Salient features, combined detectors and image flipping: an approach to Haar cascades for recognising horses and other complex, deformable objects

Steve North
EASE
(Exeter Anthrozoology as Symbiotic Ethics) working group
University of Exeter, UK
s.north2@exeter.ac.uk

ABSTRACT

The author describes a new 'shortcut' approach to automatically detecting horses in still images and video: salient features, combining and flipping. Horses are complex, deformable (non-rigid) target objects with high levels of intra-class shape variability. A prototype Haar cascade detector was trained to detect what the author calls a 'salient feature'. This a distinctive, minimally changing physical attribute that is easily recognisable from multiple viewpoints. The detector's target object is: 'horse ears' and it only required a total training time of 91 minutes. It was evaluated in combination with an existing, 'asymmetric' detector (trained only to recognise right-facing horses). By combining the existing horse detector with the author's salient feature ears detector, the hit rate for true positives was increased by 50% (relative to the existing detector's performance). Flipping each test image (or video frame) around its vertical axis increased the hit rate by 83% (relative to the unflipped results) for the existing, asymmetric detector, when tested on an image dataset of horses facing in both directions.

Author Keywords

animal-computer interaction; horses; computer vision; machine learning; automated detection

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; H.5.1 Multimedia Information Systems (video); I.2.10 Vision and Scene Understanding (Motion & Video analysis); I.4.8 IMAGE PROCESSING AND COMPUTER VISION: Scene Analysis (Motion & Object recognition)

ACI2017, November 21–23, 2017, Milton Keynes, United Kingdom

ACM ISBN 978-1-4503-5364-9/17/11 $15.00
https://doi.org/10.1145/3152130.3152143

INTRODUCTION

Training computer vision detectors to recognise complex, non-human animals (hereafter referred to as 'animals') in still images and video, traditionally requires a large dataset of example images, and a great deal of time. In addition, such detection is often limited to recognising only the face of a target species. OpenCV provides example detectors for the faces of domestic species, such as cats and dogs [12].

This paper starts from the position that procedures for speeding up the meticulous workflow (currently required for detector training) might prove beneficial across a broad range of animal-related disciplines. For example: Animal-Computer Interaction (ACI), ethology, equitation science, veterinary science and anthrozoology.

It should be noted that, while this is fundamentally a technical paper, its specific intention is to contribute to the sum of knowledge for the field of ACI.

The problem of automated identification of animals is highly relevant to ACI [13][5]. In ACI (as in Human-Computer Interaction), the user's awareness of the interface is not always a prerequisite. An interface is anything which allows data to be transferred between the user (in our case, an animal) and the computer system. A system can respond to a user's behaviour, without the user consciously choosing to interact (a door opens automatically, a light comes on etc.). In the same way, an interface for animals may operate below their level of perception, but still create a conduit for inter-species communication.

Automated identification of animals provides a key building block for applications such as: tracking, automated behaviour detection and ubiquitous computing in animal housing environments. In turn, this may enable the development of more purposed systems for challenges such as: automated feeding and enrichment.

Looking forward, detection is likely to prove a fundamental component of future systems to support inter-species communication (in the sense that detection helps to 'translate' animal behaviour into human terms).

Why horses?

There are several reasons why horses are the subject of this work. Firstly, the author is very familiar with this species and has access to willing 'horse participants'. Secondly, much of the existing ACI work on detection relates to smaller companions that live in our houses (mainly cats and dogs). Thirdly, horses make for a challenging target object.

In their 2016 paper on detecting horses [17], Uddin & Akhi note that detecting this species proves a particularly challenging problem because:

"The size, colour and breeds are different…A horse is a non-rigid body. In other words, the shape and size of a horse varies greatly, and therefore the model of a horse is much more complex than that of rigid objects. Illumination and weather conditions vary greatly". [17]

A new 'shortcut' approach to the automatic detection of horses

In the following sections, this paper will suggest an approach to help streamline the automatic detection of horses in still images and video.

There are three elements to this:

- Identifying a species' salient feature(s)
- Combining Haar cascade detectors [18]
- Flipping images during detection to increase the efficiency of asymmetrically trained detectors

Detecting a salient feature - the advantages of a Haar Cascade which targets a distinctive, minimally changing physical attribute

Detecting the entire body of a horse (in addition to recognising the face and close-up elements of the anatomy) presents a challenge. Its difficulty, reflects the high level of intra-class shape variability (and lack of symmetrical viewpoints) found in quadruped species. Human animals are slightly easier to detect than many other animals. This is because our upright, two-legged posture reduces our complexity, when viewed from different angles.

It is suggested that, before starting to train a detector, time is spent studying the nature and anatomical distinctiveness of your target species. Are there aspects to the anatomy that remain visually consistent from many different viewpoints? Are these aspects also subject to very little change when the animal is in motion or undergoing deformation? Such an anatomical aspect is what the author intends by the term: 'salient feature'.

Building Haar detectors for one (or more) salient features, may reduce the need to train multiple detectors for all possible views of a complex animal.

'The ears': selecting a distinctive, salient horse feature undergoing minimal change from a variety of viewpoints

When looking for a distinctive anatomical feature in the horse, clearly identifiable from many viewpoints, the ears are the most obvious first choice.

From a computer vision perspective, they are a pair of vaguely vertical, long, thin flattened oval shapes. They might be described as approximately parallel, but their presentation frequently resembles a v-shape, with the separating distance closer at the base. A horse's ears are often (but not always) articulated in unison, because their functional anatomy and behavioural use determines it. If not in unison, their appearance, following individual ear movement, can still resembles a v-shape.

Figure 1 Horse ears as a salient feature

As they are capable of a wide-range of movement (to track sound), they often appear similar (as two objects pointing upwards), even when the horse is viewed from different vantage points (see Figure 1). As a veterinary / horse owner's textbook states:

"Each ear can be swiveled independently through 180 degrees, or laid back, shutting it off. Such mobility is achieved with 16 auricular muscles attached to the base of the pinna. Humans have only three such muscles, all of which are vestigial. Easily visible at the top of the head, a horse's ears are used to signal emotional state and intent...In response to directional sounds, a horse flicks an ear towards the source, or, if the sound is coming from the front, pricks both ears forward" [4].

Of course the horse may have other salient features suitable for detection, but the ears were chosen as an initial target to investigate the principle.

Combining detectors - the advantages of applying specialist Haar cascades concurrently

Although most detectors are trained to be 'specialists': face, full body front, eyes etc., the advantages of applying each sequentially seem to have been largely overlooked (or perhaps taken for granted?). There does not seem to be any literature or programming examples for combining Haar cascades. That said, there may be reasons that this has been avoided. Perhaps, there are concerns that the speed of detection might be affected by sequential detection on a single image / video frame? However, this might be somewhat mitigated by running the detection operations on separate, concurrent software 'threads'. The application of multiple detectors offers a clear advantage: recognising a target object that would otherwise have been missed. For most applications, it is unlikely that more than two or three detectors would be combined. Also, if the 'salient feature' approach described above is used, then the total number of combined detectors could be minimised.

Image and video frame vertical flipping – the advantages for detecting complex, deformable objects

When trying to automatically detect a horse (with its complex, deformable morphology), a standard approach might be to develop a Haar detector for each of the possible viewpoints. It is not possible to modify the underlying code of an existing, trained detector, so that it can recognise the same target object, when reflected around an axis.

Uddin & Akhi [17] describe a Haar detector for right-facing horses (from the viewer's perspective, when looking at the test image). For the purposes of this paper, this detector will be referred to as: 'horse_uddin_and_ayaz_2016_right_facing_side_view'. As will be demonstrated in later sections of this paper, this detector is completely unable to detect left-facing horses (even when these are exactly the same horses, in the same images that the detector is able to detect when right-facing).

Detectors similar to 'horse_uddin_and_ayaz_2016_right_facing_side_view' may be described as being 'asymmetric'.

In order to detect left-facing horses, a second detector would need to be trained, taking many hours, or possibly days. In this paper, it is proposed that an alternative approach is to simply flip the test image or video frame (using software), at the point of detection.

The vertical flipping of an image (in this context) describes a rotation around the vertical axis (see Figure 2). There is some confusion about the naming for image flipping operations. For example, does 'flip vertical' refer to the axis of rotation (vertical), or the plane that rotation takes place in (horizontal)? The OpenCV library for Python [14] (as used for the development discussed in this paper) describes an image rotation around the vertical axis as 'flip vertical' and so, for consistency with the software, that is how the term may be interpreted in this paper.

Figure 2 Flipping around the vertical axis

This approach is transferrable to any Haar detector that has been trained to detect only one reflection of a target object. With a simple software operation, it is possible to negate the need to train a second detector. If applying the combined detector method described above, this further limits the total number of detectors required.

TRAINING THE HAAR CASCADE DETECTOR: HABIT_HAAR_CASCADE_HORSE_EARS_1

Having identified the ears as a suitable salient feature for detection of horses, it was now decided to train a Haar cascade detector for this purpose.

The first step was to collect many examples of images containing this feature. Prior to commencing work on the ear detector, the author had already developed an extensive dataset of general horse images. This will now be briefly described.

Developing an extensive horse image dataset

'HABIT horses still image detector training dataset 10978 v2.0' [6] was developed to provide training images of horses for computer vision and machine learning applications.

It contains a total of 10,978 images, consisting of:

- 6183 x positive images of horses (showing many horse breeds, ages, genders, viewpoints, scales, occluded, multiple instances of target object etc.). The positive images were sourced from various places, including: the authors's own images, Google Image searches, ImageNet [2] and the Weizmann Horse Databases [1][15][16].
- 4795 x negative images (not showing horses). The negative images were sourced from ImageNet [2].

It should be noted that a further dataset was developed at this time. Whereas, dataset 10978 is intended for TRAINING horse detectors (in all poses and viewpoints, not just the ears), it is suggested that: 'HABIT horses still image detector testing dataset 200 v1.0' [9] is best reserved for TESTING prototype detectors.

To train the horse ear detector, a subset of 10978 (containing clear views of horses with ears) was prepared: 'HABIT horse ears still image detector training dataset 904 v1.0' [10]. Dataset 904 has a total of 904 images, of which 200 are positive and 704 are negative.

Training the 'horse ears' detector, using Python and OpenCV

Dataset 904 [10] was used to build and train the horse ears detector ('habit_haar_cascade_horse_ears_1' [7]), using Python [14] and OpenCV [12]. The time taken for the detector training to complete 11 stages (0-10) was: 1 hour 21 minutes and 7 seconds.

At the end of Stage 10, it was using 16 Haar features, with a HR of 1 and a FA rate of 0.163.

POS count: consumed 185 of 185

Neg count: acceptanceRatio 1000 : 8.96157e-006

STILL IMAGE EVALUATION OF THE HORSE EARS DETECTOR

The evaluation of the salient features, combined detectors and image flipping approach, required a software benchmarking tool. This software will now be introduced.

The detector benchmarking tool

'HABIT image detection using haar cascade files v1.0' [8] is a benchmarking and test utility for Haar detectors in the OpenCV-compatible XML format.

The app will automatically detect one or more XML cascade files placed in its 'cascades' directory. It will also detect all positive and negative images placed in the appropriate directories. An output data text file is saved in the data directory, with the results of the test.

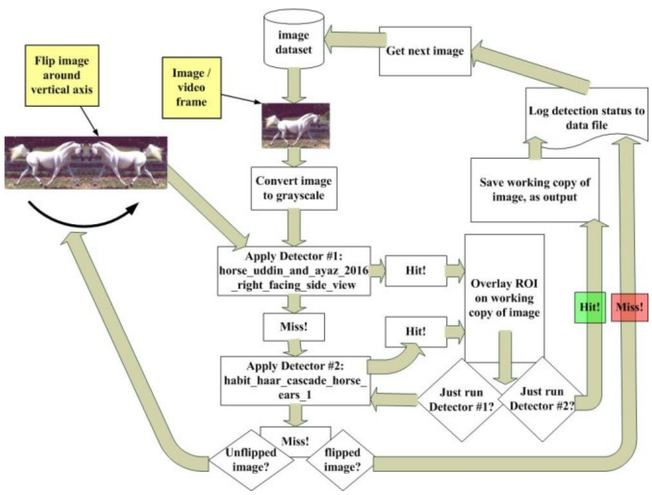

Figure 3 How the software processes the image dataset

Figure 3 illustrates the workflow coded within the benchmarking app, during the evaluation.

The datasets used in the evaluation

Two datasets were used for the purposes of this evaluation:

1. INRIA horses: a dataset for object class detection v1.03 [3]. This consists of 217 right-facing images of horses.

2. INRIA horses HABIT vertical flipped positives v1.0 [11]. This is the author's modified version of INRIA horses, where all of the right-facing images of horses (the 'positives') were: (i) duplicated (ii) flipped around the vertical axis and (iii) combined with the original, right-facing positive image set. This increases the number of positive images to: 434. There are still 223 negative images.

RESULTS

Table 1 shows the results from the evaluation run on 'HABIT haar cascade horse ears 1 v1.0' [7], using the app: 'HABIT image detection using haar cascade files v1.0' [8].

The first row of Table 1 shows the results for testing with the original, unmodified INRIA horses v1.03 dataset [3], with the benchmarking app configured for 'flipping' (see earlier). This was mode was left configured, as it is the default for the benchmarking app. As the horses in the unmodified INRIA horses v1.03 dataset are all right-facing, this should have had no impact on testing the asymmetric detector (for right-facing horses) 'horse_uddin_and_ayaz_2016_right_facing_side_view' [17]. Flipped or unflipped the results would have been unchanged.

The second and third rows show the results for the INRIA horses dataset (but with both left and right facing horse images) [11]. For row two, the benchmarking app was configured for 'unflipped'. Whereas, for row three, it was configured for flipped, the differences between rows two and three indicate whether the app flipping the images, is reflected in the results.

	Detectors							
	Detector #1: horse_uddin_and_ayaz_2016_right_facing_side_view [4]		Detector #2: HABIT haar cascade horse ears 1 v1.0 [15]		Combined detectors (unique)		Results (when combined with salient features detector)	
Image Datasets	%hits	%false alarms	%hits	%false alarms	%hits	%false alarms	% point increase on %hits for #1	% point increase on %false alarms for #1
INRIA v1.03 [19] WITH app flipping	24.88%	0%	15.67%	11.21%	35.94%	11.21%	11.06%	11.21%
INRIA horses HABIT vertical flipped positives [20] WITHOUT app flipping	12.44%	0%	9.68%	8.07%	21.19%	8.07%	8.75%	8.07%
INRIA horses HABIT vertical flipped positives WITH app flipping	22.81%	0%	14.52%	11.21%	34.33%	11.21%	11.52%	11.21%

Table 1. Results for evaluation of the HABIT haar cascade horse ears 1 v1.0 detector

In Table 1, the heading '%hits' indicates how many 'true positives' were reported. This relates to images of horses that were detected as horses. The heading '% false alarms' indicates how many 'false positives' were detected. This is images of something 'non-horse' (the negative test images) that were incorrectly identified as containing a horse.

The 'Combined detectors (unique)' section gives the new % hits rate when #2's (unique - not also detected by detector #1) %hits are added to detector #1's %hits. It presents the same for %false alarms. It should be noted that Table 1, column two, shows the total %hits and %false alarms for detector #2, not the net percentage of unique hits.

The right-hand side column ('Results') details the percentage point increases on the %hits and %false alarms. This is the values for the combine detectors (unique) column, minus the values for detector #1. This should not be confused with the overall percentage increase, relative to the original %hits values for detector #1 (not shown in Table 1, but calculated in the next subsection).

The impact of combining a salient feature Haar detector with one that is asymmetrically trained

In Table 1, row three shows the results for combining detector #1 (right-facing horses only) with the salient feature detector (#2), when applied to the dataset with both left and right facing horses ('INRIA horses dataset HABIT vertical flipped positives v1.0') - image flipping was enabled in the benchmarking app.

The difference between the %hits for detector #1 (22.81%) and the %hits for the combined detectors (34.33%) is an increase of 11.52 percentage points. The percentage increase between using just #1 and using #1 & #2 combined is:

11.52 / 22.81 = 0.50 * 100 = 50

The impact of image flipping

The difference for detector #1 ('horse_uddin_and_ayaz_2016_right_facing_side_view'), when applied to the dataset with both left and right facing horses ('INRIA horses dataset HABIT vertical flipped positives v1.0') with the benchmarking app configured first for 'unflipped' (row two) and then row 3 ('flipped) is now discussed.

The % hits for detector #1 increased by 10.37 percentage points (22.81% - 12.44%) between 'unflipped' and 'flipped'. The percentage increase of 'flipped' (relative to 'unflipped') is:

10.37 / 12.44 = 0.83 * 100 = 83%

Detector #1 seemed well-trained to avoid false positives (% false alarms was 0% in all three tests / rows). Combining detector #1 with #2 (which seems less reliable on false positives) did result in (for row three - 'INRIA horses HABIT vertical flipped positives v1.0' with app flipping) an increase of 11.21 percentage points for % false alarms.

Further work

It is intended to increase the pool of available detectors for horses. The horse ears detector ('HABIT haar cascade horse ears 1 v1.0') may need to be retrained to slightly reduce the % false alarms percentage and to improve detection against darker backgrounds. A more formal evaluation of this detector against a video dataset may be appropriate. The concept of salient feature detectors is an interesting area for further work and other detectors may be trained for horses, using this approach. In addition, there is more to be said about the methodology for identifying salient features in animals and other complex, non-rigid target objects.

CONCLUSION

Salient features, detector combining and image flipping have all been evaluated as possible methods to streamline the workflow of developing detectors for deformable, complex animals.

A salient feature detector (horse ears) combined with an existing detector increased the hit rate by 50%. Flipping each test images vertically increased the hit rate by 83% (relative to the 'unflipped' results). Knowledge contributions from this work were:

Recognising complex, deformable targets (in this case: the horse) by pre-identifying distinctive, salient features / 'recognition shortcuts' that undergo minimal change from a variety of viewpoints. For example: some anatomical details are clearly visible, irrespective of deformation, transformation, occlusion, context or viewpoint.

Combining several HAAR cascade detectors in such a way as to recognise complex, deformable targets (in this case: the horse) from a variety of viewpoints, including full-body and also isolated anatomical details (close up of leg, hoof, face etc.). Each detector focuses on recognising a different aspect of the target. When used in combination, detection becomes more reliable.

ACKNOWLEDGMENTS

The work described in this paper builds on an exploratory project funded by the EPSRC under Platform Grant 'Living with Digital Ubiquity ' reference: EP/M000877/1.

Special thanks to MS Uddin and AY Akhi for providing access to their horse detector, referred to in this paper as 'horse_uddin_and_ayaz_2016_right_facing_side_view' [17].

REFERENCES

1. E. Borenstein and S. Ullman. 2002. The Weizmann Horse Database. Retrieved 4 July 2017 from: http://www.msri.org/people/members/eranb/
2. Li Fei-Fei, Kai Li, Olga Russakovsky, Jonathan Krause, Jia Deng and Alex Berg. ImageNet: an image dataset. 2017. Retrieved 4 July 2017 from: http://image-net.org/
3. Vittorio Ferrari, Frederic Jurie and Cordelia Schmid. 2007. INRIA horses: a dataset for object class detection v1.03. Retrieved 4 July 2017 from: http://calvin.inf.ed.ac.uk/datasets/inria-horses/
4. J. McEwen. 2000. Ultimate Horse Care: The Complete Veterinary Guide. Wiley. 9781582451701
5. Steve North, Carol Hall, Amanda Roshier and Clara Mancini. 2015. HABIT: Horse Automated Behaviour Identification Tool – A Position Paper. In Proceedings of the proceedings of ACI@BHCI (Animal Computer Interaction Workshop), British HCI 2015 (Lincoln. UK. 13 July 2015), BCS, UK. http://dx.doi.org/10.13140/RG.2.1.3395.0881
6. Steve North. 2017. HABIT horses still image detector training dataset 10978 v2.0. Available from: http://dx.doi.org/10.5281/zenodo.822849
7. Steve North. 2017. Software Program: HABIT Haar cascade horse ears detector 1 v1.0.doi:10.5281/zenodo.823354. Available from: https://doi.org/10.5281/zenodo.823354
8. Steve North. 2017. Software Program: HABIT image detection using haar cascade files v1.0.doi:10.5281/zenodo.824260. Available from: https://doi.org/10.5281/zenodo.824260
9. Steve North. 2017. HABIT horses still image detector testing dataset 200 v1.0. http://dx.doi.org/10.5281/zenodo.822853.
10. Steve North. 2017. HABIT horse ears still image detector training dataset 904 v1.0. Available from: http://dx.doi.org/10.5281/zenodo.823168
11. Steve North. 2017. INRIA horses v1.03 HABIT vertical flipped positives dataset v1.0. Available from: http://dx.doi.org/10.5281/zenodo.824826
12. OpenCV team. 2017. Software Program: OpenCV (Open Source Computer Vision Library). Available from: http://opencv.org
13. Patricia Pons, Javier Jaen and Alejandro Catala. 2017. Assessing machine learning classifiers for the detection of animals' behavior using depth-based tracking. Expert Systems with Applications 86(11/15/). 235-246. http://dx.doi.org/https://doi.org/10.1016/j.eswa.2017.05.063
14. Python Software Foundation. 2017. Software Program: Python - programming language. Available from: https://www.python.org/
15. Jamie Shotton. 2007. Multiscale Weizmann Horse Detection Dataset. Retrieved 4 July 2017 from: http://jamie.shotton.org/work/data.html
16. Jamie Shotton. 2007. Single-scale Weizmann Horse Detection Dataset. Retrieved 4 July 2017 from: http://jamie.shotton.org/work/data.html
17. MS Uddin and AY Akhi. 2016. Horse Detection Using Haar Like Features. International Journal of Computer Theory and Engineering. 8. 5. 415
18. P. Viola and M. Jones. 2001. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001.511-518. http://dx.doi.org/10.1109/CVPR.2001.990517