Automatic Lesser Kestrel's Gender Identification using Video Processing

Javier M. Mora-Merchan[1], Enrique Personal[1], Diego Francisco Larios[1],
Francisco Javier Molina[1], Juan Carlos Tejero[2] and Carlos Leon[1]
[1]*Escuela Politécnica Superior, University of Seville, Spain*
[2]*Escuela Técnica Superior de Ingeniería Informática, University of Malaga, Spain*
{

Keywords: Video Processing, Pattern Recognition, Animal Surveillance, Gender Classification.

Abstract: Traditionally, animal surveillance is a common task for biologists. However, this task is often accompanied by the inspection of huge amounts of video. In this sense, this paper proposes an automatic video processing algorithm to identify the gender of a kestrel species. It is based on optical flow and texture analysis. This algorithm makes it possible to identify the important information and therefore, minimizing the analysis time for biologists. Finally, to validate this algorithm, it has been tested against a set of videos, getting good classification results.

1 INTRODUCTION

Nowadays, it is easy to find in the literature a lot of work in which biologists follow new systems to gather information from natural environments, e.g (Larios et al., 2013a). In this sense, one of the main research lines is focused on animal behavior, such as: anurans (Luque et al., 2016), birds (Larios et al., 2013b), etc. Specifically for birds, there exists several studies focusing on distinguish between the different species (e.g. (Zottesso et al., 2016) which implements a bird classification using imaging processing over graphic representation of the audio spectrogram of their songs, or (Pang et al., 2014) that uses computer vision techniques for bird specie discrimination based on the difference in features of the birds' parts). However, in bird observation, it is typical to watch their inner nest activity, it being useful to distinguish between male and females behavior inside of it. Obviously, in order to make this distinction, it is necessary that there exist some kind of distinguishable visual features between them.

In the case of some birds, like the Lesser Kestrel (*Falco naumanni*), this difference can be found in its plumage (see Figure 1).

Specifically, as can be seen in Figure 1, the Lesser Kestrel is a small falcon. The male has a bluish gray head, uniform rusty back, but the breast and belly have black spots. It has uniform rusty scapulars, gray band on greater wings coverts and black primary feathers. Additionally, its tail is gray with

Figure 1: Lesser Kestrels (Gray, 2016) (two females and a male).

a black sub-terminal band. Conversely, female and younger ones have a more uniform appearance (typical in other common falcon), rusty with black barring and streaking, and being paler underneath.

These characteristics are typically exploited by biologists to distinguish visually between males and females of this falcon species more easily. However, breeding behavior study does not refer to an isolated identification. It requires a continuous video observation of the inner nest activity for each individual (distinguishing by gender). Obviously, it is a tedious task, especially for huge amount of videos, in which it is necessary to discard great amount of useless information, with hours of bird inactivity or directly with the empty nest. Additionally, this problem is accentuated for a colony, where this study must be repeated for several nests. Therefore, an automation of this video analysis is a great help for biologists allowing them to save time.

In this sense, computer vision is proposed as an excellent solution for this task. Proof of this fact is the algorithm proposed in this paper, which is able to identify the gender of cited bird specie through a background identification algorithm and a texture classification.

Specifically, this paper is organized as follows: Section 2 describes the proposed algorithm to Kestrel gender identification. A video analysis and its results are shown in Section 3 as a case study. Finally, Section 4 sums up the conclusions, final remarks and presents future work.

2 PROPOSED SOLUTION

As has been mentioned previously, there are two visual characteristics that allow experts to classify between male and female kestrels:

- ***Color Analysis:*** Male individuals show gray and reddish brown tones not present in the female individuals. In this sense, a hue histogram study would make gender identification possible.
- ***Texture Roughness Analysis:*** Male individuals have a plumage of plain colors while the females have one with black barring and streaking. It is translated into a more rough texture, or what is the same, with high frequency components.

The first of these options, in spite of being simpler, would require color images. However, color video sensors generally require a good illumination to work adequately. Unfortunately, the illumination in the nest is very limited and forces the use of monochromatic sensors (see Figure 2). This choice makes color analysis not useful and leads to texture analysis as the most suitable solution.

This is why we opted for texture analysis. Thus, the images are taken frontally from an elevated position. Due to this, it is possible to assume that the kestrels will be recorded with the wings folded, while they are inside the nest.

(a) Male (b) Female

Figure 2: Inner nest Lesser Kestrel video captures.

Based on these constraints, the proposed identification system has been divided into four stages; bird location, Kestrel's back extraction, energy analysis of different frequency bands and comparison with models in each of which the following tasks are performed:

2.1 Kestrel Location

In this first stage, the goal is to determine the bird position in each video frame. Assuming that each one is composed by a static background (the nest) and moving elements (the kestrels), the proposed strategy extracts the background and to estimate the movement centroid later.

In this sense, a basic background extraction algorithm consists of comparing each frame with the background (nest frame without any bird). Any difference between both would be considered as a movement. However, this approach would be very sensitive to changes at illumination (if a stable light source is not available) or camera noise.

As an alternative to this problem, an adaptive gaussian mixture model for background subtraction (Zivkovic and van der Heijden, 2006), (Zivkovic, 2004) was applied. It establishes a statistical model with the probability that each pixel of a frame is part of the background. This fact makes it possible that this model fits variation in lighting conditions (with slow dynamics), not being affected by moving objects (foreground, with faster dynamics).

Once all pixels with movement of the video frame (mobile pixels) are obtained (as part of a bird), the centroid calculation of them is the next step. It is calculated through the general moment expression (defined by equation 1), in which IMG is a binary image of the object (movement points) to be analyzed.

$$m_{ji} = \sum_x \sum_y (IMG(x,y) \cdot x^j \cdot y^i) \qquad (1)$$

Calculating first moment of area (m_{10} and m_{01}), and assuming that m_{00} is equal to the number of pixels associated with the movement, the centroid position is defined by $\bar{x}$ and $\bar{y}$ coordinates, both calculated according to:

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}} \qquad (2)$$

An example of this centroid can be seen in Figure 3, where it is on a male kestrel.

To improve the robustness, an additional constraint has been added to this estimation to validate it. Specifically, the number of mobile pixels (m_{00}) must be between two limits. On the one hand, the lower limit (n_L) make it possible to filter situations of low

Figure 3: Centroid of a movement.

movement or variations due to the sensor noise. On the other hand, the upper limit (n_H) filters situations in which the large part of a frame is motion (typical in the camera iris adjustment phenomena, when the luminance changes, or at the beginning of the analysis process). These limits were set to 0.5% and 25% of frame pixels respectively, using a 3σ approximation, which is based on a statistical study of mobile pixels over the valid image set (with $\overline{n} = 12.53\%$ and $\sigma = 4.02\%$). Out of these limits, the frame is discarded because it is considered unreliable.

2.2 Kestrel's Back Extraction

Once the kestrel position has been identified, the next step is to search a part of them in which there exist a clear difference between male and female texture. In this sense, based on the previous described Lesser Kestrel appearance, and assuming that the video is captured from an elevated position, the upper parts (mainly neck and back) are a good representation for this purpose (see Figures 1 and 2).

Thus, an alternative had been to apply an algorithm that accurately determined the kestrel shape around the calculated centroid (and even the existence of any bird around it). Instead, it has been empirically demonstrated that for this application, a square of 200 pixels centered on the centroid (2/7 of a frame approximately) captures the desired information accurately. It is possible mainly because of the depth of the nests is small and so that, the diminution in the kestrels size by perspective effects can be neglected. An example of this fact is can be seen in Figure 4, where representative parts of both genders have been correctly captured by the camera.

Obviously, this clipping technique may capture fragments of the nest wall and floor. H owever, as will be seen later, this fact does not significantly affect to the classification r esults. Additionally, in this sense, some models of images without any kestrel will

(a) Male (b) Female

Figure 4: Lesser Kestrel's back clip.

been included in the next classification stage. This fact makes it possible to strengthen this classification, discarding those regions in which there is suspicion of a bad detection.

2.3 Energy Analysis of Different Frequency Bands

The Two-Dimensional Fourier Transform of an image (FT) is an excellent tool to evaluate the importance of repetitive patterns in it. Focus on the module information ($|FT_{(u,v)}|$), each point of this transformation (defined by u, v) informs about three different aspects of these patterns:

- ***The distance to the origin*** ($\sqrt{u^2+v^2}$) depicts the spacial frequency value (higher distance implies higher frequency).
- ***The direction to the origin*** ($\arctan(v/u)$) indicates the orientation of the pattern.
- ***The FT value itself*** ($|FT_{(u,v)}|$) indicates pattern relevance in the original image.

Specifically, this work uses the Normalized Fourier Transform (NFT, see equation 3), which has the advantage of being invariant with linear changes in illumination (Nixon and Aguado, 2012).

$$NFT_{(u,v)} = \frac{FT_{(u,v)}}{\sqrt{\sum |FT_{(u,v)}|^2 - |FT_{(0,0)}|^2}} \qquad (3)$$

However, it must be taken into account that seeking patterns are not strictly regular, the barred females not always being at the same distance. Additionally, the direction of these is not constant and depends directly on the individual and its orientation.

Therefore, to minimize these problems, the proposed system considers adjacent frequency bands (or spectrum representation areas) instead of singular points in it. Furthermore, to make this analysis invariant against direction changes, an axial symmetry from the origin has been applied. It defines concentric annulus centered on the origin as representative

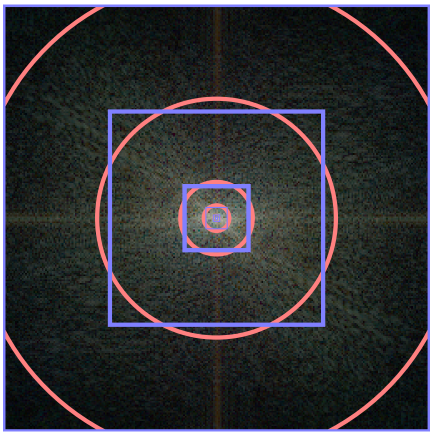

Figure 5: Spectrum representation and frequency bands; ideal (red) and real (blue).

regions for each frequency band (delimited areas by red circles in Figure 5), regardless of their orientation.

In this sense, the energy of a band (e, defined by equation 4) is an excellent descriptor to determine the importance (or how much presence does it have) of each frequency band.

$$e = \sum_{\forall(u,v)\in band} \left[NFT_{(u,v)}\right]^2 \tag{4}$$

In order to increase the system discrimination capacity, instead of focusing on the specific frequency band of relevant information (seek patterns), the analysis of multiple bands has been preferred. A profile of energy bands provides more information than a single value.

Additionally, the high correlation in gray values leads to higher energy values in low frequency bands (Theodoridis and Koutroumbas, 2008). Due to this, an exponential increase in the frequency band widths (or radius of red circles) has been chosen, in order to keep the energies in a similar order between them. Specifically, each decade has been divided into. Additionally, $NFT_{(0,0)}$ value has also been excluded, because this point represents the average gray value, which is irrelevant in the purposed texture analysis. As a last consideration, efficiency considerations suggest the substitution of circular bands limits (in red in Figure 5) by concentric squares with the same area (in blue in Figure 5). This change simplifies the computational cost of energy band calculation. Obviously, the modification results in directional distortion. The energy partially is now partially sensitive to texture orientation. However, that for the proposed application this distortion is not significant. Nevertheless, the better computational efficiency compensate this drawback.

Figure 6 shows an example of several energy profiles. As has been commented throughout this section,

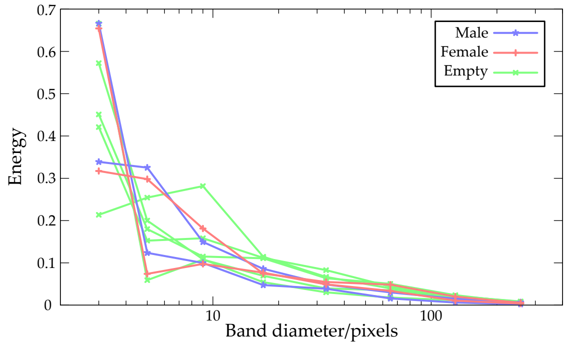

Figure 6: Examples of energy profiles for classes.

the result of this stage is an energy profile per frame, which will be used in the next classification state to distinguish the kestrel gender.

2.4 Classification

This final stage analyzes the energy profile of each frame and retrieves an estimation of the kestrel gender. In this sense, a k-Nearest Neighbors (k-NN) algorithm (Cover and Hart, 1967) implements this estimation. This method consists in comparing each energy profile under study with a set of reference profiles (or models) related to the different target classes. This classifier is a lazy supervised learning with an easy and efficient implementation. It can re-use erroneously classified data to improve the model performance.

For this algorithm, it is necessary to define a metric, which characterizes the likeness to each pattern profile. Based on energy distribution in profiles (see Figure 6), χ^2 (defined by Equation 5) is proposed as a metric. It represents the weighted euclidean distance with each pattern (m).

$$\chi^2_m = \sum_{i=bands} \frac{(e_{test,i} - e_{m,i})^2}{e_{m,i}} \tag{5}$$

where i represents each frequency bands. It is important to consider that this classifier offers an estimation per video frame where a movement has been detected. Therefore, to analyze a full video it is necessary to obtain a final estimation from the whole analyzed frames. This global classification uses a simple estimator, choosing the final result as the most frequent estimation of the frame analysis set. Notwithstanding, only a video is cataloged as a valid class, if the difference between the majority class and the rest of classes appearances is less than 50% of the total frames. Otherwise, the video is cataloged as “undetermined”. This simple analysis increases the robustness of the classification, because it can work correctly, even if some video frames are not clear and offer an incorrect classification.

3 STUDY CASE

As has been previously mentioned, the goal of this work is to facilitate the tedious tasks of viewing large sets of video. This help can be translated directly into two actions; automatic identification of useful videos (non-empty nest detection) and automatic determination of the adult bird gender.

For the test, a sample of 150 videos from the HORUS (Doñana Biological Station, 2009) project (recorded in monochrome, with a resolution of 704x576 px) was chosen. In this sense, following the two goals mentioned above, this study was divided into two parts:

3.1 Identification of Useful Videos

Table 1 shows the comparison of all cases whether the nest was empty or not, with the estimation made by proposed algorithm (localization of the kestrel, section 2.1).

The first information obtained from this analysis is the high percentage of empty nests (88%). This fact indicates that the automatic recording system does not work properly. These false detections are mainly due to the high noise presence in conditions of low luminosity.

In order to verify the quality of this estimation, the Yates' χ^2 test (Yates, 1934) (a correction in χ^2 of Pearson for cases with a low occurrence) has been applied.

$$\chi^2 = 82.3 \quad ; \quad df = 1 \quad ; \quad P < 0.001$$

These values indicate that the probability of obtained result was due to chance is less than 0.1%. Furthermore, the λ index of Goodman-Kruskal (Goodman and Kruskal, 1979) has been also calculated.

$$\lambda = 0.6 \quad ; \quad (\sigma = 0.1633)$$

This λ value indicates an improvement of 60% when this classifier is applied (instead of not applying any). Based on these studies, it is possible to say that the estimated probability of a correct prediction is 96%. Therefore, the theory that this analysis is suitable for this application was validated.

Table 1: Useful videos identification analysis.

		Estimated	
		Empty nest	Kept nest
Real	Empty nest	132	3
	Kept nest	3	12

3.2 Identification of Kestrel Gender

From the last study, it is easy to note that there are only 15 videos of kept nests in complete video set (150 samples). Due to this, only the kestrel gender algorithm evaluation have been used, minimizing the effects of high percentage of empty nest images, in front of the reduced number of cases in which the discrimination between male and female is done.

Thus, this small amount of elements for this analysis makes the showed results should be considered with a high error margin.. Table 2 shows the comparison between real and estimated result by the proposed kestrel gender identification algorithm (see sections 2.1–2.4). In this case, the classifier may determine that the kestrel's gender may be male, female, or undetermined.

The evaluation techniques have been the same as the previous section.

$$\chi^2 = 8.58 \quad ; \quad df = 2 \quad ; \quad P = 0.0137$$

Specifically, this first analysis indicate that the probability of obtained result was due to chance is less than 1.5%.

$$\lambda = 0.6 \quad ; \quad (\sigma = 0.2722)$$

For this study, λ value indicates an improvement of 66.6% when this classifier is applied, and the estimated probability of a correct prediction is 86.6%. Therefore, the theory that this analysis is suitable for this application was validated.

As in the previous section, both tests show that the proposed algorithm is suitable for this application.

3.3 Time Execution Analysis

As discussed previously, one of the main advantages of this application is to release the biologists from monotone and tedious observation. However, this is not the only advantage, its execution speed being another one.

Specifically, from Tables 3, it is easy to note that the systems have an analysis rate of 0.074 (less than

Table 2: Gender identification analysis.

		Estimated		
		Male	Undetermined	Female
Real	Male	4	1	1
	Female	0	2	7

Table 3: Processing time analysis.

Total time of videos	1800 sec.
Average processing time*	133 sec.

*Using single core of a Intel® Core™ i5-5200

1). While it can not explicitly guarantee real-time work. This makes it possible to organize a pipeline structure (see Figure 7), which can acquire the video and estimate it associated information (nest occupation and bird gender) in the same act.

In addition, this task segmentation is also easily scalable, so that the nest analysis can be performed at the same time, accelerating the analysis process even further. This fact improves the work quality of the biologist, who usually had to watch all the videos one by one.

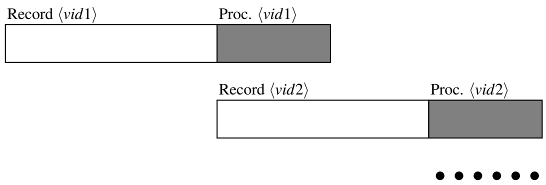

Figure 7: Pipeline structure for video processing.

4 CONCLUSIONS

As previously discussed, any tool that allows the biologist to reduce or facilitate monotonous observation tasks is useful in environmental monitoring. In this paper, video processing algorithm has been proposed for kestrel gender identification in a breeding environment (the nest). This algorithm has been tested over a video sample set, validating its correct operation for this application. In this sense, improvements in the ease and time analysis are directly obtained by biologists, allowing them to register bird activities automatically, without the need to inspect them directly. Thus, other improvements in storage needs can also be significant, being able to eliminate non-useful recordings (empty nest), typically abundant in this applications.

ACKNOWLEDGEMENTS

This work has been supported by the Consejería de Innovación, Ciencia y Empresa, Junta de Andalucía, Spain, through the excellence project eSAPIENS (reference number P10-TIC-5705). The authors would like to thank Javier Bustamente from Doñana Biological Station (CSIC) for his collaboration and support.

REFERENCES

Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27.

Doñana Biological Station (2009). HORUS project. http://horus.ebd.csic.es/wphorus/ [last visited 2017.01].

Goodman, L. A. and Kruskal, W. H. (1979). *Measures of Association for Cross Classifications*. Springer New York.

Gray, N. (2016). Male and female lesser kestrels.jpg. Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Male_and_female_Lesser_Kestrels.jpg [last visited 2017.01].

Larios, D. F., Barbancho, J., Sevillano, J., Rodríguez, G., Molina, F., Gasull, V., Mora-Merchan, J., and León, C. (2013a). Five years of designing wireless sensor networks in the doñana biological reserve (spain): An applications approach. *Sensors*, 13(9):12044–12069.

Larios, D. F., Rodríguez, C., Barbancho, J., Baena, M., Ángel, M., Marín, J., León, C., and Bustamante, J. (2013b). An automatic weighting system for wild animals based in an artificial neural network: How to weigh wild animals without causing stress. *Sensors*, 13(3):2862–2883.

Luque, J., Larios, D. F., Personal, E., Barbancho, J., and León, C. (2016). Evaluation of MPEG-7-based audio descriptors for animal voice recognition over wireless acoustic sensor networks. *Sensors*, 16(5):717.

Nixon, M. and Aguado, A. S. (2012). *Feature Extraction & Image Processing for Computer Vision*. Elsevier Science Publishing Co Inc.

Pang, C., Yao, H., and Sun, X. (2014). Discriminative features for bird species classification. In *Proceedings of International Conference on Internet Multimedia Computing and Service - ICIMCS14*. ACM Press.

Theodoridis, S. and Koutroumbas, K. (2008). *Pattern Recognition*. Elsevier LTD, Oxford.

Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test. *Supplement to the Journal of the Royal Statistical Society*, 1(2):217.

Zivkovic, Z. (2004). Improved adaptive gaussian mixture model for background subtraction. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*. Institute of Electrical and Electronics Engineers (IEEE).

Zivkovic, Z. and van der Heijden, F. (2006). Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 27(7):773–780.

Zottesso, R. H. D., Matsushita, G. H. G., Lucio, D. R., and Costa, Y. M. G. (2016). Automatic segmentation of audio signal in bird species identification. In *2016 35th International Conference of the Chilean Computer Science Society (SCCC)*. IEEE.