On Estimation of Parameters from Real Scene in Multi-Camera Environment

Ondrej KAINZ1, Dávid CYMBALÁK2

1 Dept. of Computers and Informatics, Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic
ondrej.kainz@cnl.sk, david.cymbalak@cnl.sk

Abstract. In this paper a novel methodology for estimation of parameters from real-world scene in multi-camera environment is presented. Specific focus is put on the estimation of parameters from human body, i.e. focal point is the field of anthropometry and its application in the computer vision.

We utilize and employ known proportions of human body in our system. Based on several of such parameters and together with the output of the multi-camera system we approximate several of human body parameters. Deployment of complementary object with known dimensions is to enhance the final calculations. This object is created based on 3D computational model and thus further mapping of 3D model to 2D scene is possible. Final solution may help in several use case scenarios, e.g. complementary aid in human recognition or replacement of measurements for statistical purposes.

Keywords
Anthropometry, computer vision, multi-camera system, ratio.

1. Introduction

Extracting the information from the static or dynamic image is an important process that may serve for various purposes. This paper mainly focuses on particular information to be extracted and as well particular object – i.e. human body dimensional parameters and person, respectively. Obtaining data of person from image is rather difficult task that includes several steps. In this paper we focus mainly on the techniques where known object with known dimensions is present in the image and further it is used as a complementary information for the approximation to real values. Once implemented in a real environment, obtaining of dimensional parameters of scene should be relatively fast. Several utilizations of proposed methodology are considered – statistical measurements may be carried out more efficiently and faster or being able to extract information about body dimensions from image may be used as a complementary input in person identification. Other use cases are taken into consideration as well, however these are not the principal drive for development of such system.

2. Golden Section in Human Body

Even though not always visible, there is an order in a nature and in the world of man, hence it is possible to express some of these relations by mathematics. Φ or golden ratio or section is one of such relations and may be also found in a human body but before going deeper into the topic we should define what this section is. Φ is actually irrational algebraic number. Two quantities are in golden ratio if ratio of larger, say 𝑎, to smaller, say 𝑏, is equal to ratio of 𝑏 to difference of 𝑎 and 𝑏, i.e. 𝑎/𝑏 = 𝑏/(𝑎−𝑏). It may be also “approximated by the ratio of successive terms in the Fibonacci sequence; in fact, 𝐹(𝑛+1)/𝐹(𝑛) gets closer and closer to Φ as 𝑛 tends to infinity”, [1]. Thus:

\[
\frac{1}{1-\phi} = \phi
\]

where Φ = 1.618.... A rather older research from the late 70’s was carried out by [2], the aim was to estimate several body proportions and compare it golden ratio as depicted in Fig. 1. For most of measurements authors concluded approximation to golden ratio to first decimal place of its value. Due to facial symmetry it is also possible to estimate golden ratio.

![Fig. 1. Body symmetry. Based on [3].](image-url)
Human body as a whole is not the only place that can be found to have some correlation with Φ when it comes to comparisons, as well human face contains some proportions that do correlate with this number, as well there is some symmetry and human face may be divided into sections, for illustration see Fig. 2. This proportions were utilized in the research by [4] where the estimation of height was based on facial golden proportions. In the 2 are shown two types of proportions, face divided into three thirds and the next one following golden rule. Author of study states that more consistent is the latter one.

3. State of the Art

Many cases described next are focused on the height estimation and closely to our idea is study carried out by [5] – goal in this research was the estimation of human height based on single uncalibrated image, while no reference of camera viewpoint nor were calculations of camera parameters available. Body landmarks mapping was done semi-automatically, providing the output of estimation only of height, note that no reference length was used in estimating the height. Utilization, in the case of calibrated camera, was carried out by research introduced in [6] – due to utilization of calibrated camera, intrinsic parameters were available. Again calculation of height was carried out based on top and bottom estimation of extreme points, study does no state utilization of any anthropometric parameters. Single image with human face was used by [4] with aim of estimating the height via eyes, lips and chin extraction and further statistical measurement sets and facial golden proportion deployment. In [7] authors estimated height based on pitch angle and vanishing point of camera with good accuracy. Research related to demography and utilization of calibrated camera is described in [8], authors extensively utilized anthropometric data and were estimating age and gender based on height.

3.1 Body Representation

We consider human body representation to be a crucial when estimating parameters from image. This is due to mapping of predefined human body model to static image. Such model can be expressed as a connection between edges and vertices, where edges represent constant lengths (or bones) and vertices are joints that interconnect specific bones, further they provide some degree of freedom. Model of this kind is used extensively in computer vision or graphics, e.g. [9]. Authors of just stated study designed a system referred to as SASS (Spreadsheet Anthropometry Scalin System), its function is generation of human figure having a population data as an input.

In Fig. 3 is illustrated an example of just described model consisting from vertices and edges. We consider vertex to be synovial type of joints and diarthrosis, e.g. knee or elbow (see rk, lk and re, le, respectively). Synovial types of joints are freely moveable, with specific level of movability. Note that majority of vertices are diarthrosis.

4. System Design

Entire system consists of several processes linked on to each other. Static or dynamic image may be considered, in Fig. 4 is shown the diagram describing the system basic functionality – single or multiple views may be considered. Once the image is loaded, the corrections of known distortions, e.g. lens or perspective distortions are carried out. Next step is the identification of person, in real-time environment and as well in the posterior processing was
used Histogram of Oriented Gradients (for short HOG) algorithm, which is considered to be the best solution for this system. Further part-based model detection divides human body to specific parts which enables mapping of 3D skeleton model on person. This results in first approximation of parameters are based on the outputs of mapping. Second approximation is more complex since several steps precede it – extraction of proportions that includes above introduced golden ratio proportions and as a complementary input is utilized size of known object. This object was first designed and calculated and later on it was created and deployed in the scene. Reference object is basically rectangular cuboid and is used to calibrate the camera with the goal to estimate the unknown size of objects in the image or provide complementary dimensional information. Fundamental presumption for this device is to be present in the image. Illustration of developed object is depicted in Fig. 5. 3D model of an object is loaded into to the static or dynamic scene based on extraction of feature points from the real world device, once located mapping is of 3D model into the image is carried out. This process presumes detection of human body and face was already done. Placement of an object is crucial, if placed correctly it may help to estimate plane and be used as reference dimensional point. These data may then serve as an input for metric calculation of multiple camera sources.

5. Building of Skeleton Model

Previous approaches depend heavily on utilization of static images that are used for calculation of the proportions of vertices. Our novel methodology approach takes the advantages of using video for calculating the vertices by dividing the computing capacity to multiple camera sources via searching the diffractive points of human body. These points are evaluated by optical flow algorithm while the examining the walking or moving habits using computer visions techniques, see Fig. 6.

In order to obtain the values of selected vertices of skeleton model, the process of calculation of the distance between diffractive points is done, presuming these points are detected. See in Fig. 6, selected values are initially set based on the first camera view and are further updated by next cameras view to the final model related to reference object.

Result of overall calculations is formation of vector of selected values of vertices, for Cam 1 as illustrated in the Fig. 6:

\[
H'_n = [HD_n, RY_n, LY_n, UT_n, LUA_n, RLA_n, \ldots]
\]

This vector is further utilized in the real-time for tracking and prediction of the movement of specific person. Solution was rudimentary tested in experimental laboratory environment by using the vMix software system and also the cloned HDMI input cards and scripts for evaluation of the skeleton model vector (Fig. 7).
6. Conclusion and Future Research

Framework for estimating the parameters of real-world scene in multi-camera environment was presented, having as the principal object of interest human body and its parameters, i.e. field of anthropometry.

Real-world and computational model of prototype device was proposed, implemented and is used as a reference object for the extraction of parameters from real-world scene. Proposed reference object is placed in the view field of camera system, once mapping of its 3D model is carried out, the data is used to enhance the overall calculations. Estimation of anthropometric proportions is possible due to homogeneity in the human body and both human body and facial proportions were partially estimated and further utilized from multi-camera sources. Overall system consists of several steps, part presented in this paper is one of them. When merged with the other system components the extraction of parameters is expected to be converge to real values. In the next part of research other components of system are to be deployed and tested together with proposed solution, providing the possibility to extract the parameters of the entire scene.

Acknowledgements

Research described in the paper was supervised by Assoc. Prof. Ing. Frantisek Jakab, PhD., Computer Networks Laboratory at DCI FEEI TU in Košice. We support research activities in Slovakia/This project is being co-financed by the European Union. Paper is the result of the Project implementation: University Science Park TECHNICOM for Innovation Applications Supported by Knowledge Technology, ITMS: 2622020182, supported by the Research & Development Operational Programme funded by the ERDF.

References


About Authors...

Ondrej KAINZ was born in 1988. In 2013 he graduated (MSc. in Applied informatics) from the Technical university in Kosice, Slovakia. Since the very same year he is a PhD student at the Department of Computers and Informatics of the Faculty of Electrical Engineering and Informatics at the Technical university of Kosice. His scientific research interests include computer vision, e-learning, human-computer interfaces, computer graphics, computer networks, biological engineering and body area network.

David CYMBALÁK was born in 1987. In 2011 he graduated (MSc) at the Department of Computers and Informatics of the Faculty of Electrical Engineering and Informatics at Technical University in Kosice. Since 2011 he is PhD student at the Department of Computers and Informatics. His scientific research is focusing on streaming technologies, e-learning solutions or object tracking methods. In addition, he also investigates questions related with the optimization of mobile access to multimedia sources.