



Project Management and Software Development

A Vision Based Anthropometric Scanner

Intermediate Presentation

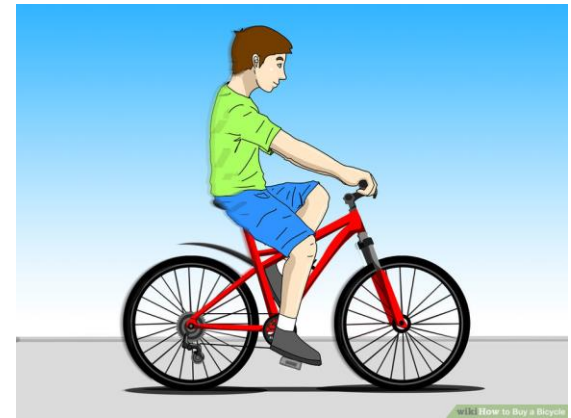
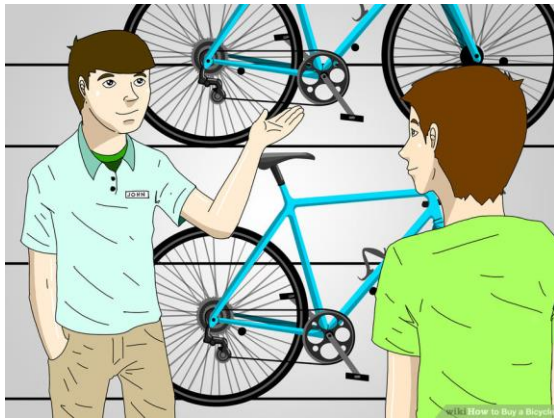
Ahmed El-Gazzar

Supervisor : Benjamin Busam



4 weeks ago...

Buying a bicycle

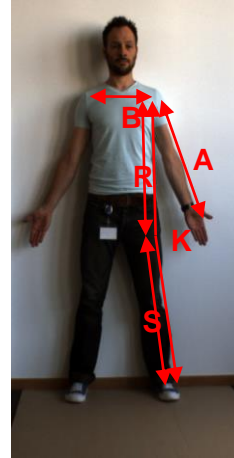
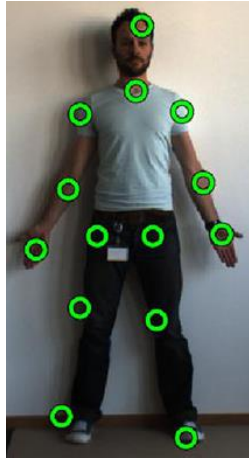
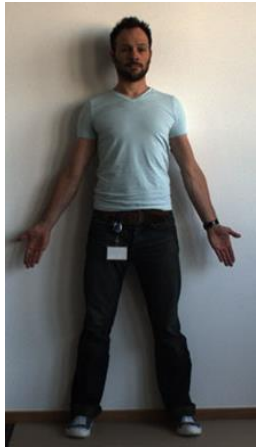


<http://www.wikihow.com/Buy-a-Bicycle>

It's not that simple!



Project Overview



Week 1...

Literature Review

- Bicycle fitting
- Camera calibration
- Human pose estimation via DNNs



A vision based anthropometric Scanner

1330 IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 22, NO. 11, NOVEMBER 2000

A Flexible New Technique for Camera Calibration

Zhengyou Zhang, Senior Member, IEEE

Abstract—We propose a flexible new technique to calibrate a camera. It only requires the camera to observe three pattern planes at two or three different orientations. Either the camera or the plane pattern can be freely moved. The motion need not be the same for the two planes. The proposed procedure consists of a closed-form solution, followed by a nonlinear refinement based on the maximum likelihood criterion. Both internal and external parameters can be estimated. Compared with traditional techniques, which need expensive and tedious hardware to be used to test the proposed technique, our very good results have been obtained. Compared with traditional techniques, the proposed technique is easy to use and flexible. It addresses 3D computer vision more than laboratory environments in real-world use. The corresponding software is available from the author's Web page.

Index Terms—Camera calibration, calibration from planes, 3D pattern, flexible plane-based calibration, absolute scale, perspective mapping, vanishing lines, closed-form solution, maximum likelihood estimation, flexible setup.

1 MOTIVATIONS

CAMERA calibration is a necessary step in 3D computer vision in order to extract metric information from 2D images. Much work has been done, starting in the photogrammetry community [2], [4] to cite a few, and more recently in computer vision [5], [6], [23], [7], [25], [24], [14], [6] to cite a few. We can classify those techniques roughly into two categories: photogrammetric calibration and self-calibration.

• **Three-dimensional reference object-based calibration.** Camera calibration is performed by observing a calibration object whose geometry in 3D space is known with very good precision. Calibration can be done very efficiently [5]. The calibration object usually consists of two or three planes orthogonal to each other. Sometimes a plane undergoing a precisely known translation is also used [2]. These approaches require an expensive calibration apparatus, and an elaborate setup.

• **Self-calibration.** Techniques in this category do not use any calibration object. Just by moving a camera in a static scene, the rigidity of the scene provides in general two constraints [14], [4] on the camera's internal parameters from one-camera displacement by using image information alone. Therefore, 4 images are taken by the same camera with fixed internal parameters, correspondence between those three images are sufficient to recover both the internal and external parameters which allow us to reconstruct 3D structures up to a similarity [15], [12]. While this approach is very flexible, it is not yet mature [1]. Because there are many parameters to estimate, we cannot always obtain reliable results.

Other techniques exist: vanishing points for orthogonal directions [3], [20], and calibration from pure rotation [11], [20].

• *The author is with Microsoft Research, One Microsoft Way, Redmond, WA 98072-8859. E-mail: zhangym@microsoft.com.*
 Manuscript received 26 Apr. 2000; revised 25 Aug. 2000; accepted 7 Sept. 2000.
 Recommended for acceptance by A. Shaohui.
 For information on obtaining reprints of this article, please send e-mail to: tpm@computer.org, and reference IEEECS Log Number 11106.

0892-3530/00/11106-08\$14.00/0

Our current research is focused on a desktop vision system (DVS) since the potential for using DNNs is large. Cameras are becoming inexpensive and ubiquitous. A DVS is one of the general public who are not experts in computer vision. A typical computer user will perform vision tasks only from time to time, so they will not be willing to invest money for expensive equipment. Therefore,

This ICVJ paper is the Open Access version, provided by the Computer Vision Foundation. Except for this watermark, it is identical to the version available on IEEE Xplore.

Flowing ConvNets for Human Pose Estimation in Videos

Thomas Pfister, James Charles, Andrew Zisserman

Dept. of Engineering, University of Cambridge, Cambridge CB2 3PQ, UK

PRACTICAL MANAGEMENT

Road Bicycle Fit

Marc R. Silberman, MD,* David Welner, MD,† Steven Collins, MD,‡ and Brian J. Shipk, DO

(Clin J Sp

This CVPR2014 paper is the Open Access version, provided by the Computer Vision Foundation. The authoritative version of this paper is available on IEEE Xplore.

DeepPose: Human Pose Estimation via Deep Neural Networks

Alexander Toshev, toshave@google.com, Google

Christian Szegedy, szegedy@google.com, Google

Fig. 1. Besides extreme variability in articulations, many of the joints are barely visible. We can guess the location of the right arm in the left image only because we see the rest of the pose and anticipate the motion or activity of the person. Similarly, the left body half of the person on the right is not visible at all. These are examples of the need for holistic reasoning. We believe that DNNs can naturally provide such type of reasoning.

Abstract

We propose a method for human pose estimation based on Deep Neural Networks (DNNs). The pose estimation is formulated as an DNN-based regression problem towards body joints. We present a cascade of such DNN regressors which results in high precision pose estimates. The approach has the advantage of reasoning about pose in a holistic fashion and has a simple but yet powerful formulation which capitalizes on recent advances in Deep Learning. We present a detailed empirical analysis with state-of-art or better performance on four academic benchmarks of diverse real-world images.

1. Introduction

The problem of human pose estimation, defined as the problem of localization of human joints, has enjoyed substantial attention in the computer vision community. In Fig. 1, one can see some of the challenges of this problem – strong articulations, small and barely visible joints, occlusions and the need to capture the context. The main stream of work in this field has been motivated mainly by the first challenge, the need to search in the large space of all possible articulated poses. Part-based models tend themselves naturally to model articulations ([1], [3])

and in the recent years a variety of models with efficient inference have been proposed ([6], [8]).

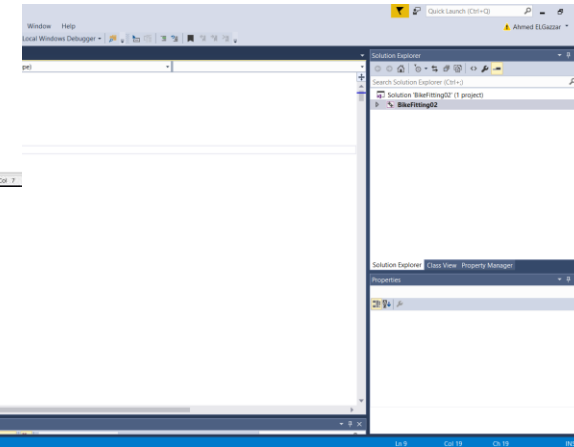
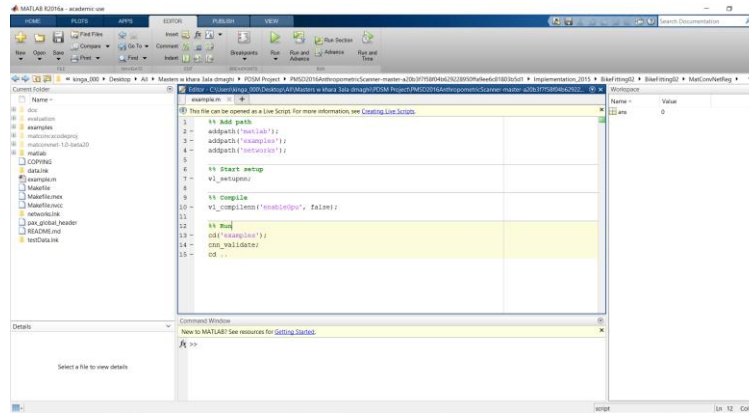
The above efficiency, however, is achieved at the cost of limited expressiveness – the use of local detectors, which reason in many cases about a single part, and most importantly by modeling only a small subset of all interactions between body parts. These limitations, as exemplified in Fig. 1, have been recognized and methods reasoning about pose in a holistic manner have been proposed [15], [20] but with limited success in real-world problems.

In this work we describe to this holistic view of human pose estimation. We capitalize on recent developments of deep learning and propose a novel algorithm based on a Deep Neural Network (DNN). DNNs have shown outstanding performance on visual classification tasks [4] and more recently on object localization [2], [7]. However, the question of applying DNNs for precise localization of articulated objects has largely remained unanswered. In this paper we attempt to cast a light on this question and present a simple and yet powerful formulation of holistic human pose estimation as a DNN.

We formulate the pose estimation as a joint regression problem and show how to successfully cast it in DNN settings. The location of each body joint is regressed to using as an input the full image and a 7-layered generic convolutional DNN. There are two advantages of this formulation. First, the DNN is capable of capturing the full context of each body joint – each joint regressor uses the full image as a signal. Second, the approach is substantially simpler to formulate than methods based on graphical models – no need to explicitly design feature representations and detectors for parts, no need to explicitly design a model topology and interactions between joints. Instead, we show that a generic convolutional DNN can be learned for this problem. Further, we propose a cascade of DNN-based pose predictors. Such a cascade allows for increased precision of joint localization. Starting with an initial pose estimation, based on the full image, we learn DNN-based regressors which refine the joint predictions by using higher resolution sub-images.

* Show state-of-art results or better than state-of-art on

Week 2...



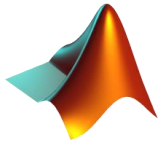
Getting in action

- Procedure breakdown
- Understanding previous codes
- Debugging & Executing



Procedure breakdown

Network



- Training
- Testing
- Packing

Software
Protocol



- Taking picture
- Pose estimation
- Calculations

Visualization



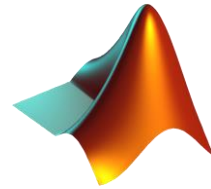
- Visualizing the results



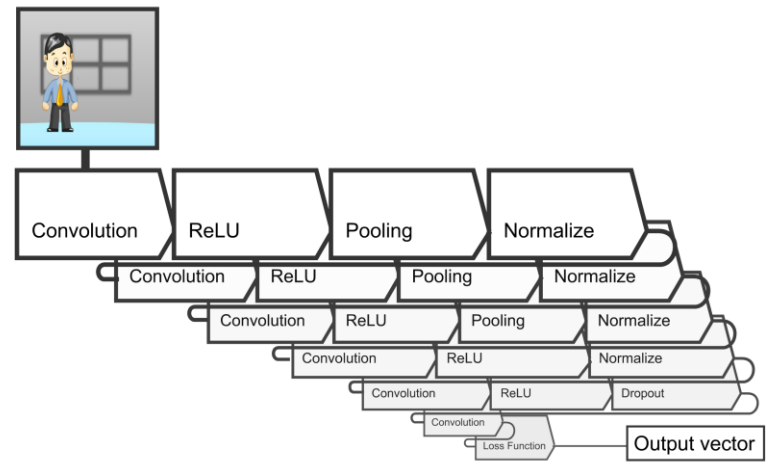
Week 3...

Network

- Implemented in Matlab using MatConvNet.



- Convolutional Neural Network.

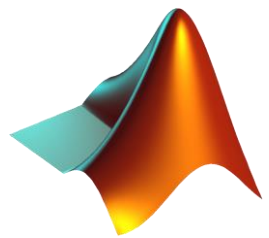


- Trained using the Leeds Sports Pose Dataset.






Week 4...

Software Protocol

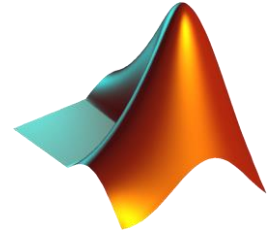


Camera Calibration

-  Tens of images with Bicycles.
-  Tires detected.
-  Tires radiuses and centers distance calculated.



Software Protocol



Real pose estimation



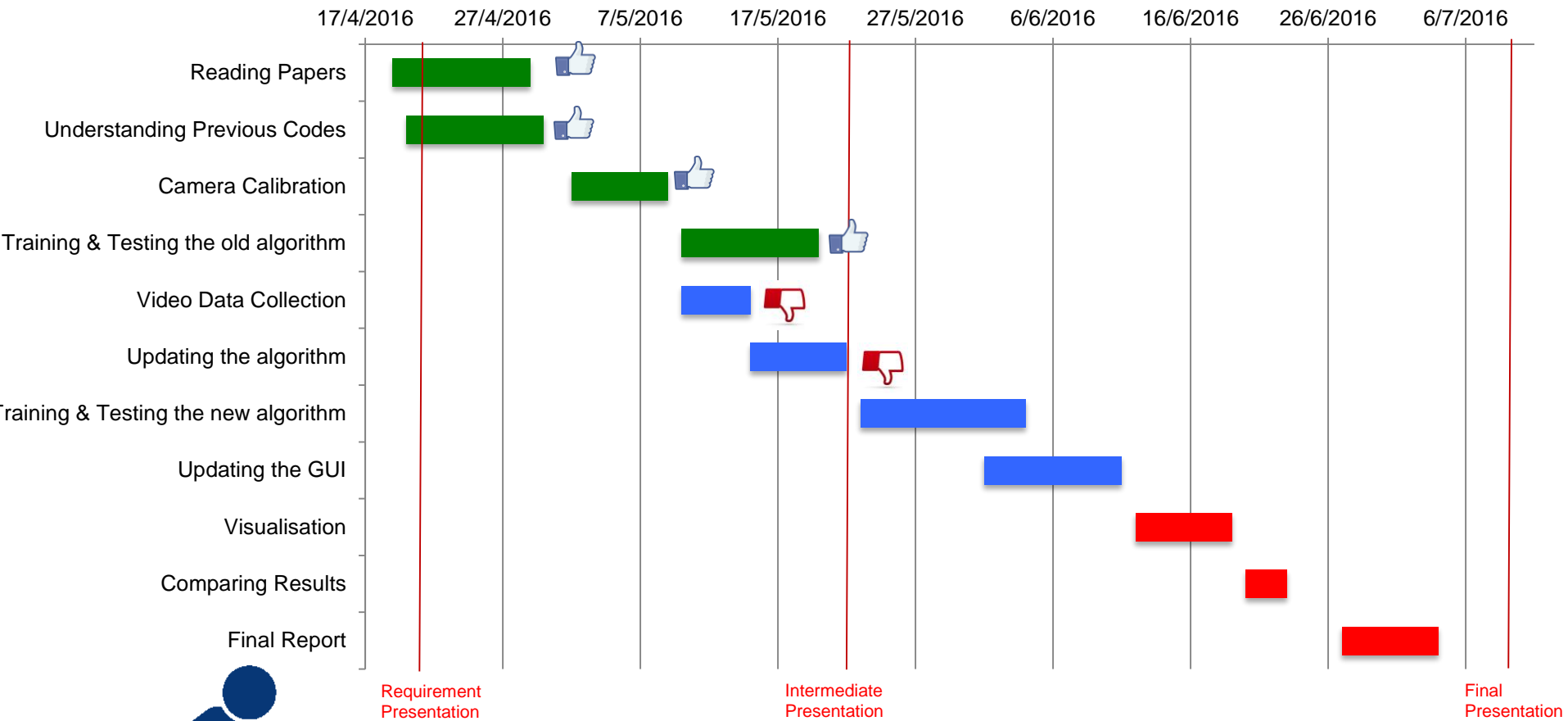
Bicycle detected



Calculation in relation to bicycle real measurements.



Gantt Chart



Created with Microsoft Excel



 **Thank you!** 

