Eq. (1)

$$|(t_i + x) - q_j| + \lambda|\phi(t_i + x) - \phi(q_j)|. \tag{1}$$

Thus the directional chamfer distance for placing the template at location x is defined as

$$d_{DCM}^{(T,Q)}(x) = \frac{1}{|T|}\sum_{t_i \in T} \min_{q_j \in Q} |(t_i + x) - q_j| + \lambda|\phi(t_i + x) - \phi(q_j)| \tag{2}$$

λ denotes the weighting factor between location and orientation terms.

3.2 Flower Tracking

The best detection co-ordinates that are obtained through matching between ground truth and proposed method are used to represent the target region and then a joint color-texture histogram method is adapted for a more distinctive and effective target representation. The major uniform Local binary patterns (LBP) patterns are used to identify the key points in the target region and then form a mask for joint color-texture feature selection. The mean shift algorithm [26] can be used for visual tracking. The simplest such algorithm would create a confidence map in the new frame based on the color histogram of the object in the previous frame, and use mean shift to find the peak of a confidence map near the object's previous position.

The LBP operator labels a pixel in an image by thresholding its neighborhood with the center value and considering the result as a binary pattern [27, 28], $LBP_{8,1}$ ($P = 8$, $R = 1$)(P = #neighbors; R = radius). By varying P and R, we have the LBP operators under different quantization of the angular space and spatial resolution, and multi resolution analysis can be accomplished by using multiple $LBP_{P,R}$ operators. The superscript "*riu2*" means that the rotation invariant "uniform" patterns have a U value of at most 2. The $LBP^{riu2}_{8,1}$ model has nine uniform texture patterns; each of the $LBP^{riu2}_{8,1}$ uniform pattern is regarded as a micro-texton. The local primitives detected by the $LBP^{riu2}_{8,1}$ model include spots, flat areas, edges, line ends and corners, etc. In target representation, the micro-textons such as edges, line ends and corners, named as "major uniform patterns" are used to represent the main features of the target while, spots and flat areas, which are named as "minor uniform patterns" are representing minor textures. Thus, main uniform patterns are extracted from the target.

The RGB channels and the LBP patterns are jointly extracted to represent the target and they are embedded into the mean shift tracking framework. To obtain the color and texture distribution of the target region, the distribution of color and texture of the target model is of ($8 \times 8 \times 8 \times 5$) dimension where first three dimensions (i.e. $8 \times 8 \times 8$) represent the quantized bins of color channels and the fourth dimension (i.e. 5) represents the modified LBP texture patterns (Ning et al. [29]).

4 Experimentation

We created a dataset with 10 flower videos. Figure 2 shows an example frame from each flower videos. These are consist flowers commonly occurring in and around Mysore city, Karnataka, India. The videos are taken to study the effect of our proposed method with large intra class variation.

For experimentation we created ground truth for both detection and tracking phase. In detection phase we asked human expert to draw manually a minimum

Fig. 2 Shows one example frame from each flower videos created

Fig. 3 Set of flowers video frames with minimum *rectangular box* fixed by the proposed detection method on flower region

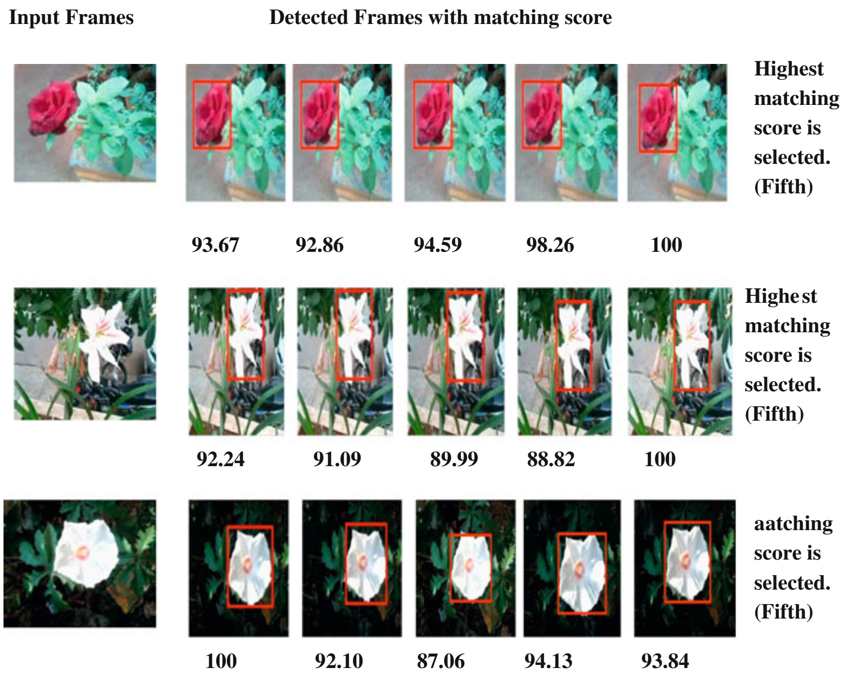

Fig. 4 Frames detected along with their matching scores for different input frames

Fig. 5 Misidentification of flowers. **a** Input frames. **b** Output of the proposed method with matching score

Fig. 6 *Column 1* shows a set of frames. *Column 2* Shows the selected frame for tracking through maximum matching score. *Column 3* Tracking matching score between the coordinates obtained by the proposed method and corresponding human experts

rectangular box on flower region which is treated as ground truth for our experimentation. The sketch is given as input it is detected in the respective frames. Figure 3 shows some samples of flower images containing minimum bounding rectangular box fixed by the proposed method. We calculate percentage of matching by comparing the co-ordinates obtained by the expert and detection

Table 1 Overall matching scores for the output of the proposed method w.r.t ground truth created by 5 human experts

Flower videos no.	Human expert 1	Human expert 2	Human expert 3	Human expert 4	Human expert 5
1	95.63	96.04865	94.79122	96.73176	96.51785
2	100	100	100	99.75758	100
3	94.3199	95.45312	98.00295	96.60815	98.26666
4	85.58392	86.23154	88.19397	87.74507	89.79025
5	96.41151	100	93.56215	100	99.92304
6	90.83918	94.79543	89.83116	95.26265	94.8868
7	100	100	100	100	100
8	87.1321	85.45818	85.87476	91.15284	94.3773
9	89.89978	90.02119	88.89532	91.4034	91.69119
10	81.4677	86.31115	86.81036	84.20956	84.65697

co-ordinates. The co-ordinates of best matching score between proposed method and human expert is selected for tracking. Figure 4 shows how best frame is selected based on their matching score. On the other hand Fig. 5 shows some samples with misidentification of flower and the percentage of overlapping with human experts. In tracking phase the best matching score co-ordinates is used to identify the target region for tracking. We asked five human experts to draw manually a minimum rectangular box on flower region which is treated as ground truth. We calculate percentage of matching by comparing the co-ordinates obtained by the experts and tracking co-ordinates. Figure 6 shows some samples flower images bounding box fixed by the proposed method, human expert1, human expert2, human expert3, human expert4 and human expert5 along with their matching score. Overall, the matching score of bounding box fixed by proposed method and human experts is shown in Table 1. The Proposed method achieves 93.46 % matching score across five human experts for 10 flower videos.

5 Conclusion

In this work we have used the fast directional chamfer matching to identify the flower region in videos using sketches provided by user. The coordinates of the best detected frame are used to track the region of flower using color and texture information throughout the video. We have conducted experimentation on our own dataset. To corroborate the efficiency of the proposed method we have created ground truth where five human experts have identified the flower region by drawing rectangular bounding box manually. Later we matched the bounding box drawn by the proposed method with bounding box of human experts to study the error analysis. In future we intend to develop a multi tracking of flower naming system based on sketches of the flowers.

References

1. Linneaus C (1759) Systemae naturae. Impensis Direct Laurentii Salvii
2. Das M, Manmatha R, Riseman EM (1999) Indexing flower patent images using domain knowledge. IEEE Intell Sys 14:24–33
3. Kass M, Witkin A, Terzopoulos D (1987) Snakes: active contour models. Int J Comp Vis 1(4):321–331
4. Chan TF, Vese LA (2001) Active contours without edges. IEEE Trans Image Process 10:266–277
5. Mortensen E, Barrett WA (1995) Intelligent scissors for image composition. In: Proceedings of ACM SIGGRAPH, pp 191–198
6. Saitoh T, Aoki K, Kaneko T (2004) Automatic recognition of blooming flowers. In: The Proceedings of 17th international conference on pattern recognition 1:27–30
7. Boykov Y, Jolly MP (2001) Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In: Proceedings of international conference on computer vision (ICCV–01), vol 2, pp 105–112
8. Rother C, Kolmogorov V, Blake A (2004) Grabcut: interactive foreground extraction using iterated graph cuts. In: Proceedings of the ACM SIGGRAPH conference on computer graphics, vol 23(3), pp 309–314
9. Kumar MP, Torr PHS, Zisserman A (2005) OBJ CUT. In: Proceedings of the IEEE conference on computer vision and pattern recognition, San Diego, vol 1, pp 18–25
10. Nilsback ME, Zisserman A (2006) A visual vocabulary for flower classification. In: The proceedings of computer vision and pattern recognition, vol 2, pp 1447–1454
11. Nilsback ME, Zisserman A (2004) Delving into the whorl of flower segmentation. In: The proceedings of British machine vision conference, vol 1, pp 27–30
12. Calderero F, Marques F (2010) Region merging techniques using information theory statistical measures. IEEE Trans Image Process 19(6):1567–1586
13. Calderero F, Marques F (2008) General region merging approaches based on information theory statistical measures. In: The 15th IEEE international conference on image processing, pp 3016–3019
14. Ning J, Zhang L, Zhang D, Wu C (2010) Interactive image segmentation by maximal similarity based region merging. Pattern Recognit 43(2):445–456
15. Yoshioka Y, Iwata H, Ohsawa R, Ninomiya S (2004) Quantitative evaluation of flower color pattern by image analysis and principal component analysis of *Primula sieboldii* E. Morren Euphytica, pp 179–186
16. Nilsback ME, Zisserman A (2008) Automated flower classification over a large number of classes. In: The proceedings of sixth indian conference on computer vision, graphics and image processing, pp 722–729
17. Guru DS, Sharath YH, Manjunath S (2011) Textural features in flower classification. Math Comput Model 54(3–4):1030–1036
18. Guru DS, Sharath YH, Manjunath S (2010) Texture features and KNN in classification of flower images. IJCA, Special Issue on RTIPPR (1), pp 21–29
19. Guru DS, Sharath YH, Manjunath S (2011) Classification of flowers based on whorl region. In: 5th Indian international conference on artificial intelligence, pp 1070–1088
20. Varma M, Ray D (2007) Learning the discriminative power invariance trade-off. In: Proceedings of international conference on computer vision
21. Fukuda K, Takiguchi T, Ariki Y (2008) Multiple classifier based on fuzzy c-means for a flower image retrieval. In: Proceedings of international workshop on nonlinear circuits and signal processing, pp 76–79
22. Cho SY, Chi Z (2005) Generic evolution processing of data structures for image classification. IEEE Trans Knowl Data Eng 17(2):216–231

23. Goller C, Kuchler A (1996) Learning task-dependent distributed representations by back-propagation through structure. In: Proceedings of IEEE international conference on neural networks, pp 347–352
24. An-xiang H, Gang C, Jun-li L, Zhe-ru C, Dan Z (2004) A flower image retrieval method based on ROI feature. J Zhejiang Univ Sci, pp 764–722
25. Ming-Yu L, Oncel T, Ashok V, Rama C (2010) Fast directional chamfer matching. Comput Vision Pattern Recognit, pp 1696–1703
26. Comaniciu D, Ramesh V, Meer P (2003) Kernel-based object tracking. IEEE Trans Patt Anal Mach Intell 25(5):564–575
27. Ojala T, Pietikäinen M, Mäenpää T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Patt Anal Mach Intell 24(7):971–987
28. Ojala T, Valkealahti K, Oja E, Pietikäinen M (2001) Texture discrimination with multi-dimensional distributions of signed gray level differences. Pattern Recogn 34(3):727–739
29. Ning J, Zhang L, Zhang D, Wu C (2009) Robust object tracking using joint color-texture histogram. Int J Pattern Recognit Artif Intell 23(7):1245–1263