Research Statement

Jeffrey Ho
Department of Computer Science and Engineering
University of California, San Diego
La Jolla, CA 92122-0114
jho@cs.ucsd.edu

January 18, 2004

The advances in computer hardware and peripheral equipments in the past few years have made it feasible to acquire, store and process large amount of visual data (images). Specifically, the improvements in (video) camera technology have made it both economical and convenient to acquire various kinds of image data, while the advances in processor and hardware memory allow us to process and store them with relative ease. One immediate consequence of this progress is the recent emergence of many exciting and potentially revolutionary applications, ranging from the fascinating face recognition systems to the breathtaking virtual reality demos. The undeniable trend of development is to endow the computers with the capability of understanding and appreciating the visual world in which we all live. From a more academic viewpoint, this also makes computer vision one of the most exciting areas of research in computer science, with numerous interesting problems and fascinating applications awaiting our solutions and discoveries.

My main research interest has been in studying some of the traditional and fundamental problems in computer vision, namely, object recognition, visual tracking and image clustering. Great emphasis is always placed on the practicality of the solution. However, a robust and efficient algorithm is usually not the final goal for my research; instead, I strive for a deeper and more conceptual understanding of each vision problem in its own unique way.

Previous and Current Research

Some of the more successful efforts from the past few years are summarized below. The unifying theme for the following somewhat disparate computer vision problems is linearity. In each case, our (my colleagues and I) main contribution is in formulating the correct linear problem to solve. There are two important advantages for preferring linear technique. First, its usually makes the algorithm conceptually more transparent and easier to implement. Second, it generally demands less computational resources and therefore; the algorithm's performance (speed) can be greatly enhanced.

Illumination and Face Recognition [5, 6, 11] Arguably, face recognition is the computer application that can invoke curiosity and fascination from the general public. From a technical standpoint, this technology is also essential for the development of more advanced and sophisticated HCI (Human-Computer Interaction) applications. However, after years of research, a robust face recognition algorithm with human-like performance is still elusive. The main challenge is in designing a comprehensive and specific model that allows the system to correctly recognize people under different imaging conditions. In particular, images of a person taken under different external illumination conditions can appear disparate, and a robust face recognition system should be capable of perceiving the superficial dissimilarity caused by the illumination variation.

In the works cited above, we develop a systematic and principled method for handling external illumination variation for face recognition. Based on the concept of Lambertian reflectance, we propose a linear model (subspace) to model the image variation caused by illumination variation. Our method offers a computational solution for determining the basis of the linear subspace, which is consisted of a small number of images taken under some prescribed illumination conditions. Extensive evaluations of our algorithm have shown that the proposed linear model is indeed effective for face recognition under significant illumination variation.

Video Face Recognition[4, 10] A logical extension of the previous work is to design recognition algorithms that recognize people in video sequences. Potential applications of this technology are numerous, ranging from the security surveillance at airports to sophisticated humanoid robots such as Honda's ASIMO. Besides illumination variation, the problem's difficulty is now compounded by the frequent and unpredictable pose variation occurred in the video sequence. Furthermore, the addition of the temporal dimension also marks the point of departure from the more traditional and static recognition problems studied in machine learning. Therefore, a more refined and elaborate algorithm is required to engage these new challenges.

In our solution, a small number of linear subspaces are used to model individual pose states and the unpredictable pose variation is then modelled by Markov-type transitions between the subspaces. The algorithm also exploits the important idea of "temporal coherence" by coupling Markov-type transitions and recognition together under the usual Bayesian framework. Experiments have shown that the addition of the temporal element in the recognition problem actually makes it more robust and stable.

Image Clustering[3, 2] The recognition problems we discussed above are supervised learning problems. A clustering problem, on the other hand, can be considered as an unsupervised learning problem and a clustering algorithm tries to detect some underlying patterns among the usually large collection of input data, and these detected patterns can be used to construct representations for the observed data. The input of our image clustering algorithms is a large collection of unlabelled images (in the hundreds and sometimes in the thousands), which contains images of different objects (ranging from 10 to 100) taken under various imaging conditions. The algorithm is able to automatically cluster the image collection according to the object, i.e. each cluster of images consists of only images of the same object.

The image clustering algorithms we proposed is purely computational. Conceptually, our algorithms rely on detecting global and local linear structures hidden among the input images. In addition, because of the natural action of 2D affine group on images, our algorithms is required to be invariant with respect to this affine symmetry. Experiments with large collections of human face images and images of natural objects demonstrate that our algorithms are quite effective in clustering large collections of images.

Visual Tracking[1, 9] Visual tracking has always been an intensively studied subject in computer vision literature. Its importance is supported by the fact that tracking algorithms usually form the basis of a wide variety of vision-related applications. Again, the challenge here is to design an algorithmic model that can anticipate diverse and occasionally adverse changes in the external environment that can affect the tracking result. However, a subtle and important difference between recognition and tracking is that the models required by the former are usually specific while the latter asks for more flexible and adaptive models.

Our main contribution to this time-honored subject is the "discovery" that a simple linear model can be used to form the basis of a robust tracker. The secret of making the linear model both flexible and adaptive is to use previous tracking results to update the model according to the uniform norm (instead of the usual L^2 norm) in the image space. Based only on this simple linear model and eschewing the usual complicated probability estimates and non-linear optimizations, good tracking results have been obtained that demonstrate the robustness of our tracker against challenging imaging conditions which include drastic illumination variation, partial occlusion and extreme pose variation.

Mesh Compression[8, 7] In computer vision and graphics, mesh is the standard geometric construct for storing and manipulating 3D data. Our interest in mesh compression was spurred by the collection of massive meshes (containing millions of vertices) produced by the renowned Michelangelo project. Based on the earlier work on mesh compression, our two main contributions are 1) an algorithm for partitioning the mesh into sufficient small submeshes such that each submesh can be independently fitted into memory and compressed separately and 2) a compression/de-compression scheme for gluing (along the boundaries of the submeshes) the submeshes back into the original mesh.

Future Objectives

The aforementioned works represent the bulk of my research output in the past three years (some are also on-going) as a post-doctoral research associate under the supervision of Prof. David Kriegman. These projects constitute the obligatory exercises for my transition from mathematics to computer vision. They provide valuable

opportunity for me to get acquainted with the literature, sharpen my intuition and adjust my mode of thinking, from the penchant for the continuous to the pragmatism of the discrete. In the future, I intend to push the frontier of computer vision science further by widening the existing repertoire of vision-related applications and by broadening its theoretical foundation. Among the projects I am currently pursuing or intend to pursue in the future, the following two broad and general directions seem most interesting and potentially far-reaching:

Distributed and Omnipresent Vision Computer vision is traditionally considered as a branch of AI. Not surprisingly, many problems studied in computer vision have their roots in the desire to emulate certain visual cognitive capabilities of a human. Recognition, tracking and 3D reconstruction are good examples. However, with the rapid advances in hardware technology, this anthropocentric and somewhat limited paradigm may no longer be the only source of inspiration for computer vision. For example, with the explosive growth of the internet and web-based cameras, we have potentially the capability of viewing the world with numerous different perspectives (literally!) everywhere and instantaneously. The available visual data are much greater in quantity and varied in quality than the visual data that can be gathered by a person's eyes. Therefore, the important research question we should ask is how should the immense computing power and communication capacity be harnessed to produce/create new kinds of intelligence and their manifestations (applications).

One such "intelligence" can be the "collective intelligence" exhibited by a collection of smaller units of intelligence. Colonies of ants and bees are prominent examples in the nature, while the self-reconfigurable robots provide an analogous concept in robotics. In computer vision, a camera and a computer constitute physically the simplest visual intelligent unit and the network provides the necessary infrastructure for communication. The fundamental question is then to identify/invent new applications (intelligence) and to explore other possibilities using a large collections of well-connected visual intelligent units. Under this new paradigm, many interesting questions and issues can be raised. How should the visual knowledge gathered at each unit be represented and communicated with others? How should pieces of the visual knowledge be processed and aggregated in order to produce certain type of global awareness of the visual world? What is a good working definition of global awareness of the visual world? And many more. These questions are fascinating to ponder and they provide the test grounds for generating new ideas and directions for future research. In addition, the connection with other branches of computer science, notably distributed and parallel computing, is inescapable and the interaction would likely benefit all disciplines.

Large-Scale Vision: With the amount of visual data available to us increasing at an explosive rate, it is imperative to enlarge the capacity of the existing algorithms to deal with this new challenge. Until now, most of the algorithms developed in computer vision (as well as in machine learning) are trained, evaluated and tested on data samples numbering in the hundreds and sometimes, in the thousands. The size of data is rarely in the range of the tens of thousands or more. Furthermore, they are usually not designed for handling large amount of data, and when applied to a large collection of data, they either become unstable or are simply inefficient. In many ways, working with data collections whose sizes differ in several orders of magnitude is similar to doing physics at different scales. Many cherished concepts and proven practices may have to be re-examined and modified in order to meet the new demands and reality.

For instance, several different advances and improvements are necessary in order for our face recognition system to perform reliably and efficiently with a database of size in the tens of thousands or millions. First, there is the problem of storing the data; hence, a new compression scheme or a simplified representation is required. For an efficient system, a new and faster search algorithm (used in the recognition process) is called for, and this may entail a fundamental re-design of the recognition algorithm altogether. There is also the problem of feature extraction. With so many possible samples, what are other significant features one can extract from the images to enhance the recognition performance? And finally, the problem of how to evaluate and test this new algorithm also has to be addressed, both theoretically and empirically. That is, how does one know the performance of the algorithm is stable with respect to the data size? It is clear that there are many other pertinent questions, some significant, some interesting and some fundamental. With the present developmental trend, it is inevitable that they have to be addressed and solved sooner or latter in order to make further progresses.

In summary, I believe that computer vision has reached a critical stage in its development. With the aid of better hardware technology, many interesting and vision-related applications have appeared in the past few years. The importance and impact of computer vision to the world and to the technology development in particular will be increasingly felt and recognized. Many exciting applications and fascinating problems are still awaiting our

discoveries and solutions, and they provide a fertile ground for new and exciting research. Similar to many other aspiring academics, the ultimate objective of my career is to produce a body of decent works that can enrich my chosen discipline and hopefully at the end, they can make some differences in this ever-changing world.

References

[1] Jeffrey Ho, Kuang-Chih Lee, David Kriegman, "Visual Tracking with Learned Linear Subspaces", submitted to *IEEE Conf. On Computer Vision and Pattern Recognition*, Washington D.C., U.S.A., 2004.

[2] Jongwoo Lim, Jeffrey Ho, Ming-Hsuan Yang, Kuang-Chih Lee, David Kriegman, "Image Clustering with Metric, Local Linearity and Affine Symmetry", submitted to *European Conf. On Computer Vision*, Prague, Czech Republic, 2004.

[3] Jeffrey Ho, Ming-Hsuan Yang, Jongwoo Lim, Kuang-Chih Lee, David Kriegman, "Clustering Appearances of Objects Under Varying Illumination Conditions," *IEEE Conf. On Computer Vision and Pattern Recognition*, 2003, vol.1 , pp. 11-18.

[4] Kuang-Chih Lee, Jeffrey Ho, Ming-Hsuan Yang, David Kriegman. " Video-Based Face Recognition Using Probabilistic Appearance Manifolds," *IEEE Conf. On Computer Vision and Pattern Recognition*, 2003, vol. 1, pp. 313-320.

[5] Kuang-Chih Lee, Jeffrey Ho, David Kriegman,"Nine Points of Lights: Acquiring Subspaces for Face Recognition under Variable Illumination," *IEEE Conf. On Computer Vision and Pattern Recognition*, 2001, vol. 1, pp. 519-526.

[6] Jeffrey Ho, Kuang-Chih Lee, David Kriegman, "On Reducing the Complexity of Illumination Cones," *IEEE Workshop on Identifying Objects Across Variations in Lighting: Psychophysics and Computation*, 2001, pp. 56-63.

[7] Jeffrey Ho, Kuang-chih Lee, David Kriegman, "Compressing Large Polygonal Models," *IEEE Conference on Visualization*, 200, pp. 357-362.

[8] Jeffrey Ho, Kuang-chih Lee, David Kriegman, "Compressing Large Polygonal Models," *SIGGRAPH* Technical Sketch, 2001, pp. 159.

[9] Jeffrey Ho, Kuang-chih Lee, David Kriegman, "Visual Tracking Using Uniform Reconstruction Error Norm."

[10] Kuang-chih Lee, Jeffrey Ho, Ming-Hsuan Yang, David Kriegman, "Visual Tracking and Face Recognition Using Probabilisitic Appearance Manifolds."

[11] Kuang-chih Lee, Jeffrey Ho, David Kriegman, "Acquiring Linear Subspaces for Face Recognition under Variable Lighting," To Appear in *IEEE Trans. Pattern Analysis and Machine Intelligence.*