RESEARCH STATEMENT
Carlos D. Castillo
April, 2011

I am interested in all aspects of computer vision. In general my research is about detecting and recognizing objects. My doctoral dissertation is on the topic of face recognition across pose and illumination. The type of methods I apply are at the intersection of computer vision and machine learning.

Face recognition is a fundamental problem in computer vision. There has been a lot of progress in the case of images taken under controlled conditions. There are many approaches for handling, variation of illumination and expression. There are also several approaches to handling pose variation. However, there is still a lot of room for improvement. When multiple confounding factors occur simultaneously the problem is often termed unconstrained face recognition. Progress in unconstrained face recognition would be important in many applications, for example: surveillance, security, the analysis of personal photos and other domains in which we cannot control the conditions under which the images are taken.

Existing systems achieve excellent results when images are taken under controlled conditions, so that there is no variation in viewing conditions. Recently, there has been a good deal of work on recognition in the case of variations in viewing conditions that occur over a short period of time, such as variations in pose or lighting. Variations that occur over longer periods of time (such as aging and weight gain) are harder to study.

Face recognition across pose is the task of determining whether two faces seen from different viewpoints and perhaps illuminated differently are of the same person. Recognizing faces across pose is important in safety and surveillance applications and also appears frequently in day to day tasks such as personal photo organization.

The general direction of my research is in the area of direct image comparison for face recognition. I believe that direct image comparison is a very powerful method for face recognition in unconstrained settings.

Summary of Previous and Ongoing Research

In this section I will describe the research projects I've worked on. Here the projects are organized as independent units, while in reality they are more of a continuum. There are four projects, the first two are closer to "core" computer vision and the final two are closer to machine learning.

Stereo Matching for Face Recognition Across Pose.

The methods I have developed to approach the problem of face recognition across pose are based on stereo matching. Stereo matching is a well understood problem and provides a firm foundation to build on. At the core of stereo matching is the concept of correspondences. These correspondences turn out to be fundamental to recognize faces across pose.

In making stereo matching methods more appropriate for face recognition I have developed a method that accounts for slanted surfaces. Accounting for slant seems to be necessary to handle large pose variation (for example to compare a frontal view to a to profile view).

We have built a fast, practical method for stereo matching in the presence of medium pose variation [4]. Also, we have built a method that is robust to large and very changes in viewpoint and illumination when matching very slanted objects [2].

We have evaluated this approach both in controlled settings (like PIE and MultiPIE) obtaining with both methods state of the art results. Our results are almost perfect for azimuthal pose differences of up to 30 degrees, and from there on the results and from there on the results gracefully degrade.

Dense Wide-baseline Matching with Varying Illumination

Using stereo matching for face recognition is an application of stereo matching where the illumination can vary significantly, Using our insights from using stereo matching to compare faces, we have developed a formulation that would allow to adapt 2-D Markov Random Field based stereo formulation for wide baseline dense matching with variation in illumination.

Figure 1: Several outdoor wide-baseline images.

We have been studying a family of energy functions that fully account for both slant and tilt and have properties that make it amenable to optimization.

Our initial results show that this energy function is very robust to changes in illumination that occur in wide baseline stereo and our results compare favorably with other methods to handle wide baseline dense stereo matching.

Descriptor-based Learning for Face Verification

One inherent limitation of the stereo-based methods I have developed is that they weigh each location in the face equally. There are strong reasons to believe that this is not a good idea; differences around the eyes should be more significant than differences in the cheeks, for example.

I have developed methods to integrate large-scale spatial learning into our stereo-based face recognition work. This formulation will allow us to learn how to compensate for slight-variations in the images that are not being explicitly accounted for by the (pose+illumination) model described in the previous item. These variations include: expression, aging, weight variation, etc.

We have evaluated this approach both in controlled settings (like PIE and MultiPIE) and in unconstrained settings using data sets like Labeled Faces in the Wild and the results are encouraging. To learn more about this method, see [5].

Variations and Traits in Machine Learning

We have studied the problem of meta features and features in machine learning (I call these variations and traits, respectively). We first started thinking about this problem when it appeared in our own work on face recognition: for example if we are comparing two faces and we know that one of the two individuals has his eyes closed we would like to compare images differently. Here the eyes open feature is a variation and the similarity score is a trait.

This problem is very general and appears in many areas. For example, in Natural Language Processing this problem is called domain adaptation. In domain adaptation, we are given a large training set from the Wall Street Journal (called source dataset) and a smaller training set from PC Magazine (called target dataset), and the classifier will be evaluated using data from PC Magazine. For example for part of speech tagging there are words that will have the same use in the source and target domain. For these words having a large source training set will (should) help. However there are some words that will be used differently in the source and target domain, consider, for example, the word "monitor" that will likely be used in WSJ as a verb and in PC Magazine as a noun. The learning mechanism should be able to account for that. Here, in our formulation, the document description is a trait and the document origin is a variation.

We have proposed a general formulation for this problem based on the support vector machine framework. Our formulation is principled and natural and builds on the well founded support vector machine framework.

Summary of Research Plans

For the near-term future there are several broad areas that I'm hoping to pursue:

1. **Face Recognition with Non-homogeneous Galleries:** In most if not all face recognition methods across pose, illumination and/or expression the variation in the gallery is fixed but arbitrary. Non-homogeneous galleries are beyond the state of the art. I would like to develop data-driven methods to handle non-homogeneous galleries robustly. I think would be a good application of the variations/traits work I've been doing.

2. **Generalized b-matching in Clustering and Classification:** In many applications of similarity-based classification and clustering the following issue appears: "a hub" that is similar to everything, an average looking face, an average signal, etc. I plan to study a novel class of algorithms to enforce a principled type of sparsity constraint based on generalized b-matching. This project may fold in nicely with the non-homogeneous galleries work I plan on doing.

3. **Max-margin Learning with the Distance Transform:** The distance transform is one of the most classic ways of handling shape variation in image matching. I want to evaluate ways in which this can be integrated with max-margin classifiers. This would make it a very appealing way of handling shape variation.

4. **Parallel Stochastic Gradient Descent:** Massive data classification problems appear everywhere, in vision, bioinformatics, NLP, everywhere. However, it has some appealing features that make it amenable to parallelization. I plan to study how stochastic gradient descent can be run efficiently in many cores.

In the long term I would like to do research to understand the role of three key aspects for image representation: obtaining correspondences, accounting for deformation and performing learning. These may seem like fairly disparate topics, but I think that they're very closely connected. I also believe that these problems are fundamental in the process of building intelligent systems that are capable of performing visual object detection and recognition.

Additionally, I want to continue working on machine learning problems that arise in application areas such as in computer vision. The learning problems in vision have the common feature of being very difficult to approach due to the nature of the data. Data in computer vision is very high dimensional and also in general comes with a lot of structure.

References

[1] Carlos D. Castillo and David W. Jacobs *Foreshortening and Window Shape in Graphcut Stereo*, ICCV 2011, Under review.

[2] Carlos D. Castillo and David W. Jacobs *Wide-Baseline Stereo for Face Recognition with Large Pose Variation*, . In *CVPR*. IEEE Computer Society, 2011.

[3] Carlos D. Castillo and David W. Jacobs. Using stereo matching for 2-d face recognition across pose. In *CVPR*. IEEE Computer Society, 2007.

[4] Carlos D. Castillo and David W. Jacobs. Using stereo matching with general epipolar geometry for 2-d face recognition across pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Dec. 2009.

[5] Thomas Berg, Neeraj Kumar, Alexander C. Berg, Peter N. Belhumeur, Carlos D. Castillo and David W. Jacobs, *Attributes and Stereo Matching Costs for Face Verification*, Technical Report, 2010.