22^{nd} *Computer Vision Winter Workshop*
Nicole M. Artner, Ines Janusch, Walter G. Kropatsch (eds.)
Retz, Austria, February 6–8, 2017

From trajectories to behaviors: an algorithm to track and describe dancing birds

Leonardo Oliva[1,2], Alessia Saggese[2], Nicole M. Artner[1], Walter G. Kropatsch[1], and Mario Vento[2]
[1]Pattern Recognition and Image Processing Group (PRIP), TU Wien, Austria
[2]Dept. of Information Eng., Electrical Eng. and Applied Mathematics (DIEM)
Faculty of Engineering, University of Salerno, Italy
loliva@prip.tuwien.ac.at

Abstract. *The movement of the males in some species of birds during the courtship determines the probability to be chosen by the female for mating. In order to analyze the behaviors of the birds, biologists nowadays perform a manual annotation of videos displaying the courtship phase. This a tedious and time consuming task. Thus, there is a strong interest in the development of algorithms able to automatically process these videos in order to analyze the behaviors. In this paper, we propose a novel approach able to track the movement of the males of a particular species of birds, namely the Golden Collared Manakin. Furthermore, we describe their mating dance by means of synthetic parameters, which are useful for the biologist. Both the tracking and the parameters used for describing the dance could be easily re-adapted to similar types of birds. The proposed approach has been tested on a set of videos from the biologists and the obtained results confirm the effectiveness of the proposed approach.*

1. Introduction

The Golden Collared Manakin (Manacus vitellinus) [3] is a small bird (in average less than five inches in length), which lives in the Panama forests. Males and females can be distinguished due to the color of their coat. The coat of the males shows a bright, yellow bloom (see Fig. 1), while the females are uniformly olive-green. These birds are well known for their particular and spectacular courtship displays. The male Golden Collared Manakins are a lekking species, as they form an aggregation of individuals who perform competitive displays. During the mating period, which lasts about 8 months, several males are visited by females at a precise locality, a small space on the ground (the so called court), which is cleared by the male and will become its focal point during these months. There is only one male per court. The males perform elaborate courtship displays that include rapid visual and loud acoustic signals produced by extraordinary physical movements involving extensive neuromuscular coordination. They alternately jump between two or more branches in their court. These little birds are a lekking species [1], so their elaborate displays are thought to be a result of a sexual selection: they want to attract the female as much as possible. However, little is known about differences in display between male Manakins, and how they relate to the success of the courtship. For these reasons, an automatic solution to extract significant parameters of the birds courtship from videos is needed. To achieve this, first we need a tracking system that is accurate enough to collect the positions of the male bird during its courtship dance. Until now biologists have been using small sensors to gather telemetry data from small birds [2, 23, 18] such as heart rate. Unfortunately, these sensors do not deliver the position of the birds

Figure 1. Male Golden Collared Manakin.

over time. A manual annotation of the bird's position would be tedious and time consuming for the biologist. For this reason, we propose a novel and efficient tracking approach of the moving birds, in particular of the male courtship dance. The data collected by the proposed tracking algorithm is then analyzed to extract parameters suitable to describe the dance of the birds. We propose to model the birds jumps by a parabolic motion, so that we can express the whole jump with a set of parameters that represents (1) the initial impulse that the bird puts for jumping, (2) the initial angle and (3) the range of the jump. There are some tracking methods in literature, for animals, such as mammals [16], but also for people [8]. Even though they work well in different situations, such tracking methods cannot be used directly for our problem since the bird changes its shape by opening its wings, so we can not use any shape features, also we can not use background subtraction only to build the foreground mask. As summarized in [26], there are several tracking algorithms for birds [21, 19] that use background subtraction methods to extract the objects of interest. These methods are meant for the tracking of birds that are flying in the sky, so the background is mostly static blue. In our case, Golden Collared Manakins are living in Panama forest, so besides the motion of the birds, there are also a lot of moving objects in the background, such as leaves and branches. Hence, we cannot rely on basic background subtraction only, because the moving background will introduce noise. There are even tracking techniques that make use of GPS devices [24], but these are not suitable for birds like Manakins since they would reduce their maximum life span [15]. Our approach combines several techniques that, together with contextual information, are capable to build a suitable foreground mask and successfully track the male birds. The foreground mask is built by using both Mixture of Gaussians and HSV thresholding, while the tracking process is done by employing a Kalman filter and similarity indexes. The remaining part of this paper is organized as follow: in Sec. 2 our method is presented, in Sec. 3 there are the evaluation results and finally in Sec. 4 the Conclusions.

2. The proposed approach

In this section, we describe the proposed approach in more detail. Section 2.1 is about the detection phase, which is necessary to deal with the challenging environment of the jungle in Panama, and the tracking of the bird frame by frame and through occlusions. In Section 2.2 the trajectories of the birds are analyzed to separate their jumps and to determine the order of the branches visited in the court.

2.1. Bird Tracking

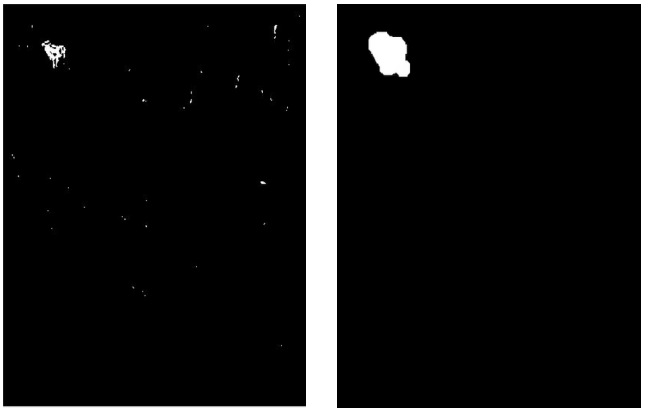
A way to build a tracker could be to model the relationship between the appearance of the target and its corresponding pixel values [14], so that we can estimate the position of the tracked object over the time. In general, a tracking algorithm should be robust against changes in the target object (pose, appearance, etc.), ambient illumination, noise (both from camera and from environment) and occlusions. In our case, the main challenges are the varying illumination (see Fig. 2), the ambient noise such as moving branches and leaves, motion blur and the deformations/articulations of the birds. To address these difficulties, we use a combination of foreground detection (detection phase) and Kalman tracking (tracking phase). The aim of the *detection phase* is to build a foreground mask with which the male bird can be segmented from the background. In the *tracking phase*, tracked objects are referred to as *track*, new objects extracted from the current frame are called *blob*, and I_M is the foreground of a frame after the application of the foreground mask. Let $\phi(I_M, t_{k-1}, b_k)$ be a function that associates the track t_{k-1} from the previous frame to a blob b_k of the current frame based on the current foreground I_M in frame k. A Kalman filter is employed to predict the position of track t_{k-1} into the next frame k [6]. To compute the association ϕ, a similarity matrix can be used, where the rows are the *tracks* and the columns are the *blobs*. Every element in this matrix gives the similarity between a blob b_k and the track t_{k-1}.

2.1.1 Detection phase

There are a lot of methods to build a foreground mask F_M: mean and median based [13, 5], mixture of Gaussians (MOG) [22], Kernel Density Estimators (KDE) [9], Eigenbackgrounds [17], Mean Shift based estimation [4], etc. Every method has proven to perform well in different situations [20], but looking at our environment, the MOG is suited best to model the background, since the environment shows slow movements of the branches and leaves. MOG can adapt to slow movement in the background. As stated in [22], MOG does not require prior knowledge about the environment. To simplify the task,

(a) (b)

Figure 2. Examples of different illumination conditions: (a) bright scene (b) dark scene.



(a) (b)

Figure 3. (a) Foreground mask F_{HSV} after HSV thresholding. (b) Final foreground mask F_M after morphological operations. The mask is computed on the frame showed in Fig. 2(a)

we can use the knowledge that the head of male bird of Golden Collared Manakin is very bright and saturated, while the courts are usually less saturated and darker. So, every pixel of the foreground mask is obtained as logical AND of two masks:

$$F_M(x,y) = F_{MOG}(x,y) \wedge F_{HSV}(x,y), \qquad (1)$$

where $F_{MOG}(x,y)$ are the pixel values of the foreground mask obtained with MOG, and F_{HSV} are the pixel values of the foreground mask obtained with HSV thresholding. We use an HSV representation for the frames and apply a threshold on the H and S bands. Since our object of interest has well-known color-characteristics [3], it is possible to restrictively threshold in both the H and the S band. Let $H_m(x,y), S_m(x,y), F_{HSV}(x,y)$ be the pixel values of the H band mask, the S band mask and the foreground mask of HSV, which is built as follows:

$$F_{HSV}(x,y) = \begin{cases} 1 & H_m(x,y) = 1 \wedge S_m(x,y) = 1 \\ 0 & otherwise \end{cases} \qquad (2)$$

Both $H_m(x,y)$ and $S_m(x,y)$ masks are obtained by thresholding the two bands with an upper and a lower values found empirically. The fixed thresholds are not a problem since, as said before, the bird has well+known color characteristics that does not change over the time. The result of this operation is shown in Fig. 3(a). Note that using HSV thresholding we are able to segment, and then track, the head of the bird. As expected, this mask is noisy, because of highly saturated parts in the environment. The noise could be reduced with morphological erosion. Nevertheless, false positive are expected in saturated scenes as in Fig. 2(b). Therefore, the MOG is applied on the original frame I. A mixture of two Gaussians is chosen, as the background is mostly bimodal (brown and green). In [22], it is stated that the initialization of MOG is an important step. For this, we compute the median image over 10 random frames, and repeat this process 10 times, obtaining 10 median frames that most likely will contain only background. These 10 median frames are used to initialize the background model, which is then updated according to [22]. F_{MOG} results from applying MOG and after applying Eq. 1 the foreground mask F_M is acquired. Then, morphological opening is applied to F_M to enlarge the blob and remove remaining noise. From this, the final foreground mask $F_M(x,y)$ results, as shown in Fig. 3(b).

2.1.2 Tracking phase

In the tracking phase, the task is to determine $\phi(I_M, t_{k-1}, b_k)$, the function that associates the track t in frame $k-1$ to the blob b in frame k, given a masked frame $I_M(x,y) = I(x,y) \wedge F_M(x,y)$, where I is the original frame. For this, a similarity measure is necessary to quantify the similarity between the blobs b and the tracks t, so that we can associate them. Note that this paper presents preliminary work on tracking the male bird. The following features are extracted from the region of the male bird:

- Normalized Histogram of H band, $f^h \in \mathbb{N}^{256}$
- Distance of the currently detected blob from the predicted position of Kalman filter, $f^d \in \mathbb{R}$

We chose Histogram instead of mean value so that we can have more values to compare.Shape features are excluded since the bird can change its shape significantly by opening its wings. Furthermore, features based on edges are not suitable, because the bird moves fast, which results in motion blur. For the similarity measure, let $i \in [1, ..., n]$ be the number of blobs and $j \in [1, ..., l]$ the number of tracks in frame k. The similarity is calculated as follows:

$$\sigma_{ij} = \sqrt{\alpha_h(\sigma_{ij}^h)^2 + \alpha_d(\sigma_{ij}^d)^2}, \tag{3}$$

where σ_{ij}^h is the histogram similarity and σ_{ij}^d is the distance similarity normalized to a maximum distance d_{max} found empirically. α_h and α_d are weight factors. For the experiments, these weight factors were set to 1 so that both features have the same influence, but for other applications these weights might be adapted.

$$\sigma_{ij}^h = 1 - X^2(f_i^h, f_j^h) \in [0, 1] \tag{4}$$

$$\sigma_{ij}^d = 1 - \frac{f^d}{d_{max}} \in [0, 1] \tag{5}$$

X^2 is the Chi-Square distance that is often employed to compare histograms [8, 7, 27, 25]. Finally, the assignment between a track and an object is done if the similarity index showed in Eq. 3 is bigger than a fixed number that could be tuned in function of the tracked object. Our experiments found that 0.9 is a good choice for this application.

2.2. Analysis and description of motion

In this section, we define parameters that could be useful for the biologists to characterize the courtship of a bird. We propose parameters that can describe the jumps of a bird in a compact way, without the need of storing all the positions of the birds trajectory. Furthermore, with these parameters, the biologists can relate the movements of the male with the movements of female.

The following parameters are used to describe the jumps and the environment (see Fig. 4 for a visualization of the parameters):

jumps $(MS_1, G_1, \alpha_1), ..., (MS_n, G_n, \alpha_n)$

frequency of jumps $jmpF \in \mathbb{R}$

branch sequence $B_1, B_2 \rightarrow B_2, B_4 \rightarrow B_k, B_j$

center of court (x_c, y_c)

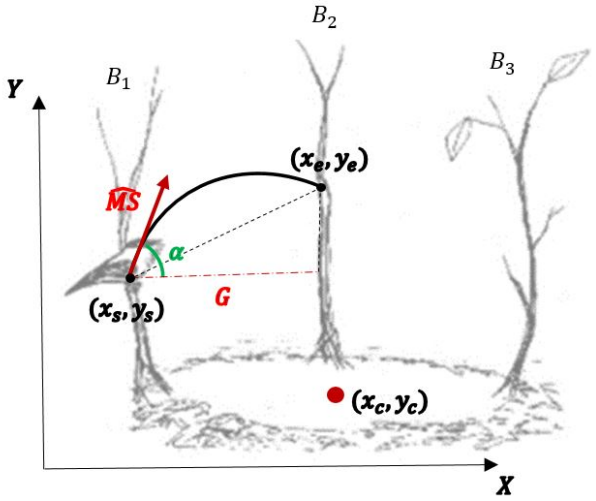


Figure 4. Computed parameters of a jump. ($\hat{MS}$ = Vector defined by the couple (MS, α))

The jumps of the birds are described by parabolas with the triplet (MS, G, α), where MS (Manakin Sprint) is the initial impulse, G the range and α the angle. To obtain the range G, the distance in X-direction is measured from start to end of the jump. Given the start point (x_s, y_s) and the end (x_e, y_e), G is computed as:

$$G = x_e - x_s \tag{6}$$

Note that in this way G also encodes the direction of the jump, so if $G < 0$ the jump is to the left, otherwise to the right. To separate the different jumps in the time domain, the presented approach searches for frames, where the bird is still (not moving). With this, the trajectory is split into segments (one for each jump) and the number of jumps is determined. The jump frequency is obtained using:

$$jmpF = \frac{no.\ of\ jumps}{elapsed\ time\ in\ minutes} \tag{7}$$

We use non-linear regression to fit a parabola to the trajectory data (2D positions over time). Since we are currently working with a single video camera, the real 3D motion is projected to the 2D image plane. It can be shown that under certain assumptions a parabola in the 3-D space projects onto a parabola in the image plane. The real 3D world can be described in a space XYZ, where XY coincide with the axis of image plane. Let's consider that a bird moves along a parabola in the space $X'Y'Z'$, where $X'Y'$ coincide with the plane of the parabola. In the most general case, the space $X'Y'Z'$ is rotated and translated with the respect to the XYZ space. The

general equation of a conic in $X'Y'Z'$ space is given by

$$ax^2 + bxy + cy^2 + dx + ey + f = 0. \quad (8)$$

To simplify the notation, the matrix form can be used:

$$p'^T C p' = 0, \quad (9)$$

where $C \in \mathbb{R}^{4x4}$ and $p' \in \mathbb{R}^{1x4}$ in homogeneous coordinates. Now the projection of any p' belonging to the conic on XYZ space can be found, that is

$$p = Rp' \rightarrow p' = R^{-1}p, \quad (10)$$

where $R \in \mathbb{R}^{4x4}$ is a matrix in homogeneous coordinates that takes into account both rotations and translations. Substituting Eq. 10 in Eq. 9 one obtains

$$p^T (R^{-1})^T C R^{-1} p = 0, \quad (11)$$

which is still a conic with the new coefficients matrix $C' = (R^{-1})^T C R^{-1}$. There are some degenerated configurations of motion, where the conic (see Eq. 11) is not a parabola. For example when the bird moves directly in the direction of the projection. Anyhow, for most configurations, where the point of view is almost orthogonal to the motion plane, we can consider that the discriminant equation $b'^2 - 4a'c'$ [11] is close to 0. Hence, we still obtain an approximation of a parabola. Note that this is valid if the space $X'Y'Z'$ is considered, where the motion is happening.

The equation of a parabola is used

$$y = \hat{a}x^2 + \hat{b}x + \hat{c} \quad (12)$$

to fit the collected motion data. An example of the obtained approximation is shown in Fig. 5.

In general it is possible to derive the initial impulse MS and the angle α from the equations of the parabolic motion. (Note that the fitting is done using positions from the tracking phase, which are in pixels.) Eq. 13 of parabolic motion uses quantities such as the gravity acceleration g that are $\frac{meters}{seconds^2}$, the horizontal component of velocity v_{0x} and the vertical component of velocity v_{0y} that are $\frac{meters}{seconds}$ and initial positions (x_0, y_0) that are in $meters$ as well. Since there is no calibration information available for our current video data, all information is expressed in pixels. Hence, it is not possible to directly use Eq. 13. We had to come up with a way to use the parameters

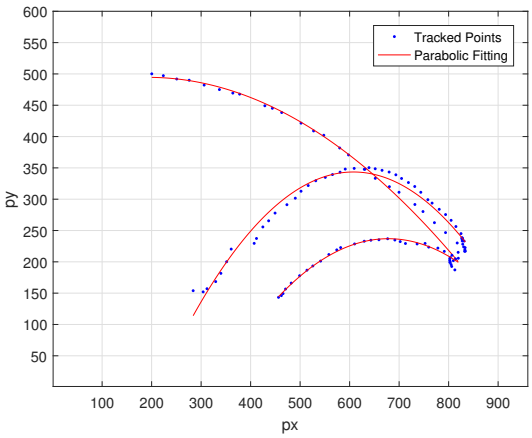


Figure 5. Fitted parabolas.

of the fitting to determine MS and α. From the equation of parabolic motion [12]:

$$y = -\frac{g}{2v_{0x}^2}x^2 + \left(\frac{gx_0}{v_{0x}^2} + \frac{v_{0y}}{v_{0x}}\right)x - \frac{gx_0}{v_{0x}^2} - \frac{v_{0y}}{v_{0x}}x_0 + y_0 \quad (13)$$

we can define

$$\begin{cases} MS = \frac{g}{v_{0x}^2} \in \mathbb{R} \\ \alpha = \arctan\left(\frac{v_{0y}}{v_{0x}}\right) \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \end{cases} \quad (14)$$

After that, we can solve the system using the parameters obtained from the fitting of the parabola:

$$\begin{cases} \hat{a} = -\frac{MS}{2} \\ \hat{b} = MSx_0 + \tan(\alpha) \end{cases} \quad (15)$$

Solving the system in Eq. 15 gives us MS and α. Note that using calibration parameters, we could do a fitting on real positions expressed in meters rather than in pixels, where this substitution is still valid. In fact, using calibration parameters, g is known to be $9.8\frac{m}{s^2}$, so we could directly use MS to derive v_{0x} by dividing it by g. To obtain the sequence of visited branches, a semi-automatic method is used. The user is asked to manually mark the branches the bird visits (see Fig. 6). When the bird stops, the minimum distance between the tracked position of the bird and all positions of the branches given by the user is determined to identify the branch on which the bird is sitting/resting. The result is shown in Fig. 6.

Finally, to obtain the center of the court, all trajectories of the bird in the video are analyzed. We propose to look for the bounding box enclosing all trajectories (all tracked positions). The center line along the y-axis approximates the geometric center of the court. Note that this is the center in the image plane. Based on the court center (see Fig. 7), it is

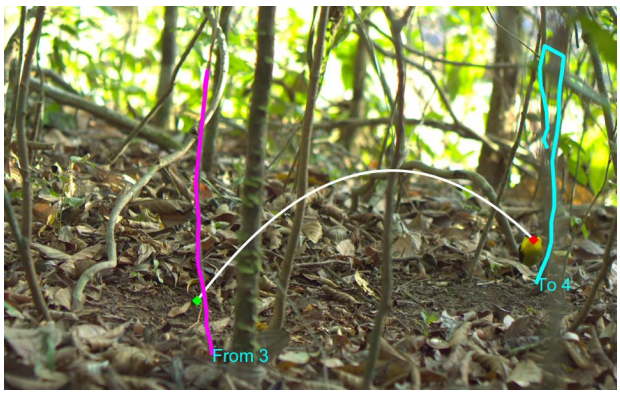


Figure 6. Parabola associated with starting (green point) and ending (red point) position of the bird's head. Manually annotated branches by the user.

Figure 7. Court center shown in a frame of a video (dashed, black line).

possible to normalize the positions of the bird, so that little variations of the camera position can be compensated between different recordings. This would allow the comparison of the courtship dance of different videos.

3. Experimental results

In this section, we report about the preliminary results achieved by the proposed approach. In particular, Sec. 3.1 gives a brief description of the data set, Sec. 3.2 shows the results achieved by the proposed tracking algorithm, while Sec. 3.3 presents the results of the description of the bird motion.

3.1. Data set description

Biologists provided a set of videos featuring the males dancing and other videos featuring both male and female. All videos were recorded using a camera with a Sony IMX174 CMOS sensor, which is capable of high speed video recording. The frame rate of the videos ranges from 99 to 120 fps. The original videos are uncompressed and have a size of 1920 $\times$ 1200. Due to computation time, tests are done on videos half the original size (960 $\times$ 600). The data set is composed of 6 videos with a total of 15066 frames and a total length of 8 minutes and 48 seconds. We chose the most different videos in our data set, to take in account of different scenarios as we can see in Fig. 2. The videos were manually annotated to generate the ground truth for the evaluation phase.

Table 1. Performance of tracking algorithm.

Video	Frames	Precision	Recall	**F1-Score**
1	5930	0.941	0.938	**0.940**
2	1263	0.927	0.924	**0.925**
3	889	0.846	0.849	**0.847**
4	2845	0.819	0.817	**0.818**
5	2405	0.901	0.887	**0.894**
6	1734	0.927	0.896	**0.904**

3.2. Evaluation of Tracking Algorithm

To evaluate the quality of the tracking algorithm, we compared our results with the ground truth by using the indices adopted within the Pascal VOC Challenge [10]. Let B_k^t and B_k^g be the bounding box found by tracking and the manually annotated ground truth bounding box respectively at frame k. The criterion to accept (true positive) a tracked object in frame k is

$$s = \frac{area(B_k^t \cap B_k^g)}{area(B_k^t \cup B_k^g)}. \tag{16}$$

With this, a bounding box is considered valid when $s > 0.5$, i.e. the bounding boxes are overlapping for more than 50% of their area. For every suitable bounding box, we consider the classic Precision, Recall and F1-Score, which is a harmonic mean between Precision and Recall. The obtained results are given in Tab. 1. The tracking algorithm performs better when the illumination conditions are darker, while it drops slightly when illumination conditions are brighter. This happens because of the thresholding on the S band: since we have more saturated elements in the scene, it is possible to acquire false positives. An example video of tracking can be found on our website[1]. The main goal of the tracking algorithms is to provide motion data for the description of the jumps. Therefore, the next section shows that the obtained positions are accurate enough for that purpose.

[1] http://www.prip.tuwien.ac.at/research/birds/TrackingOutput.mp4

Table 2. Performance of detecting the parabolas. (RJ = Real Jumps; DJ = Detected Jumps)

Video	RJ	DJ	**TP**	**FN**	**FP**	**Accuracy**
1	38	38	**38**	**0**	**0**	**1.00**
2	4	4	**4**	**0**	**0**	**1.00**
3	3	3	**3**	**0**	**0**	**1.00**
4	19	14	**14**	**5**	**0**	**0.73**
5	11	11	**11**	**0**	**0**	**1.00**
6	11	9	**9**	**2**	**0**	**0.81**

Table 3. Evaluation of the couple (MS^{-1}, α). (v = video; noj = number of jumps; μ = mean of (MS^{-1}, α); σ = standard deviation of (MS^{-1}, α); angles are in radians.

v	noj	from	to	μ	σ
1	12	B. 1	B. 2	(19.82, 3.39)	(6.03, 1.55)
	12	B. 2	B. 1	(29.71, -2.14)	(7.38, 0.69)
	3	B. 1	B. 4	(595.76, 0.44)	(80.90, 0.11)
	3	B. 4	B. 1	(565.28, -0.41)	(20.47, <0.1)
	2	B. 4	B. 2	(232.47, -1.10)	(<0.1, <0.1)
	5	B. 2	B. 4	(298.69, 0.67)	(24.35, 0.12
2	2	B. 3	B. 4	(220.86, 1.33)	(<0.1, <0.1)
	2	B. 4	B. 3	(279.09, -0.53)	(<0.1, <0.1)
4	2	B. 6	B. 2	(306.14, -0.52)	(65.86, 0.22)
5	2	B. 1	B. 3	(424.42, -0.27)	(30.22, <0.1)
	3	B. 2	B. 1	(64.19, 1.19)	(25.44, 0.56)
6	4	B. 1	B. 2	(254.11, -0.70)	(28.00, 0.13)

3.3. Evaluation of description of motion

In Table 2, the detected jumps are compared with the real jumps of the bird, which were annotated by hand. Furthermore, it reports about *True Positive (TP)*, *False Negative (FN)* and *False Positive (FP)* jumps for every video. The *Accuracy* was computed as percentage of correctly detected jumps:

$$Accuracy = \frac{Detected\ Jumps}{Real\ Jumps}. \qquad (17)$$

As stated before, the tracking is accurate enough to accomplish the detection of most jumps. In Table 3, it is shown that the couple (MS, α) is a reliable quantity to discriminate the jumps and to describe the dance of the birds. To undermine this, the mean and the standard deviation of (MS^{-1}, α) for jumps, from one branch to another branch, were measured. Note that we evaluate MS^{-1} because MS would be very small, since it has v_{0x}^2 as denominator, also we use only jumps from a branch to another that are, in number, superior to two, so that we can compute the mean and the standard deviation. As can be seen in Table 3, there is slight deviation between measurements of the couple (MS^{-1}, α). This is because the bird does not always jump in the same manner. Nevertheless, the mean values of (MS^{-1}, α) shows that it discriminates the different kinds of jumps. Opposite jumps, such as jump from B. 1 to B. 2 and then back again from B. 2 to B. 1, look similar by looking to the mean of MS^{-1} because the bird, on average, puts the same initial impulse to do them. While by looking to the mean of α it's clear that the angle encodes the direction of the jump, so opposite jumps has opposite sign. Also, there are angles that are outside the range $[-\frac{\pi}{2}, \frac{\pi}{2}]$ because the $\arctan$ function has the same values into the 1st and 3rd quadrant, and also the same into the 2nd and 4th quadrant. We are able to discriminate this cases with the sign of G. As a final test, we checked if the triplet is able to predict where the bird will land. For this, the final position of the bird (x_F, y_F) is computed with the time laws of the parabolic motion given by Eq. 13. This is compared with the final position given by the fitting. Our evaluation reported an error in pixels which was always in the range of $[0; 0.3]$ pixels. This result is coherent with the fact that, using the parameters and substituting them in Eq. 12, we can return to the same parabola that we obtained from fitting. Hence, the parameters (MS, G, α) are correctly describing the parabolic motion of the bird.

4. Conclusions

In this paper, a novel approach to track and describe the dance during courtships of male Golden Collared Manakins is proposed. Experimental evaluations on 15066 frames with different illumination conditions showed that the proposed tracking algorithm is able to track the bird with a high enough accuracy in the wild during its courtship dance. Situations in which there are highly saturated objects could degrade the performance of tracking, but this has a negligible effect on the detection of the jumps. The defined parameters are able to discriminate the different jumps of the bird. In addition, it is possible to determine the parameters (MS, G, α) from the tracking data, which describe the motion of the bird. Our approach is able to automatically extract the sequence of branches visited by the bird, which can be very time consuming to do by hand. The proposed parameters for the description of the motion are based on the physical model of parabolic motion. Hence, they can be used for other species that present similar behaviors. Currently, the proposed approach is only applied to 2-D images, and because of this we

only get approximative results. As future work, we plan to setup a multi-camera system to acquire the 3-D position of the birds.

References

[1] J. Barske, B. A. Schlinger, and L. Fusani. The presence of a female influences courtship performance of male manakins. *The Auk*, 132(3):594–603, 2015. 1

[2] I.-A. Bisson, L. K. Butler, T. J. Hayden, L. M. Romero, and M. C. Wikelski. No energetic cost of anthropogenic disturbance in a songbird. *Proceedings of the Royal Society of London B: Biological Sciences*, 276(1658):961–969, 2009. 1

[3] F. M. Chapman. *The courtship of Gould's manakin (Manacus vitellinus vitellinus) on Barro Colorado Island, canal zone*. American Museum of Natural History, 1935. 1, 3

[4] D. Comaniciu and P. Meer. Mean shift analysis and applications. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 1197–1203. IEEE, 1999. 2

[5] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting moving objects, ghosts, and shadows in video streams. *IEEE transactions on pattern analysis and machine intelligence*, 25(10):1337–1342, 2003. 2

[6] E. V. Cuevas, D. Zaldivar, and R. Rojas. Kalman filter for vision tracking. Technical Report B 05-12, Freie Universität Berlin, Institut fr Informatik, 2005. 2

[7] O. G. Cula and K. J. Dana. 3d texture recognition using bidirectional feature histograms. *International Journal of Computer Vision*, 59(1):33–60, 2004. 4

[8] R. Di Lascio, P. Foggia, G. Percannella, A. Saggese, and M. Vento. A real time algorithm for people tracking using contextual reasoning. *Computer Vision and Image Understanding*, 117(8):892–908, 2013. 2, 4

[9] A. Elgammal, D. Harwood, and L. Davis. Nonparametric model for background subtraction. In *European conference on computer vision*, pages 751–767. Springer, 2000. 2

[10] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010. 6

[11] J. R. Fanchi. *Math refresher for scientists and engineers*. John Wiley & Sons, 2006. 5

[12] G. Klimi. *Exterior Ballistics with Applications: Skydiving, Parachute Fall, Flying Fragments*. Xlibris, 2008. 5

[13] B. Lo and S. Velastin. Automatic congestion detection system for underground platforms. In *Intelligent Multimedia, Video and Speech Processing, 2001. Proceedings of 2001 International Symposium on*, pages 158–161. IEEE, 2001. 2

[14] E. Maggio and A. Cavallaro. *Video tracking: theory and practice*. John Wiley & Sons, 2011. 2

[15] G. Peniche, R. Vaughan-Higgins, I. Carter, A. Pocknell, D. Simpson, and A. Sainsbury. Long-term health effects of harness-mounted radio transmitters in red kites (milvus milvus) in england. *Veterinary Record-English Edition*, 169(12):311, 2011. 2

[16] P. Perner. Motion tracking of animals for behavior analysis. In *Visual Form 2001*, pages 779–786. Springer, 2001. 2

[17] J. Rymel, J. Renno, D. Greenhill, J. Orwell, and G. A. Jones. Adaptive eigen-backgrounds for object detection. In *Image Processing, 2004. ICIP'04. 2004 International Conference on*, volume 3, pages 1847–1850. IEEE, 2004. 2

[18] N. Sapir, M. Wikelski, M. D. McCue, B. Pinshow, and R. Nathan. Flight modes in migrating european bee-eaters: heart rate may indicate low metabolic rate during soaring and gliding. *PLoS One*, 5(11):e13956, 2010. 1

[19] K. Shafique and M. Shah. A noniterative greedy algorithm for multiframe point correspondence. *IEEE transactions on pattern analysis and machine intelligence*, 27(1):51–65, 2005. 2

[20] A. Sobral and A. Vacavant. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding*, 122:4–21, 2014. 2

[21] D. Song and Y. Xu. Monocular vision-based detection of a flying bird. Technical report, Citeseer, 2008. 2

[22] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2. IEEE, 1999. 2, 3

[23] S. S. Steiger, J. P. Kelley, W. W. Cochran, and M. Wikelski. Low metabolism and inactive lifestyle of a tropical rain forest bird investigated via heart-rate telemetry. *Physiological and Biochemical Zoology*, 82(5):580–589, 2009. 1

[24] E. W. Stienen, P. Desmet, B. Aelterman, W. Courtens, S. Feys, N. Vanermen, H. Verstraete, M. Van de Walle, K. Deneudt, F. Hernandez, et al. Gps tracking data of lesser black-backed gulls and herring gulls breeding at the southern north sea coast. *ZooKeys*, (555):115, 2016. 2

[25] M. Varma and A. Zisserman. A statistical approach to material classification using image patch exemplars. *IEEE transactions on pattern analysis and machine intelligence*, 31(11):2032–2047, 2009. 4

[26] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *Acm computing surveys (CSUR)*, 38(4):13, 2006. 2

[27] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *International journal of computer vision*, 73(2):213–238, 2007. 4