



## Multi-scale Reflection Invariance

Henderson, C; Izquierdo, E; SAI Computing

To be published by SAI Computing

For additional information about this publication click this link.

<http://qmro.qmul.ac.uk/xmlui/handle/123456789/13049>

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact [scholarlycommunications@qmul.ac.uk](mailto:scholarlycommunications@qmul.ac.uk)

# Multi-scale Reflection Invariance

Craig Henderson, Ebroul Izquierdo  
Multimedia and Vision Lab  
Queen Mary University of London  
Email: {c.d.m.henderson, ebroul.izquierdo}@qmul.ac.uk

**Abstract**—In this position paper, we consider the state of computer vision research with respect to invariance to the horizontal orientation of an image – what we term *reflection invariance*. We describe why we consider reflection invariance to be an important property and provide evidence where the absence of this invariance produces surprising inconsistencies in state-of-the-art systems. We demonstrate inconsistencies in methods of object detection and scene classification when they are presented with images and the horizontal mirror of those images. Finally, we examine where some of the invariances are exhibited in feature detection and descriptors, and make a case for future consideration of reflection invariance as a measure of quality in computer vision algorithms.

**Keywords**—*reflection invariance, image orientation, mirror symmetry, object detection, image description*

## I. INTRODUCTION

Human perception is generally invariant to horizontal reflection with respect to recognising objects and scenes as if looking in a mirror. We observe that computer vision algorithms are more sensitive to the reflection of an image and that invariance to this has not received any attention in contemporary research. In this position paper, we introduce a property of *reflection invariance*, specifically studying horizontal reflection as an introduction to the concept, although discussion is appropriate for general reflection about alternatives lines of symmetry.

We suggest *reflection invariance* as an important property in designing and implementing algorithms, and metric in measuring the success of vision algorithms and applications. Just as scale invariance seeks to neutralise the size of a feature to avoid bias in scale, we propose *reflection invariance* to avoid bias in mirror reflection about an arbitrary axis. It is important that algorithms should be consistent in applications such as object recognition and scene classification, and we demonstrate that current state-of-the-art methods do not exhibit consistency when an image is reflected horizontally.

The rest of the paper is organised as follows. In Section II we describe our idea of *reflection invariance* in the context of other popular invariance measures, and describe its relevance to key areas of computer vision research in Section III. Section IV briefly suggests some root causes of invariance and we conclude our findings in Section V.

## II. ORIENTATION AND REFLECTION

Low-level keypoint features describe a neighbourhood of a few pixels, where the co-location of pixel intensities is an important attribute used to describe the feature. Most feature descriptors, including the most popular SIFT [1] and HoG [2],

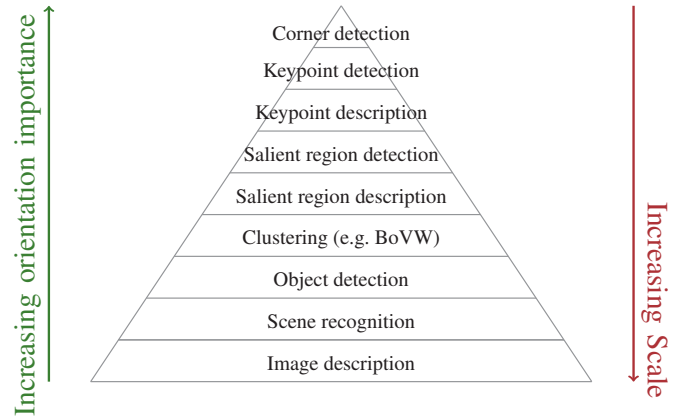


Fig. 1: Pyramid of Scales and Orientation Significance: as the scale increases, the importance of orientation diminishes

use the orientation of pixel gradients in a colour space or channel in some way to detect and represent distinct feature characteristics. These algorithms are inherently sensitive to orientation, however others are sensitive only in practice, caused by poor implementation choices and mathematical rounding errors that accumulate to affect the result and cause dependence on image orientation.

A collection of descriptors can be composed to describe a distinctive pattern or region, such as in the popular *Bag of Visual Words* method [3]. In such a collection, the orientation of individual features *relative to each other* is important, but the orientation of the collection as a whole is less significant. As the scale of description increases further, orientation becomes less important and indeed becomes a limitation when considering high-level features in an image. The significance of orientation can therefore be considered inversely proportional to the scale of description, with its influence diminishing with the increase in distance from the pixel detail (Figure 1).

Reflection has the same scale of sensitivity as rotational-orientation. Consider an example of scene recognition. A human would describe a city-scape scene, and identify a familiar city regardless of the horizontal reflection of the image; if the image is reflected about its vertical centre, this mirrored image would still be recognisable to a human and would not influence their description or identification. Computer vision algorithms, however, are more sensitive and often produce different results for these images.

The challenge is to generalize the description as the scale increases, with orientation becoming less relevant to the point where it is irrelevant at image scale.

### III. REFLECTION SENSITIVITY IN STATE-OF-THE-ART METHODS

#### A. Low level features

Feature detectors fulfil the common need to identify interest points within an image. Information at these positions is extracted into a *descriptor* – a fixed length vector of numeric or binary values – that can be used, for example, to match similar features in applications such as image retrieval, alignment, stitching, and classification.

Many research papers combine the two stages of detection and description into a single step, but each are independent. The invariance properties of detectors and descriptors are important, and in work to date are consistent. An algorithm that provides for feature detection and feature description can provide invariance to scale, rotation, illumination or affine regions in both steps.

In considering invariance to horizontal reflection, we assess the two separately and propose that it is not necessary – or even desirable – for a method to be consistent in a reflection invariance in detection and description. The goal of feature detection is to find keypoints or regions in an image that contain *interesting* information. The definition of *interesting* is specific to the goal of the detector, but it is reasonable to expect that a location that is *interesting* in an image should also be *interesting* in the same image that is horizontally reflected.

*Feature detectors* To be reflection invariant, a feature detector must show that the set of keypoints or regions found in an image are equivalent to those found in the a mirror reflection of the image [4]. In that study, an analysis of feature detectors with respect to reflection invariance concluded that corner detectors are stable, and the most popular detectors SIFT and SURF are very unstable in detecting consistent feature points in images and their mirror reflections (Table I).

*Feature descriptors* Conversely, the orientation of a feature is an important and discriminating attribute, and extracted descriptors should generally maintain local orientation so that established methods of feature matching, for example, can accurately measure the magnitude and position of a feature vector in high-dimensional space. However, reflection invariance in low-level descriptors can be especially useful for detecting intra-image lines of symmetry, such as water reflections in scene analysis. Research has explored reflection-invariant HoG [5] and, more frequently, SIFT-based methods such as RIFT [6], MI-SIFT [7] and MIFT [8]. Generally, rotational invariance can be achieved by finding the dominant gradient and rotating the image patch so that the gradient is always in the same direction. RIFT, for example, divides normalized patches into four concentric rings of equal width, from each of which eight gradient orientation histograms are computed. The orientation is measured at each point relative to the direction pointing outward from the centre, thus maintaining rotation invariance.

#### B. Alignment and localization

In a recent work, [9] assessed object part localization and observed that the state-of-the-art methods augment the training set with mirrored images, but they did not result in bilaterally symmetric results. The authors introduced the term *mirrorability*

TABLE I: Conclusions of the invariance characteristics of ten feature detectors from [4]

Detector	Invariant
BRISK	No
FAST	Perfect
GFTT	Yes, after matching
HARRIS	Yes, after matching
ORB	No
SIFT	No
STAR	Perfect
SURF	No
MSCR	Somewhat
MSER	Somewhat

and a *mirror error* that correlated with localization errors in human pose estimation and face alignment.

#### C. Deep learning

While the recent adoption and development of neural network techniques have undoubtedly produced impressive results in computer vision tasks, and object and scene recognition in particular, they are not at all robust to variation in data. Studies have shown that changing an image in a way imperceptible to humans can cause a deep neural network (DNN) to label the image as something else entirely [10] and that it is easy to produce images that are completely unrecognizable to humans, but that state-of-the-art DNNs believe to be recognizable objects with 99.99% confidence [11].

Recently published research on a scene recognition system [12] includes an online demonstration. Figure 2 shows a set of four images and their mirror reflections (top row) with the *information regions* that the author's online demo produce. The information regions are salient areas that the system has identified in its quest to understand and describe an image. We find compelling the difference in the information regions and suggest that this demonstrates a bias to the horizontal orientation of the image.

Table II shows the detailed results of the scene recognition. The system determines the environment, semantic categories and SUN scene attributes [19]. The category column summarizes the highest scoring semantic category. Despite the differences in salient areas of the images, the overall categorization has not been affected. Each image and its mirror image are categorized the same in these examples. However, there are differences in the detail, which illustrate inconsistencies that, in boundary cases, could change the categorization. The semantic categories are rated with a likelihood. The Rock Arch – a stock image from the author's own demonstration – reduces in likelihood by 0.01 in the mirror image, the Palace of Westminster [13] is classified exactly the same in each pair, Tower Bridge – another stock image from the author's own demonstration – appears less like a skyscraper and more like an office building in the reflected image than in the original, and the City of London skyline [14] increases its likelihood of being an abbey in the reflection image. The inconsistency in the ratings, albeit small, further strengthen our resolve that computer vision systems are commonly biased to image horizontal orientation. It is also interesting to note that images from the author's own

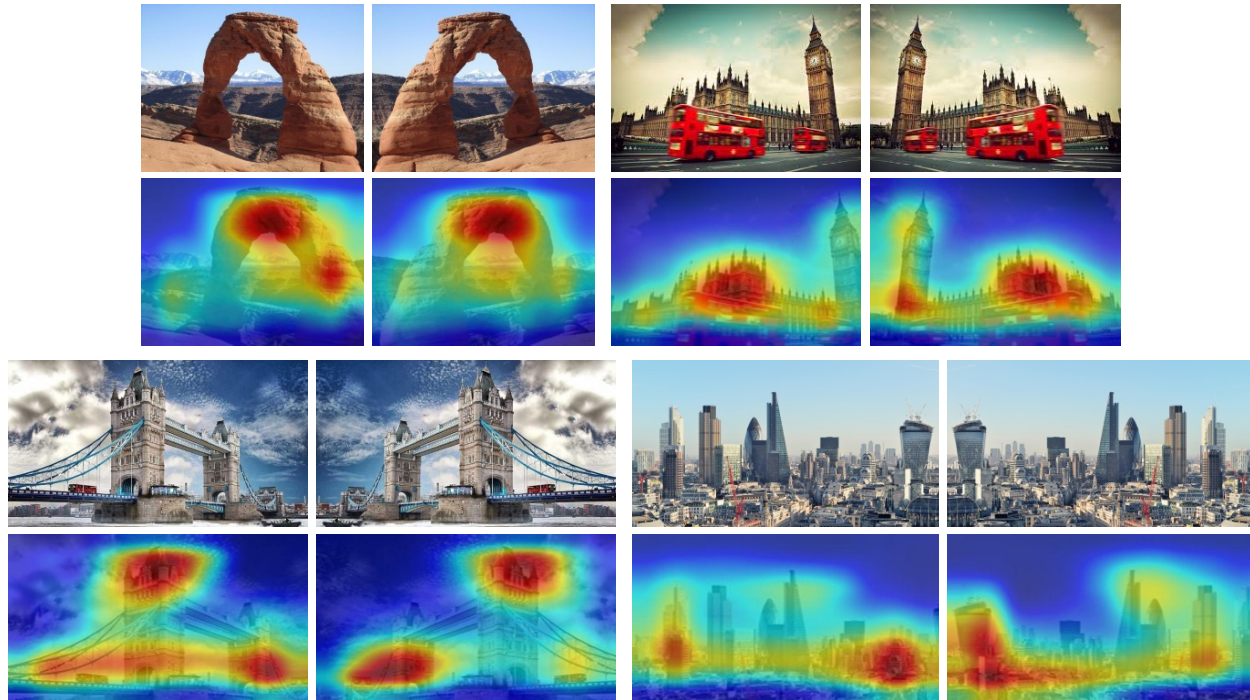


Fig. 2: *Informative regions* of images and their mirror, identified by [12]. Note that the informative regions are not mirror images, suggesting the algorithms are sensitive to the horizontal orientation of the image.

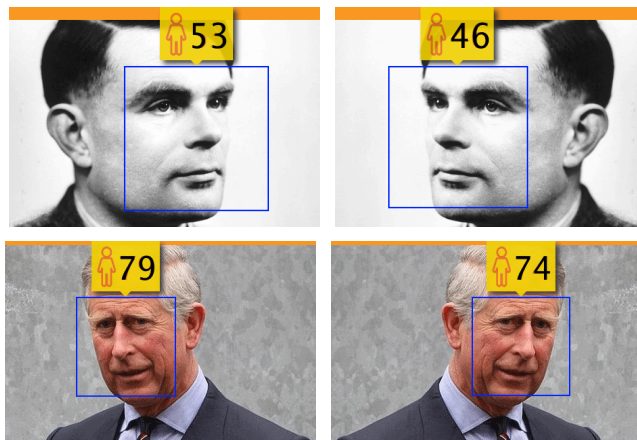


Fig. 3: Microsoft's How-Old.Net demonstration [16] attempts to guess a person's age from a photograph image. These two examples demonstrate that the system is sensitive to image orientation – and not head orientation – as the ages are quite different for each pair.

```









1      using namespace cv;
2
3      Mat src = imread("image.png",
4                      CV_LOAD_IMAGE_GRAYSCALE);
5
6      Mat fpt;
7      src.convertTo(fpt, CV_32F,
8                    SIFT_FIXPT_SCALE, 0);
9
10     Mat fpt_r;
11     flip(fpt, fpt_r, 1);
12
13     auto sigma = 1.24899971;
14     GaussianBlur(fpt, fpt, Size(), sigma, sigma);
15     GaussianBlur(fpt_r, fpt_r, Size(),
16                 sigma, sigma);
17
18     assert(countNonZero(fpt - fpt_r) == 0);

```

Fig. 4: Example C++ code to test reflection invariance of a Gaussian filter in OpenCV. Using 32-bit floating point arithmetic – CV\_32F on line 6 – will often result in an assertion failure on line 15 indicating that a Gaussian filter on a horizontally flipped image does not produce the same as the result as applying the same filter to the original image. Changing to use 64-bit double precision arithmetic – CV\_64F – produces identical results on all of our test images, with no assertion failures.



TABLE II: Predictions from Deep Learning Scene Recognition system [12]

	Environment	Semantic categories	SUN scene attributes	Category
	outdoor	rock_arch:0.75, arch:0.24	naturallight, openarea, ruggedscene, climbing, rockstone, directsun, sunny, dry, vacationing, touring, natural, warm	rock_arch
	outdoor	rock_arch:0.74, arch:0.25	naturallight, ruggedscene, rockstone, openarea, climbing, directsun, sunny, dry, vacationing, touring, warm, natural	rock_arch
	outdoor	tower:0.50, bridge:0.25, viaduct:0.12	man-made, clouds, openarea, naturallight, mostlyverticalcomponents, metal, vacationing, touring, nohorizon, directsun, sunny, congregating	tower
	outdoor	tower:0.50, bridge:0.25, viaduct:0.12	man-made, clouds, openarea, naturallight, mostlyverticalcomponents, metal, vacationing, touring, nohorizon, praying, directsun, sunny	tower
	outdoor	skyscraper: 0.72, tower: 0.13, office_building: 0.06	mostlyverticalcomponents, openarea, man-made, naturallight, directsun, sunny, far-awayhorizon, clouds, metal, driving, transportingthingsorpeople	skyscraper
	outdoor	skyscraper:0.66, tower:0.13, office_building:0.11	mostlyverticalcomponents, openarea, man-made, naturallight, directsun, sunny, driving, transportingthingsorpeople, clouds, far-awayhorizon, metal	skyscraper
	outdoor	abbey:0.64, palace:0.16	man-made, clouds, openarea, mostlyverticalcomponents, naturallight, vacationing, touring, praying, nohorizon, electricindoorlighting, metal	abbey
	outdoor	abbey:0.66, palace:0.15	clouds, man-made, openarea, mostlyverticalcomponents, naturallight, praying, vacationing, touring, nohorizon, metal, electricindoorlighting	abbey

demonstration score higher in the semantic categorization than images from other sources.

We used a second neural network based object recognition system, *The Wolfram Language Image Identification Project* [15], to test classification of our images, this time using different sizes of the same image. Table III shows the results; the Rock Arch is classified differently in its original orientation at a small scale, the Palace of Westminster was classified consistently at each scale, Tower Bridge is classified differently in its original orientation at a large scale and the London Skyline is classified differently in its mirror orientation at a large scale. These results show that this system is sensitive to scale, and that the scale change also influences the invariance to horizontal reflection.

Finally, Microsoft's much publicised How-Old.net [16] asks "How Old Do I Look?" and uses machine learning to guess the answer to the question from a photograph. We used photographs of Alan Turing [17] and Prince Charles [18] and observed the difference in age that was guessed for each image and its

reflection (Figure 3). In both cases, the ages decreased in the reflected image (*right*), despite the orientation of the head being different in each case.

This inconsistency in results is perhaps more surprising as the image orientation affects the guess of the person's age, but the system does not appear to be intrinsically biased towards the orientation of the head itself. On close examination, the bounding boxes of the identified *faces* are different sizes – smaller in the reflected image in both cases – by 5 pixels in each *x*- and *y*-axis in the case of the photograph of Alan Turing and 1 pixel in each axis in the case of Prince Charles. The detected face of Alan Turing is in a consistent corner position relative to the visible ear, and the detected face of Prince Charles is consistent in the opposite top corner. We therefore conclude that the face detection algorithm used in the system is sensitive to head orientation and this may affect the subsequent learned system of age estimation, which may or may not be orientation-sensitive itself.

TABLE III: Object recognition results from the online *Wolfram Language Image Identity Project*

Resolution	Original	Mirror
550 × 412		
	arch	arch
244 × 183		
	broken arch	arch
736 × 490		
	fire truck	fire truck
275 × 183		
	building	building
607 × 338		
	bascule	church
329 × 183		
	church	church
4370 × 2383		
	oil refinery	industrial park
336 × 183		
	oil refinery	oil refinery

#### IV. ALGORITHMS AND IMPLEMENTATIONS

Many algorithms described in the research literature – especially saliency based feature detectors – are not inherently sensitive to orientation. Nonetheless, no mention is made of reflection invariance in the papers, suggesting a general unawareness of this property. Consequently, we have observed several cases where commonly used, freely available code – including reference implementations from the original authors – have an invariance worsened by, or *caused by*, choices made in the implementation. For example, algorithms that use a Difference-of-Gaussian pyramid for sub-pixel feature detection can inadvertently increase their reflection dependence by the use of 32-bit floating point arithmetic for intermediate calculations. Using the popular OpenCV [20] tool kit for C++, we tested the `GaussianBlur()` function that convolves an image

with a specified Gaussian kernel. We found that using 32-bit arithmetic produces reflection-sensitive convolutions for many images that we tested (not shown), but using 64-bit arithmetic all convolutions of our test images were reflection invariant (Figure 4).

Conceptually, one would expect salient regions to be less biased to horizontal orientation, because they use neighbourhood colour and intensity measures and are less dependent on pixel gradients. However, common implementations of salient region detectors such as *maximally stable extremal regions* (MSER) [21] can suffer in the initial step of the algorithm blurring the image with a Gaussian kernel. In their saliency detector reference implementation, [22] exhibit orientation sensitivity due to many reasons including floating point errors in colour quantization which are realized differently dependent on the order in which the data is processed, which is determined by the image orientation. Increasing floating point arithmetic to double-precision 64-bit calculations correct the quantization sensitivity to reflection invariance.

#### V. CONCLUSION

We have proposed *reflection invariance* to be an important consideration when designing and implementing algorithms. It is evident from the cited contemporary research projects that many inconsistencies exist within applications of scene classification, object detection and age-guessing when systems are presented with images and their horizontal reflections. In each of our examples, the systems have produced results that are different for each reflected image orientation. We have described where some of the sensitivity is exhibited in feature detection and descriptors, and the interested reader is referred to [9] for a detailed analysis and experiments in alignment and localization.

#### ACKNOWLEDGEMENTS

This work is funded by the European Union’s Seventh Framework Programme, specific topic “framework and tools for (semi-)automated exploitation of massive amounts of digital data for forensic purposes”, under grant agreement number 607480 (LASIE IP project).

#### REFERENCES

- [1] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov 2004.
- [2] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, vol. 1. IEEE, 2005, pp. 886–893.
- [3] J. Sivic and A. Zisserman, “Video Google: a text retrieval approach to object matching in videos,” in *Proceedings Ninth IEEE International Conference on Computer Vision*, vol. 2, Oct 2003, pp. 1470–1477.
- [4] C. Henderson and E. Izquierdo, “Symmetric Stability of Low Level Feature Detectors,” 2015. <http://doi.org/10.13140/RG.2.1.2029.7043>
- [5] A. Kanezaki, Y. Mukuta, and T. Harada, “Mirror reflection invariant HOG descriptors for object detection,” in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, Oct 2014, pp. 1594–1598.
- [6] S. Lazebnik, C. Schmid, and J. Ponce, “A sparse texture representation using local affine regions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1265–1278, Aug 2005.

- [7] R. Ma, J. Chen, and Z. Su, "MI-SIFT," in *Proceedings of the ACM International Conference on Image and Video Retrieval - CIVR '10*. New York, New York, USA: ACM Press, Jul 2010, p. 228.
- [8] X. Guo and X. Cao, "MIFT: A framework for feature descriptors to be mirror reflection invariant," *Image and Vision Computing*, vol. 30, no. 8, pp. 546–556, Aug 2012.
- [9] H. Yang and I. Patras, "Mirror, mirror on the wall, tell me, is the error small?" in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15)*, Jun 2015, pp. 4685–4693.
- [10] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," Dec 2013.
- [11] A. Nguyen, J. Yosinski, and J. Clune, "Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'15)*, Jun 2015.
- [12] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning Deep Features for Scene Recognition using Places Database," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 487–495.
- [13] Palace of Westminster, UK. Online at <https://s-media-cache-ak0.pinimg.com/736x/d2/5a/0e/d25a0ed9bb2e788ae9c9ec59cc52670c.jpg> Accessed December 2015
- [14] City of London skyline, *Wikipedia*. Online at [http://upload.wikimedia.org/wikipedia/commons/d/da/The\\_City\\_London.jpg](http://upload.wikimedia.org/wikipedia/commons/d/da/The_City_London.jpg) Accessed December 2015
- [15] The Wolfram Language Image Identification Project. Online at <https://www.imageidentify.com> Accessed December 2015
- [16] Microsoft "How Old Do I Look?" Online at <http://how-old.net> Accessed December 2015
- [17] Photograph of Alan Turing from *Wikipedia*. Online at <https://kpfa.org/wp-content/uploads/2015/05/Dr-Alan-Turing-2956483.jpg> Accessed December 2015
- [18] Photograph of Prince Charles from *The Telegraph*. Online at [http://i.telegraph.co.uk/multimedia/archive/01422/princeCharles\\_1422434c.jpg](http://i.telegraph.co.uk/multimedia/archive/01422/princeCharles_1422434c.jpg) Accessed December 2015
- [19] J. Xiao, K.A. Ehinger, J. Hays, A. Torralba, and A. Oliva, "SUN Database: Exploring a Large Collection of Scene Categories," *International Journal of Computer Vision*, 2014.
- [20] G. R. Bradski, "The OpenCV Library," *Dr. Dobbs's Journal of Software Tools*, 2000.
- [21] P.-E. Forssén and D. G. Lowe, "Shape descriptors for maximally stable extremal regions," in *Proceedings of the IEEE International Conference on Computer Vision*, 2007.
- [22] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'11)*, Jun 2011, pp. 409–416.