

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/319847034>

Mid-level Image Representation for Fruit Fly Identification (Diptera: Tephritidae)

Conference Paper · October 2017

DOI: 10.1109/eScience.2017.33

CITATIONS

2

READS

350

4 authors, including:



Sandra Avila

University of Campinas

60 PUBLICATIONS 909 CITATIONS

SEE PROFILE



Roberto Zucchi

Escola Superior de Agricultura "Luiz de Queiroz"

155 PUBLICATIONS 2,137 CITATIONS

SEE PROFILE



Fabio A. Faria

Universidade Federal de São Paulo

32 PUBLICATIONS 258 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Moscas-das-frutas (Diptera: Tephritidae) e índice de infestação em goiabas no norte de Minas Gerais. [View project](#)



Computação Aplicada ao Desenvolvimento de Métodos e Ferramentas para a Análise Forense de Documentos Digitais Utilizando Características de Iluminação [View project](#)

Mid-level Image Representation for Fruit Fly Identification (Diptera: Tephritidae)

Matheus Macedo*, Sandra Avila†, Roberto A. Zucchi‡, and Fabio A. Faria*

*GIBIS Lab., Institute of Science and Technology, Federal University of Sao Paulo, Sao Jose dos Campos, Brazil,

Email: matheus.macedo.leonardo@gmail.com and ffaria@unifesp.br

†RECOD Lab, Institute of Computing, University of Campinas, Campinas, Brazil,

Email: sandra@ic.unicamp.br

‡Luiz de Queiroz College of Agriculture, University of Sao Paulo, Piracicaba, Brazil

Email: rzucchi@usp.br

Abstract—Fruit flies are of huge biological and economic importance for the farming of different countries in the World, especially for Brazil. Brazil is the third largest fruit producer in the world with 44 million tons in 2016. The direct and indirect losses caused by fruit flies can exceed USD 2 billion, putting these pests as one of the biggest problems of the world agriculture. In Brazil, it is estimated that the economic losses directly related to production, the cost of pest control and in the loss of export markets, are between USD 120 and 200 million/year. The species of the genus *Anastrepha* are among the fruit flies economically important in the America tropics and subtropics with approximately 300 known species, of which 120 are recorded in Brazil. However, few species are economically important in Brazil and are considered pests of quarantine significance by regulatory agencies. In this sense, the development of automatic and semi-automatic tools for fruit fly species identification of the genus *Anastrepha* can assist the few existing specialists to reduce the insect analysis time and the economic losses related to these agricultural pests. We propose to apply mid-level image representations based on local descriptors for fruit fly identification tasks of three species of the genus *Anastrepha*. In our experiments, several local image descriptors based on keypoints and machine learning techniques have been studied for the target task. Furthermore, the proposed approaches have achieved excellent effectiveness results when compared with a state-of-art technique.

I. INTRODUCTION

The fruit flies belong to the Tephritidae family, which comprises approximately 5,000 species. They are distributed all over the world and several species are important agricultural pests. The damages are caused by the larvae that feed inside the fruit, making them unfit for consumption and commercialization. In addition to direct damage to fruit, some species of fruit flies are of quarantine importance, that is, they hamper the international market for fresh fruit. The country where the quarantine pest does not occur imposes customs barriers for the importation of commodities from the country, in which the pest is present.

There are few studies on economic losses caused by fruit flies. For example, it was estimated that economic loss caused by the Mediterranean fruit fly, *Ceratitidis capitata* (Wiedemann), an invasive species, was estimated at USD 242 million/year in Brazil [1]. Although economic losses caused by species of

Anastrepha, fruit flies native in Brazil, are also high, there are no economic analysis on these losses.

Among the economically important fruit flies in the Americas are the species of the *Anastrepha*. This genus is the most diverse in America tropics and subtropics with approximately 300 known species, of which 120 are recorded in Brazil [2]. However, few species are economically important in Brazil namely the South American fruit fly *Anastrepha fraterculus* (Wiedemann), the West India fruit fly *Anastrepha obliqua* (Macquart), and the guava fruit fly *Anastrepha striata* Schiner. These three species are considered pests of quarantine significance by many regulatory agencies.

Identification of species is a crucial step for the development studies on biology such as distribution, damage, quarantine, and control. The identification of *Anastrepha* species are based on wing pattern, and mostly on the aculeus (the piercing part of the female ovipositor). However, the species boundaries of some fruit fly complexes are difficult to be delimited. *Anastrepha fraterculus* is the most emblematic case of a cryptic species complex in the Americas, because it is a major pest only in some areas of its occurrence, which ranges from Mexico to northern Argentina [3]. Thus, misidentifications can be of serious problem for the implementation of quarantine restrictions, integrated pest management, and other control programs [4].

New procedures of insect identification have been used such as morphometric and molecular analyses to an accurate identification of fruit flies of the genus *Anastrepha* [5], [6]. However, Martineau et al. [7] showed that few works in the literature have been proposed to identify genus *Anastrepha* through image processing and machine learning techniques. Maybe this fact is related to the high similarity between species belonging that genus *Anastrepha*.

In this sense, this work aims to propose the use of mid-level representations based on local image descriptors for fruit fly identification. Furthermore, it compares the effectiveness results of different machine learning techniques using those representations to support the development of a real-time system for fruit fly identification of the genus *Anastrepha*. This system can be a good solution for a quick and precise

identification, reducing time and costs in performing and assisting the scarce number of experts in their tasks. Finally, the proposed approaches can be incorporated into other systems already existing in the literature.

The contributions of this work are:

- A comparative study among five feature detectors, eight local image descriptors based on keypoints and nine learning techniques;
- An improvement for two well-known local image descriptors based on keypoints (SIFT and SURF descriptors);
- The use of a mid-level representation approach based on BossaNova for fruit fly identification task of the genus *Anastrepha*;
- An effectiveness analysis among our proposal based on local image descriptors and the state-of-art approach existing in the literature;

The remainder of this paper is organized as follows. Section II presents works related to insect identification approaches existing in the literature. Section III describes different feature extraction techniques used in this works for insect identification task. Section IV shows the experimental protocol we devised to validate the work while Section V discusses the results. Finally, Section VI concludes the paper and points out future research directions.

II. RELATED WORK

Many works have been proposed to identify the most varied species of insects through image processing and machine learning techniques. Among proposed works, a semi-automated classification system called Digital Automated Identification SYstem (DAISY) classifies spiders, pollen grain, and butterfly through principal component analysis (PCA) [8]. In [9], SPIDA-web (SPecies IDentified Automatically) identifies Australian spiders to distinguish 121 species using Daubechies 4 wavelet function [10]. In another work, the ABIS (Automatic Bee Identification System) recognizes bee species of genus *Bombus*, *Colletes*, and *Andrena* by use of support vector machine (SVM) and kernel discriminate analysis techniques [11], [12]. In [13], an automated system to identify insects based on wing outlines (DAIIS) combines Elliptic Fourier coefficients and SVMs to classify a sample of 120 owlflies (Neuroptera: Ascalaphidae) [14]. The goal in [15] is to develop an insect recognition system through multiple-task sparse representation and multiple-kernel learning (MKL) techniques [16]. It combines various kind of visual properties such as color, texture, shape, scale-invariant feature transform (SIFT) and histogram of oriented gradients (HOG) features [17], [18].

Despite the biological and economic importance of the Tephritidae family (fruit fly), only two papers have been found in the literature [19], [20]. The first one adopted a successful framework of classifier selection and fusion [19], which combines several global image descriptors and machine learning techniques for a multimodal classification approach, using image of wings and aculei of three species of the *fraterculus* group: *A. fraterculus* (Wied.), *A. obliqua* (Macquart) and

A. sororcula Zucchi [20]. In the second work, the authors proposed a sparse representation using SIFT features densely sampled as input for two different machine learning techniques (multi-layer Max-pooling ScSPM [21] and linear SVM [11]). It has performed experiments with three unreported genus and twenty species [22].

III. MATERIALS AND METHODS

The typical image classification pipeline is composed of the three following steps: (i) local visual feature extraction, which extracts information directly from the image pixels, (ii) mid-level feature extraction, which makes the representation more general, aggregating abstraction to the model, and (iii) supervised classification, a machine learning technique allowing the extraction of a general model from the data.

A. Local Feature Extraction

Feature extraction is the first crucial step of image analysis procedures, aiming at extracting visual properties from certain regions of the image via pixel-level operations. According to the relative area of those regions, the extracted features can be roughly classified into *global* (one feature vector for the entire image) or *local* (several feature vectors per image, computed over relatively small regions of the image).

Local feature extraction usually includes two distinct steps [23]: feature detection and feature description. Feature detection consists in finding a set of interest points, or salient regions in the image that are invariant to a range of image transformations. Feature description consists in obtaining robust local descriptors from the detected features. In the following, we briefly introduce the detectors/descriptors evaluated in this work.

1) *Scale-Invariant Feature Transform (SIFT)*: Proposed by Lowe [18], this is the most well-known and widely used local descriptor for visual recognition tasks. SIFT is both a feature detector (based upon Differences-of-Gaussians, or DoG), and a feature descriptor. As a descriptor, it computes a histogram of gradient (HoG) locations and orientations. The resulting descriptor is 128-dimensional feature vector, which is invariant to scale, rotation, affine transformations, and partially invariant to illumination changes.

2) *Speeded Up Robust Features (SURF)*: It was proposed by Bay et al. [24] as an accelerated version of SIFT. SURF is also both a detector (based upon the determinant of the Hessian matrix, also known as Fast-Hessian feature detector) and descriptor. As a descriptor, it describes a distribution of Haar-wavelet responses within the interest point neighborhood. The SURF descriptor is based on similar properties of localized information and gradient distribution as SIFT, with a complexity stripped down even further. Only 64 dimensions are used, reducing the time for feature computation and matching, and increasing simultaneously the robustness.

3) *Binary Robust Independent Elementary Features (BRIEF)*: Presented in 2010 by Calonder et al. [25], BRIEF was the first binary descriptor published. It consists mainly of generating binary strings from simple pixel intensity value

comparisons over an image patch smoothed using a Gaussian kernel. The patches are usually obtained with the Fast-Hessian detector, but it is not limited only to the use of this feature detector. We employed the STAR detector, derived from CenSurE (Center Surround Extremas) detector [26] and FAST (Features from Accelerated Segment Test) detector [27]. The bit-length of the BRIEF descriptor are 128, 256 (default), or 512 and due to their correspondence in bytes they can also be referred as BRIEF-16, BRIEF-32 and BRIEF-644, respectively.

4) *Oriented FAST and Rotated BRIEF (ORB)*: As the name itself suggest, ORB [28] combines and extends on the concepts of FAST and BRIEF, reducing sensitivity to noise and having rotational invariance. The ORB detector is essentially a multi-scale FAST with orientation, while the ORB descriptor uses a learning process to determine the spatial arrangement of binary tests, decorrelating BRIEF features under rotational invariance. This makes the nearest neighbor search during matching less error-prone. The learning algorithm search for a set of 256 uncorrelated tests, which produce a 256 bit string, the ORB descriptor size.

5) *Binary Robust Invariant Scalable Keypoints (BRISK)*: Proposed by Leutenegger et al. [29], BRISK is a fast descriptor which uses symmetric sampling pattern (composed out of concentric rings) for intensity tests. The BRISK detector is based on the AGAST (Adaptive and Generic Accelerated Segment Test) detector [30], which is an extension of a faster performance version of the FAST detector. To describe the features, pairs of pixels around the interest point are separated into two subsets: short-distance and long-distance pairs. BRISK uses the long-distance pairs to estimate the patch orientation and the short-distance pairs to construct the descriptor itself through pixel intensity comparisons. BRISK descriptor is composed of a bit-string of length 512, i.e., a 64-dimensional feature vector.

6) *Fast Retina Keypoint (FREAK)*: Inspired by the human visual system, FREAK [31] uses a retinal sampled pattern for intensity tests. Similar to BRISK, FREAK applies the same AGAST feature detector. The FREAK descriptor is constructed by evaluating 43 weighted Gaussians at locations around the interest point, leading to 903 possible pairs. A learning algorithm similar to ORB is applied to finding the 512 most relevant pairs and build the FREAK bit string.

B. Mid-level Feature Extraction through BossaNova Approach

Mid-level feature extraction aims at transforming local descriptors into a global and richer image representation of intermediate complexity [32]. The standard pipeline to get mid-level features can be broken into two steps: coding and pooling. The *coding* step quantifies the local descriptors according to a visual dictionary of k visual words, which is usually built by clustering a set of local descriptors (e.g., k -means clustering algorithm). The *pooling* step aggregates the codes obtained into a single feature vector.

In the Bag of Visual Words (BoVW) [33], [34], the most popular mid-level image representation, the coding step asso-

ciates the local descriptors to the closest element in the visual dictionary (called hard-assignment coding), and the pooling takes the average of those codes (called average pooling). Since the pooling operation compacts all the information contained in the individually encoded local descriptors into a single feature vector, that step is critical for BoVW-based representations. In general, the objective of pooling is to summarize the information contained in the individually encoded descriptors into a single feature vector, thus preserving important information and discarding irrelevant details [35].

Over the years, BoVW representation has been extended to both steps of coding [36]–[38] and pooling [39]–[41]. Avila et al. [40] introduced the BossaNova mid-level image representation. It extends upon the BoVW method by using a density-based pooling function, which computes a histogram of distances between the local descriptors found in the image and each visual word, keeping more information than the BoVW during the pooling. In this work, we have adopted BossaNova as the basis of our framework since it has shown excellent results that overcome the state-of-the-art for image recognition tasks [42]–[44]. This is the first time that it is applied to fruit fly identification.

BossaNova brings several novelties for the mid-level image representation, a density function-based pooling strategy, a local soft coding schema, and also normalizations of the final feature vector (see [40] for more details). In our experiments, we kept the default BossaNova parameter values the same as in [40].

Figure 1 shows the main pipeline using BossaNova on insect identification task.

IV. EXPERIMENTAL SETTINGS

We now discuss the experimental setup, including the dataset and details on the image acquisition.

A. Dataset

1) *Fruit Flies Samples*: According to Martineau et al. [7], the systems of insect identification might be divided in three different categories based on image acquisition: (1) Lab-based samples, which entomologists bring to laboratory the target insects and manage the image acquisition system in a more controlled environment; (2) Field-based samples, which experts might capture the images directly in outdoor environments (cultivated fields), without any particular constraints and therefore a very more difficult scenario for computer vision approaches; and (3) Multi-individuals samples, which there are more than one single insect in the same image. In our experiments, we have used lab-based samples of the specimens *A. fraterculus*, *A. obliqua* and *A. sororcula* from the collection of the Instituto Biológico of São Paulo (e.g., Figures 2(a-c)). Specimens have been collected through McPhail-type traps (Figure 2b) and reared flies from fruit as well.

2) *Image Acquisition*: Figure 3 shows the employed three-step process for treatment and acquisition of wing image: (a) The right wing of each specimen is dissected; (b) It is mounted on a microscope slide with Euparal; (c) The slide is covered

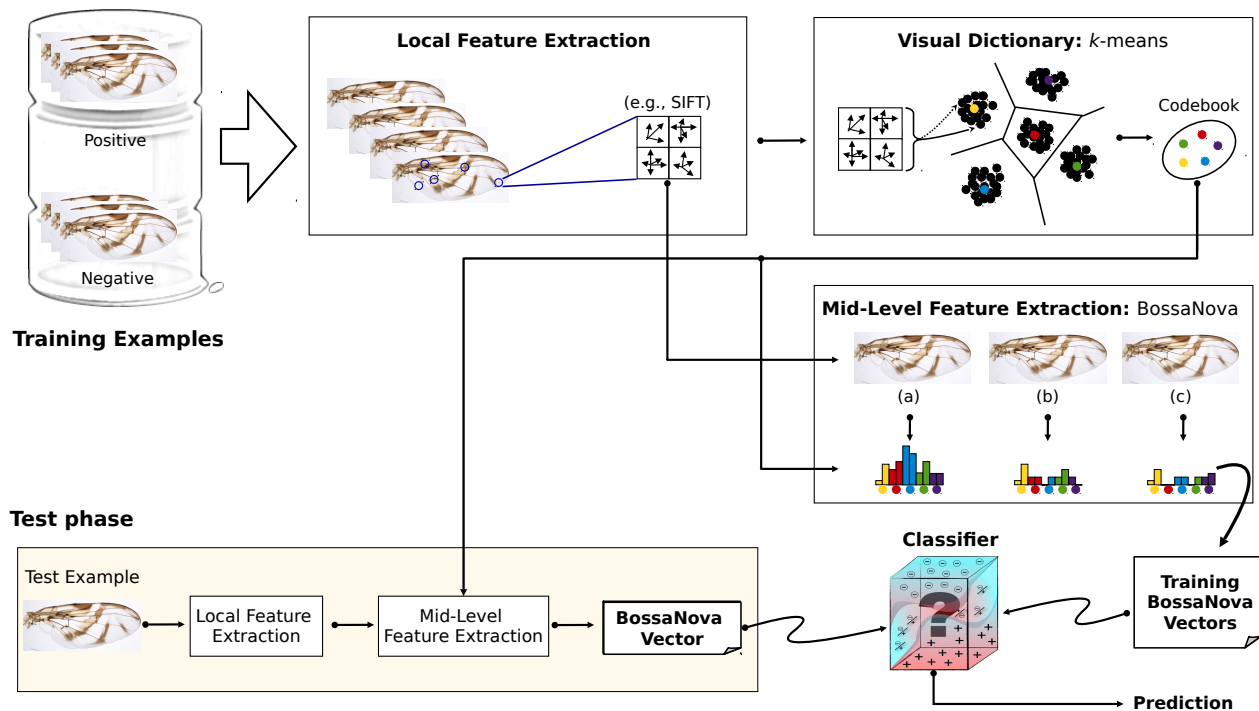


Fig. 1. The main pipeline of BossaNova. **Local Feature Extraction:** robust local descriptors (e.g., SIFT, SURF, BRIEF, ORB) are obtained from the detected features. **Mid-Level Feature Extraction:** BossaNova descriptors creates the feature vectors for the images using a visual dictionary (k -means with Euclidean distance is run over a sample local features, the final centroids are used as visual words). **Decision Model Training:** During the training-phase, the BossaNova vectors of annotated images are employed to train a decision model using a machine learning method. **Decision Model Prediction:** The trained model employs the BossaNova feature vectors of an image to predict on the positive or negative classes.

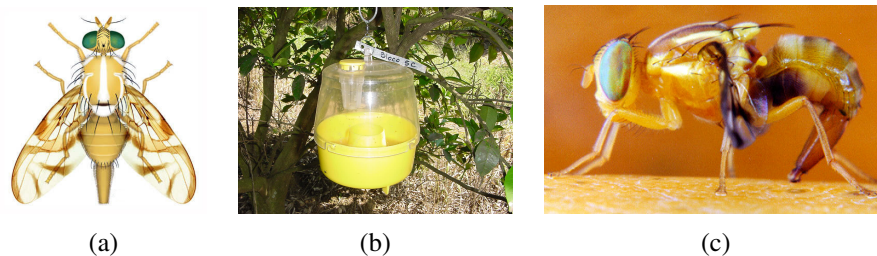


Fig. 2. (a) A fruit fly example (drawing) [45]; (b) a McPhail-type trap; and (c) a fruit fly laying eggs. Extracted from [20].

with a glass coverslip. The slides have been photographed with a Nikon DS-Fi1 camera (resolution 2560×1920) attached to a Nikon SMZ 1500 stereomicroscope (1.5X objective) [20].

The dataset used in this work is composed of 301 images and divided into three different categories: *A. fraterculus* (100), *A. obliqua* (101), and *A. sororcula* (100). It consists of pictures of specimens reared from samples of fruit trees in experimental and commercial orchards in the state of São Paulo, Brazil, stored in the Department of Entomology and Acarology ESALQ, Piracicaba, SP, Brazil and in the Biological Institute, Campinas, SP, Brazil.

B. Machine Learning Techniques

We have used nine different machine learning techniques: Multiple Layer Perceptron (MLP), Naïve Bayes (NB), Deci-

sion Tree (DT), Naïve Bayes Tree (NBT), k -Nearest Neighbor (kNN) with $k = \{1, 3, 5\}$, Simple Logistic (SL), and Support Vector Machine (SVM) using polynomial kernel.

The implementation of the machine learning techniques are available in the WEKA¹ data mining library. All machine learning techniques were used with default parameters which means we did not optimize them whatsoever.

V. RESULTS AND DISCUSSION

We have performed three different experiments with objective to support a system for fruit fly identification. In the first experiment, an analysis among five feature detection methods implemented in the OpenCV library [46] has been performed.

¹<http://www.cs.waikato.ac.nz/~ml/weka> (As of July, 2017).

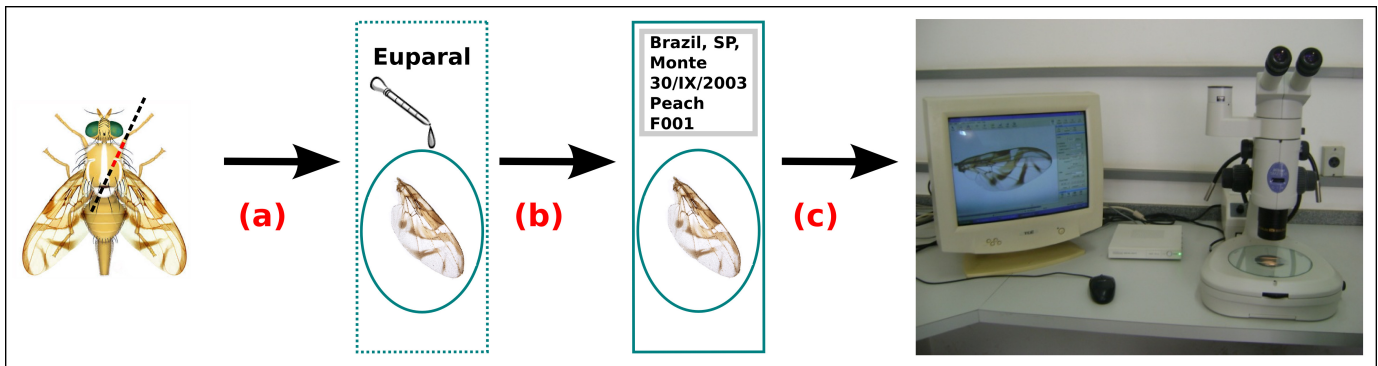


Fig. 3. The image acquisition process of wings. Extracted from [20].

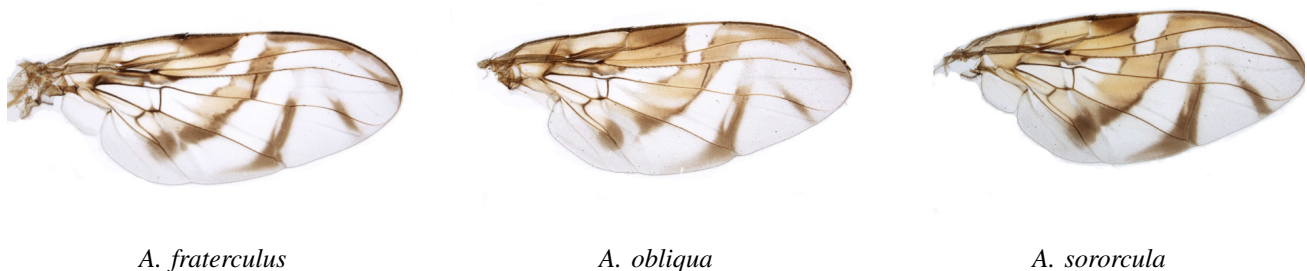


Fig. 4. Example of wings of each specie studied. Extracted from [20].

In the second experiment, a comparative study among six different local descriptors and nine learning techniques has been showed. Finally, in the third, a comparison among the best tuple (descriptor/learning technique) against the best image descriptor from the literature [47]. For all effectiveness experiments, the mean accuracies in the 5-fold cross-validation protocol have been computed.

A. Feature Detector Analysis

According to Francoy et al. [48], a classic technique used by specialists for insect identification (taxonomic classification), in the literature, is an approach based on geometric morphometric analysis. This approach computes the (x, y) coordinates average of the landmarks recorded on the insect wing vein intersections to obtain called reference configuration or tangent configuration [49]. This reference configuration is computed for all known species and used to classify new wing images.

In this sense, we believe that the quality of keypoint detection might be extremely important to the success of our proposed approach. Therefore, in this experiment, we have compared five different feature detectors, FAST [27], ORB [28], DoG (SIFT detector) [18], STAR [26], and Fast-Hessian (SURF detector) [50]. This experiment aims to verify the sensitivity of the feature detectors for the target dataset image. Recall that the implementation of the feature detectors is available in the OpenCV library.

Figure 5 shows an original image (see Figure 5-(a)) from the dataset and five examples of detected keypoints (dots in blue), one per each feature detector used in this work (See Figure 5-(b)–(e)).

As can be observed, in the Figure 5-(b), the FAST detector have detected many more keypoints than all feature detectors performed in this work. Accordingly, this detector has been adopted for our next experiments.

B. Effectiveness Analysis

In this section, we have performed two different experiments: (1) Comparative study among eight local descriptors based on keypoints (BRIEF, BRISK, FREAK, ORB, F-SIFT, F-SURF, SIFT, and SURF); and (2) Comparative study among nine learning techniques (MLP, NB, DT, NBT, kNN1, kNN3, kNN5, SL, and SVM).

Table I shows effectiveness results for all local descriptors and learning techniques. Taking into account the findings of the previous experiment, the BRIEF, BRISK, FREAK, ORB, F-SIFT and F-SURF local descriptors have been performed with FAST feature detector. SIFT and SURF descriptor used the original implementation, described in the Section III-A.

In the first experiment, we can observe that BRIEF and F-SIFT descriptors have achieved four of the best effectiveness results among nine released learning techniques (in blue). FREAK descriptor has achieved one best result using simple logistic (SL). Furthermore, we can observe that BRIEF

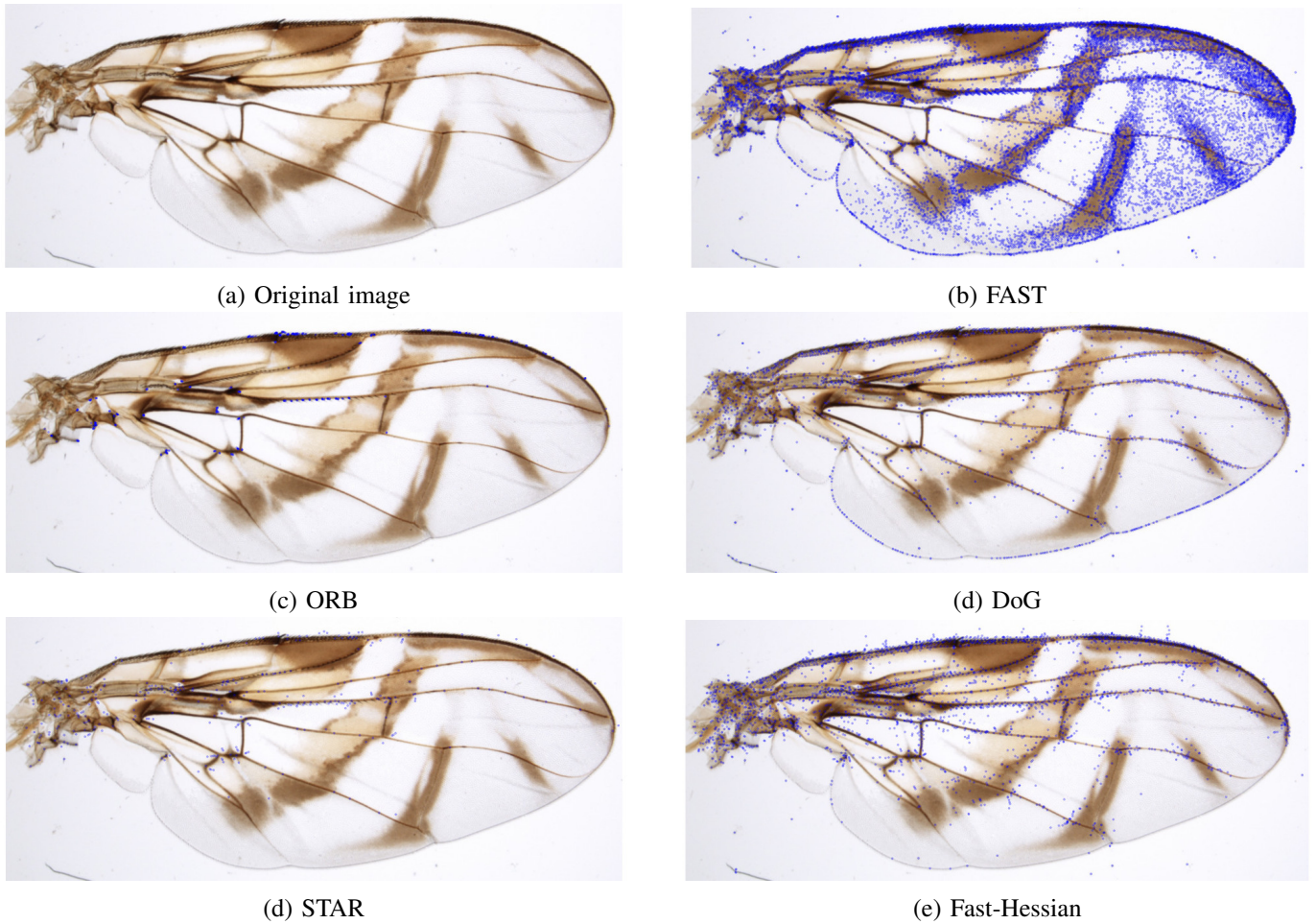


Fig. 5. Examples of keypoints detected by each feature detector for an original image from the dataset.

TABLE I

EFFECTIVENESS RESULTS (IN %) AMONG EIGHT LOCAL DESCRIPTORS AND NINE MACHINE LEARNING TECHNIQUES FOR A 5-FOLD CROSS-VALIDATION PROTOCOL. IN BLUE ARE THE BEST IMAGE DESCRIPTORS FOR EACH MACHINE LEARNING TECHNIQUE. IN GRAY CELL ARE THE BEST MACHINE LEARNING TECHNIQUES FOR EACH IMAGE DESCRIPTOR.

Descriptor	Machine Learning Techniques									Average	CI
	MLP	NB	DT	NBT	kNN1	kNN3	kNN5	SL	SVM		
BRIEF [25]	92.0	73.5	79.8	78.8	86.4	90.0	88.1	82.7	90.7	84.7	4.4
BRISK [29]	87.4	51.5	74.4	69.1	78.4	74.4	71.4	79.7	87.0	74.8	7.5
FREAK [31]	88.4	55.8	67.8	70.4	76.7	75.1	73.7	84.0	85.0	75.2	7.0
ORB [28]	90.0	61.8	75.1	74.4	86.7	84.4	82.4	82.1	90.4	80.8	6.3
SIFT [18]	84.7	53.8	62.5	61.8	67.2	68.4	68.5	75.1	84.4	69.6	7.1
SURF [24]	89.4	63.5	62.5	70.1	77.4	78.7	79.7	66.8	87.4	75.0	6.9
F-SIFT	94.7	62.1	76.7	82.1	87.4	85.1	84.4	82.4	93.7	83.2	6.7
F-SURF	84.7	49.2	65.8	66.1	76.1	74.1	72.1	80.4	83.4	72.4	7.7
Average	88.9	58.9	70.6	71.6	79.5	78.8	77.5	79.2	87.7		
CI	2.4	5.5	4.7	4.57	4.8	5.0	4.9	4.0	2.5		

descriptor achieved the best average accuracy (84.7%) with lower confidence interval (4.4).

In the second experiment, it possible to note that multilayer perceptron (MLP) technique has achieved seven of the best effectiveness results among eight local descriptors (in gray cell) released in this work. SVM technique has achieved one better effectiveness result with 90.4% of mean accuracy using ORB descriptor. In addition, MLP technique using F-SIFT descriptor was the best tuple (descriptor+learning technique) performed in this work with 94.7% of mean accuracy (in blue text and gray cell). Finally, we can verify that MLP and SVM techniques were the best learning techniques with average accuracy of 88.9% and 87.7%, respectively.

C. The Best Approaches

We also compared the best learning techniques with each local descriptor (rows in the Table I), BRIEF+MLP, BRISK+MLP, FREAK+MLP, ORB+SVM, SIFT+MLP, SURF+MLP, F-SIFT+MLP, and F-SURF+MLP. Furthermore, the baseline technique LCH+SVM proposed in [20] has been added in this experiment.

Figure 6 shows the effectiveness results among the best tuples (descriptor+learning technique) and the best baseline existing in the literature. Although F-SIFT+MLP (in blue) has achieved the best mean accuracy, when we compute the confidence interval with significance level of 0.05, it is possible to observe that there is no statistically significant difference among our seven approaches and the baseline from the literature LCH+SVM (in red). However, it is very important to note that the baseline achieved excellent effectiveness results by extracting color features from enhanced image (e.g., segmentation and dilation operations) [20]. Our approaches have been applied on the original images from the dataset. Therefore, our approaches might be used in real-time systems for insect identification tasks with no the use of any image enhancement operation.

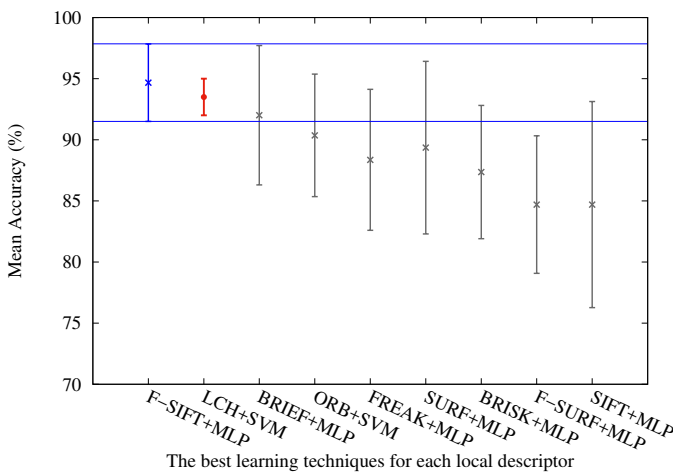


Fig. 6. Effectiveness results for each local image descriptor with 95% confidence interval (CI), i.e., a significance level of 0.05. In blue is MLP using BRIEF descriptor that has achieved the best mean accuracy.

VI. CONCLUSION

Fruit flies are of extreme importance for the world agriculture, especially in Brazil. Economic losses caused by some pests of the genus *Anastrepha* can exceed USD 2 billion, and in Brazil it is estimated between USD 120 and 200 million/year [51]. The species of the genus *Anastrepha* are the fruit flies economically important in the America tropics and subtropics with approximately 300 known species, which 120 are recorded in Brazil. However, few species are considered pests of quarantine significance by many regulatory agencies.

In this work, we proposed the use of a mid-level image representation approach for insect identification of three species of the genus *Anastrepha* using different local descriptors based on keypoints. Furthermore, we performed three robust analyses to support the development of a real-time system for fruit fly identification.

In the first, we performed an experimental analysis of the quality of keypoints detected through feature detectors. In this experiment, we analyzed the sensitivity of the detectors in the wing images. As findings of the experiment, we observed that the FAST detector achieved to detect many more keypoints on the image regions considered very important by specialists (wing vein intersections).

In the second, an effectiveness analysis among eight local descriptors and nine learning techniques was performed to verify the behavior of the tuples (descriptor+learning technique) in the fruit fly identification task. In this experiment, we observed that BRIEF and F-SIFT achieved the best results of mean accuracy among all of released local descriptors. Moreover, MLP and SVM techniques achieved to be the best learning techniques with higher average accuracy values and lower confidence interval.

Finally, in the third experiment, we compared the best learning techniques using each local descriptor to the best state-of-the-art baseline from the literature. In this experiment, we verified that even though there are statistical differences among our approach based on mid-level image representation and baseline, our approach might be directly applied on the original images with no require any enhancement operation. Therefore, we conclude that this work will support the development of a real-time system for fruit fly specie identification of the genus *Anastrepha*.

As future work, we intend to perform experiments with other local image descriptors, species, and learning techniques as classifier ensemble and deep learning approaches. Furthermore, a mobile system should be developed to assist the few experts from the biology area on their field works.

ACKNOWLEDGMENT

This work is partially financed by CNPq Universal Project (408919/2016-7). RECOD Lab. is partially supported by diverse projects and grants from FAPESP, CNPq, and CAPES. R.A.Z. is fellow of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

REFERENCES

- [1] C. M. Oliveira, A. M. Auad, S. M. Mendes, and M. R. Frizzas, "Economic impact of exotic insect pests in Brazilian agriculture," *Journal of Applied Entomology*, vol. 137, no. 1-2, pp. 1–15, 2013.
- [2] Zucchi, R. A., "Fruit flies in Brazil: *Anastrepha* species and their host plants and parasitoids," <http://www.lea.esalq.usp.br/anastrepha/>, 2008.
- [3] M. K. Schutze, M. Virgilio, A. Norrbom, and A. R. Clarke, "Tephritid integrative taxonomy: Where we are now, with a focus on the resolution of three tropical fruit fly species complexes," *Annual Review of Entomology*, vol. 62, no. 1, pp. 147–164, 2017.
- [4] B. A. McPherson, "Population genetics and cryptic species," *Area-wide Control of Fruit Flies and Other Insect Pests*, pp. 483–490, 2000.
- [5] Z. Bomfim, K. Lima, J. Silva, M. Costa, and R. Zucchi, "A morphometric and molecular study of *Anastrepha pickeli* Lima (Diptera: Tephritidae)," *Neotropical Entomology*, vol. 40, pp. 587–594, 2011.
- [6] —, "Morphometric and Molecular Characterization of *Anastrepha* Species in the spatulata Group (Diptera, Tephritidae)," *Annals of the Entomological Society of America*, vol. 5, pp. 893–901, 2014.
- [7] M. Martineau, D. Conte, R. Raveaux, I. Arnault, D. Munier, and G. Venturini, "A survey on image-based insect classification," *Pattern Recognition*, vol. 65, no. C, pp. 273–284, 2017.
- [8] A. T. Watson, M. A. O'Neill, and I. J. Kitching, "A qualitative study investigating automated identification of living macrolepidoptera using the digital automated identification system (daisy)," *Systematics & Biodiversity*, vol. 1, pp. 287–300, 2003.
- [9] K. Russell, M. Do, J. Huv, and N. Platnick, "Introducing spida-web: wavelets, neural networks and internet accessibility in an image-based automated identification system," *Systematics Association*, vol. 74, pp. 131–152, 2007.
- [10] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, Sep 1990.
- [11] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth Annual Workshop on Computational Learning Theory*, 1992, pp. 144–152.
- [12] T. Arbuckle, S. Schrder, V. Steinhage, and D. Wittmann, "Biodiversity informatics in action: identification and monitoring of bee species using abis," *International Symposium for Environmental Protection*, pp. 425–430, 2001.
- [13] H. P. Yang, C. S. Ma, H. Wen, Q. B. Zhan, and X. L. Wang, "A tool for developing an automatic insect identification system based on wing outlines," in *Scientific Reports*, vol. 5, 2015, p. vol. 1.
- [14] S. Chen, P. Lestrel, W. Kerr, and J. McColl, "Describing shape changes in the human mandible using elliptical fourier functions," *European Journal of Orthodontics*, vol. 22, no. 3, p. 205, 2000.
- [15] C. Xie, J. Zhang, R. Li, J. Li, P. Hong, J. Xia, and P. Chen, "Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning," *Computers and Electronics in Agriculture*, vol. 119, no. C, pp. 123–132, 2015.
- [16] M. Gonen and E. Alpaydin, "Multiple kernel learning algorithms," *Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, 2011.
- [17] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.
- [18] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [19] F. A. Faria, J. A. dos Santos, A. Rocha, and R. da S. Torres, "A framework for selection and fusion of pattern classifiers in multimedia recognition," *Pattern Recognition Letters*, vol. 39, pp. 52–64, 2014.
- [20] F. Faria, P. Perre, R. Zucchi, L. Jorge, T. Lewinsohn, A. Rocha, and R. da S. Torres, "Automatic identification of fruit flies (diptera: Tephritidae)," *Journal of Visual Communication and Image Representation*, vol. 25, no. 7, pp. 1516–1527, 2014.
- [21] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1794–1801.
- [22] A. Lu, X. Hou, C. Lin, and C.-L. Liu, "Insect species recognition using sparse representation," in *Proceedings of the British Machine Vision Conference*, 2010, pp. 1–10.
- [23] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [24] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [25] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in *European Conference on Computer Vision*, 2010, pp. 778–792.
- [26] M. Agrawal, K. Konolige, and M. R. Blas, "CenSurE: Center surround extremas for realtime feature detection and matching," in *European Conference on Computer Vision*, 2008, pp. 102–115.
- [27] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *European Conference on Computer Vision*, 2006, pp. 430–443.
- [28] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *IEEE International Conference on Computer Vision*, 2011, pp. 2564–2571.
- [29] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *IEEE International Conference on Computer Vision*, 2011, pp. 2548–2555.
- [30] E. Mair, G. Hager, D. Burschka, M. Suppa, and G. Hirzinger, "Adaptive and generic corner detection based on the accelerated segment test," in *European Conference on Computer Vision*, 2010, pp. 183–196.
- [31] A. Alahi, R. Ortiz, and P. Vanderghyest, "FREAK: Fast retina keypoint," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 510–517.
- [32] Y. Boureau, F. Bach, Y. LeCun, and J. Ponce, "Learning mid-level features for recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2559–2566.
- [33] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *IEEE International Conference on Computer Vision*, 2003, pp. 1470–1477.
- [34] G. Csurka, C. Bray, C. Dance, and L. Fan, "Visual categorization with bags of keypoints," in *European Conference on Computer Vision*, 2004, pp. 1–22.
- [35] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *International Conference on Machine Learning*, 2010, pp. 111–118.
- [36] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the fisher kernel for large-scale image classification," in *European Conference on Computer Vision*, 2010, pp. 143–156.
- [37] L. Liu, L. Wang, and X. Liu, "In defense of soft-assignment coding," in *International Conference on Computer Vision*, 2011, pp. 2486–2493.
- [38] H. Goh, N. Thome, M. Cord, and J.-H. Lim, "Unsupervised and supervised visual codes with restricted boltzmann machines," in *European Conference on Computer Vision*, 2012, pp. 298–311.
- [39] S. Avila, N. Thome, M. Cord, E. Valle, and A. de A. Araújo, "BOSSA: Extended BoW formalism for image classification," in *IEEE International Conference on Image Processing*, 2011, pp. 2909–2912.
- [40] —, "Pooling in image representation: the visual codeword point of view," *Computer Vision and Image Understanding*, vol. 117, no. 5, pp. 453–465, 2013.
- [41] Q. Wang, X. Deng, P. Li, and L. Zhang, "Ask the dictionary: Soft-assignment location-orientation pooling for image classification," in *International Conference on Image Processing*, 2015, pp. 4570–4574.
- [42] S. Avila, N. Thome, M. Cord, E. Valle, and A. de A. Araújo, "BossaNova at ImageCLEF 2012 flickr photo annotation task," in *Working Notes of the Conference and Labs of the Evaluation Forum*, 2012, pp. 1–6.
- [43] R. Pires, S. Avila, H. F. Jelinek, J. Wainer, E. Valle, and A. Rocha, "Automatic diabetic retinopathy detection using BossaNova representation," in *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 146–149.
- [44] C. Caetano, S. Avila, S. Guimarães, and A. de A. Araújo, "Pornography detection using BossaNova video descriptor," in *European Signal Processing Conference*, 2014, pp. 1681–1685.
- [45] Plantwise, "Empowering farmers, powering research — delivering improved food security," <http://www.plantwise.org/>, 2013.
- [46] G. Bradski, *Dr. Dobb's Journal of Software Tools*.
- [47] M. Swain and D. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.
- [48] T. M. Francoy, D. Wittmann, M. Drauschke, S. Müller, V. Steinhage, M. A. F. Bezerra-Laure, D. De Jong, and L. S. Gonçalves, "Identification of africanized honey bees through wing morphometrics: two fast and efficient procedures," *Apidologie*, vol. 39, no. 5, pp. 488–494, 2008.

- [49] F. J. Rohlf, A. Loy, M. Corti, and R. Kiltie, "Morphometric analysis of old world talpidae (mammalia, insectivora) using partial-warp scores," *Systematic Biology*, vol. 45, no. 3, p. 344, 1996.
- [50] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79–116, 1998.
- [51] R. A. Zucchi, A. Malavasi, A. S. Nascimento, and J. M. Walder, "Prejuízos das moscas-das-frutas na exportação de citros," *Revista Visão Agrícola*, no. 2, 2005.