ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 4, Issue 1, pp: (76-81), Month: January - March 2016, Available at: www.researchpublish.com

Computer Vision

Rishabh Soni

Engineering Graduate,Venkteshwar Institute of Technology, Indore, India

Abstract: **Computer vision is a field that includes methods for acquiring, processing, analyzing, and understanding images and, in general, high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions. A theme in the development of this field has been to duplicate the abilities of human vision by electronically perceiving and understanding an image. Understanding in this context means the transformation of visual images (the input of retina) into descriptions of world that can interface with other thought processes and elicit appropriate action. This image understanding can be seen as the disentangling of symbolic information from image data using models constructed with the aid of geometry, physics, statistics, and learning theory. Computer vision has also been described as the enterprise of automating and integrating a wide range of processes and representations for vision perception.**

Keywords: **Image detection, Structure from motion, Image reconstruction, Recognition, Image processing.**

I. INTRODUCTION

As humans, we perceive the three-dimensional structure of the world around us with apparent ease. Think of how vivid the three-dimensional percept is when you look at a vase of flowers sitting on the table next to you. You can tell the shape and translucency of each petal through the subtle patterns of light and shading that play across its surface and effortlessly segment each flower from the background of the scene. Looking at a framed group por- trait, you can easily count (and name) all of the people in the picture and even guess at their emotions from their facial appearance. Perceptual psychologists have spent decades trying to understand how the visual system works and, even though they can devise optical illusions1 to tease apart some of its principles, a complete solution to this puzzle remains elusive(Marr 1982; Palmer 1999; Livingstone 2008).

Researchers in computer vision have been developing, in parallel, mathematical tech- niques for recovering the three-dimensional shape and appearance of objects in imagery. We now have reliable techniques for accurately computing a partial 3D model of an environment from thousands of partially overlapping photographs. Given a large enough set of views of a particular object or fac¸ade, we can create accurate dense 3D surface mod- els using stereo matching. We can track a person moving against a complex background. We can even, with moderate success, attempt to find and name all of the people in a photograph using a combination of face, clothing, and hair detection and recognition. However, despite all of these advances, the dream of having a computer interpret an image at the same level as a two-year old (for example, counting all of the animals in a picture) remains elusive. Why is vision so difficult? In part, it is because vision is an inverse problem, in which we seek to recover some unknowns given insufficient information to fully specify the solution. We must therefore resort to physics-based and probabilistic models to disambiguate between potential solutions. However, modeling the visual world in all of its rich complexity is far more difficult than, say, modeling the vocal tract that produces spoken sounds.

The forward models that we use in computer vision are usually developed in physics (radiometry, optics, and sensor design) and in computer graphics. Both of these fields model how objects move and animate, how light reflects off their surfaces, is scattered by the atmosphere, refracted through camera lenses (or human eyes), and finally projected onto a flat (or curved) image plane. While computer graphics are not yet perfect (no fully computer- animated movie with human characters has yet succeeded at crossing the uncanny valley2 that separates real humans from android robots and

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 4, Issue 1, pp: (76-81), Month: January - March 2016, Available at: www.researchpublish.com

computer-animated humans), in limited domains, such as rendering a still scene composed of everyday objects or animating extinct creatures such as dinosaurs, the illusion of reality is perfect.

In computer vision, we are trying to do the inverse, i.e., to describe the world that we see in one or more images and to reconstruct its properties, such as shape, illumination, and color distributions. It is amazing that humans and animals do this so effortlessly, while computer vision algorithms are so error prone. People who have not worked in the field often under- estimate the difficulty of the problem. (Colleagues at work often ask me for software to find and name all the people in photos, so they can get on with the more "interesting" work.) This misperception that vision should be easy dates back to the early days of artificial intelligence, when it was initially believed that the cognitive (logic proving and planning) parts of intelligence were intrinsically more difficult than the perceptual components (Boden 2006).

II. OBJECT DETECTION IN IMAGES

Detection of objects consists broadly of four stages:

Fig. 1. Left: A sample object image used in vocabulary construction. Center: Interest points detected by the F¨orstner operator. Crosses denote intersection points; circles denote centers of circular patterns. Right: Patches extracted around the interest points

Fig. 2. The 400 patches extracted by the F¨orstner interest operator from 50 sample images.

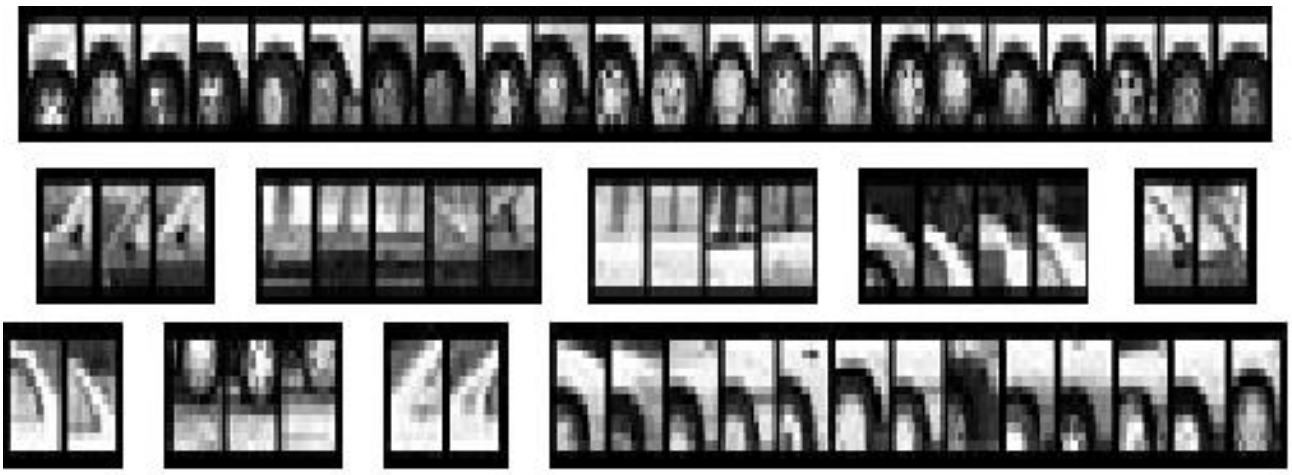


Fig. 3. Examples of some of the "part" clusters formed after grouping similar patches together. These form our part vocabulary

A. Vocabulary Construction-

The first stage in the approach is to develop a vocabulary of parts with which to represent images. To obtain an expressive representation for the object class of interest, we require distinctive parts that are specific to the object class but can also capture the variation across different instances of the object class. Our method for automatically selecting such parts is based on the extraction of interest points from a set of representative images of the target object.

B. Image Representation-

Having constructed the part vocabulary above, images are now represented using this vocabulary. This is done by determining which of the vocabulary parts are present in an image, and then representing the image as a binary feature vector based on these detected parts and the spatial relations that are observed among them.

C. Image Representation-

Having constructed the part vocabulary above, images are now represented using this vocabulary. This is done by determining which of the vocabulary parts are present in an image, and then representing the image as a binary feature vector based on these detected parts and the spatial relations that are observed among them.

D. Learning a Classifier-

Using the above feature vector representation, a classifier is trained to classify a 100×40 image as car or non-car. We used a training set of 1000 labeled images (500 positive and 500 negative), each 100×40 pixels in size.2 The images were acquired partly by taking still photographs of parked cars, and partly by grabbing frames from digitized video sequences of cars in motion. The photographs and video sequences were all taken in the Champaign-Urbana area. After cropping and scaling to the required size, histogram equalization was performed on all images to reduce sensitivity to changes in illumination conditions. The positive examples contain images of different kinds of cars against a variety of backgrounds, and include images of partially occluded cars. The negative training examples include images of natural scenes, buildings and road views. Note that our training set is relatively small and all images in our data set are natural; we do not use any synthetic training images, as has been done.

Detection Hypothesis Using the Learned Classifier Having learned a classifier4 that can classify 100 × 40 images as positive or negative, cars can be detected in an image by moving a 100 × 40 window over the image and classifying each such window as positive or negative. However, due to the invariance of the classifier to small translations of an object, several windows in the vicinity of an object in the image will be classified as positive, giving rise to multiple detections corresponding to a single object in the scene. A question that arises is how the system should be evaluated in the presence of these multiple detections. In much previous work in object detection, multiple detections output by the system are all considered to be correct detections (provided they satisfy the criterion for a correct detection; this is discussed later in Section III-B). However, such a system fails both to locate the objects in the image, and to form a correct hypothesis about the number of object instances present in the image. Therefore in using a classifier to perform detection, it is necessary to have another processing step, above the level of the classification output, to produce a coherent detection hypothesis.

III. APPLICATIONS OF COMPUTER VISION

Applications range from tasks such as industrial machine vision systems which, say, inspect bottles speeding by on a production line, to research into artificial intelligence and computers or robots that can comprehend the world around them. The computer vision and machine vision fields have significant overlap. Computer vision covers the core technology of automated image analysis which is used in many fields. Machine vision usually refers to a process of combining automated image analysis with other methods and technologies to provide automated inspection and robot guidance in industrial applications. In many computer vision applications, the computers are pre-programmed to solve a particular task, but methods based on learning are now becoming increasingly common. Examples of applications of computer vision include systems for:

- **Optical character recognition (OCR):** reading handwritten postal codes on letters (Figure 1.4a) and automatic number plate recognition (ANPR);

- **Machine inspection:** rapid parts inspection for quality assurance using stereo vision with specialized illumination to measure tolerances on aircraft wings or auto body parts (Figure 1.4b) or looking for defects in steel castings using X-ray vision;
- **Retail:** object recognition for automated checkout lanes (Figure 1.4c);
- **3D model building (photogrammetry):** fully automated construction of 3D models from aerial photographs used in systems such as Bing Maps;
- **Medical imaging:** registering pre-operative and intra-operative imagery (Figure 1.4d) or performing long-term studies of people's brain morphology as they age;
- **Automotive safety:** detecting unexpected obstacles such as pedestrians on the street, under conditions where active vision techniques such as radar or lidar do not work well (Figure 1.4e; see also Miller, Campbell, Huttenlocher *et al.* (2008); Montemerlo, Becker, Bhat *et al.* (2008); Urmson, Anhalt, Bagnell *et al.* (2008) for examples of fully automated driving);
- **Match move:** merging computer-generated imagery (CGI) with live action footage by tracking feature points in the source video to estimate the 3D camera motion and shape of the environment. Such techniques are widely used in Hollywood (e.g., in movies such as Jurassic Park) (Roble 1999; Roble and Zafar 2009); they also require the use of precise *matting* to insert new elements between foreground and background elements (Chuang, Agarwala, Curless *et al.* 2002).
- **Motion capture (mocap):** using retro-reflective markers viewed from multiple cam- eras or other vision-based techniques to capture actors for computer animation;
- **Surveillance:** monitoring for intruders, analyzing highway traffic (Figure 1.4f), and monitoring pools for drowning victims;
- **Fingerprint recognition and biometrics:** for automatic access authentication as well as forensic applications.

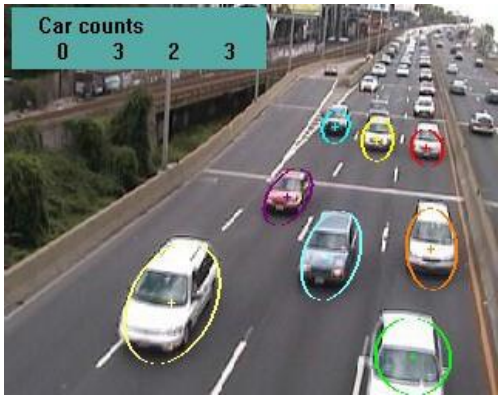
(a)

(b)

(c)

(d)

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 4, Issue 1, pp: (76-81), Month: January - March 2016, Available at: www.researchpublish.com



(e) **(f)**

Figure 1.4 Some industrial applications of computer vision: (a) optical character recognition (OCR) http://yann.lecun.com/exdb/lenet/; (b) mechanical inspection http://www.cognitens. com/; (c) retail http://www.evoretail.com/; (d) medical imaging http://www.clarontech.com/; automotive safety http://www.mobileye.com/; (f) surveillance and traffic monitoring http: //www.honey wellvideo.com/, courtesy of Honeywell International Inc.

IV. CONCLUSION

Three major trends in the future development of computer vision can already be detected. First, vision and testing are the driving forces helping in identifying faults so that their causes can be successively eliminated, ultimately permitting the successive elimination of testers and, possibly, vision systems themselves. As a consequence, more and more feedback will be included into the process, through configuration, regulation, and explanation facilities.

REFERENCES

[1] Richard Szeliski, Computer Vision: Algorithms and Applications.Draft Sept 3, 2010

[2] Shivani Agarwal, Aatif Awan and Dan Roth, Member, IEEE Computer Society, Learning to Detect Objects in Images via Sparse, Part-Based Representation

[3] S. Kranthi, K. Pranathi, A. Srisaila "Automatic Number Plate Recognition", Information Technology, VR Siddhartha Engineering College, Vijayawada, India, volume 2,2012.

[4] S. E. Palmer, “Hierarchical structure in perceptual representation,” Cognitive Psychology, vol. 9, pp. 441–474, 1977.

[5] Wachsmuth, M. W. Oram, and D. I. Perrett, “Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque,” Cerebral Cortex, vol. 4, pp. 509–522, 1994.

[6] /A. Mohan, C. Papageorgiou, and T. Poggio, “Example-based object detection in images by components,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pp. 349– 361, 2001.

[7] /A. J. Colmenarez and T. S. Huang, “Face detection with information-based maximum discrimination,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1997, pp. 782–787.

[8] H. A. Rowley, S. Baluja, and T. Kanade, “Neural networkbased face detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 23–38, 1998.

[9] Osuna, R. Freund, and F. Girosi, “Training support vector machines: an application to face detection,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1997, pp. 130–136.

[10] M. Turk and A. Pentland, “Eigenfaces for recognition,” Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71–86, 1991.

[11] B. Moghaddam and A. Pentland, “Probabilistic visual learning for object detection,” Proceedings of the Fifth International Conference on Computer Vision, 1995.

[12] Y. Amit and D. Geman, "A computational model for visual selection," Neural Computation, vol. 11, no. 7, pp. 1691–1715, 1999.

[13] M-H. Yang, D. Roth, and N. Ahuja, "A SNoW-based face detector," in Advances in Neural Information Processing Systems 12, Sara A. Solla, Todd K. Leen, and Klaus-Rober M¨uller, Eds., 2000, pp. 855–861.

[14] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2001.

[15] L. Shams and J. Spoeslstra, "Learning Gabor-based features for face detection," in Proceedings of World Congress in Neural Networks, International Neural Network Society, 1996, pp. 15– 20.

[16] C. Papageorgiou and T. Poggio, "A trainable system for object detection," International Journal of Computer Vision, vol. 38, no. 1, pp. 15–33, 2000.

[17] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object recognition with gradient-based learning," in Feature Grouping, Forsyth, Ed., 1999.

[18] Schneiderman and T. Kanade, "A statistical method for 3D object detection applied to faces and cars," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2000, vol. 1, pp. 746–751.