Available online at www.sciencedirect.com
SCIENCE DIRECT®
Image and Vision Computing 24 (2006) 343–356
www.elsevier.com/locate/imavis

Zernike velocity moments for sequence-based description of moving features

J.D. Shutler [a,*], M.S. Nixon [b]

[a] Remote Sensing Group, Plymouth Marine Laboratory, Prospect Place, Plymouth, PL1 3DH, UK
[b] Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK

Received 20 September 2004; received in revised form 14 December 2005; accepted 15 December 2005

Abstract

The increasing interest in processing sequences of images motivates development of techniques for sequence-based object analysis and description. Accordingly, new velocity moments have been developed to allow a statistical description of both shape and associated motion through an image sequence. Through a generic framework motion information is determined using the established centralised moments, enabling statistical moments to be applied to motion based time series analysis. The translation invariant Cartesian velocity moments suffer from highly correlated descriptions due to their non-orthogonality. The new Zernike velocity moments overcome this by using orthogonal spatial descriptions through the proven orthogonal Zernike basis. Further, they are translation and scale invariant. To illustrate their benefits and application the Zernike velocity moments have been applied to gait recognition—an emergent biometric. Good recognition results have been achieved on multiple datasets using relatively few spatial and/or motion features and basic feature selection and classification techniques. The prime aim of this new technique is to allow the generation of statistical features which encode shape and motion information, with generic application capability. Applied performance analyses illustrate the properties of the Zernike velocity moments which exploit temporal correlation to improve a shape's description. It is demonstrated how the temporal correlation improves the performance of the descriptor under more generalised application scenarios, including reduced resolution imagery and occlusion.

Keywords: Statistical moments; Zernike moments; Motion; Velocity moments; Gait

1. Introduction

Moving object description is a growing area of computer vision research, traditionally an arena dominated by tracking algorithms. The developments in this area were previously limited not least by the storage requirements of image sequences. With the advance of digital video (DV), and the explosion of storage capacities, the analysis and storage of image sequences has become viable, enabling increased interest. Tracking algorithms [1] generally locate the region or feature of interest in the first frame and then track it throughout the remainder of the sequence. This requires good initialisation in the first image and assumes that in later images tracked objects are not overcome by noise or occlusion. This kind of approach enables real-time performance, a major benefit of these algorithms. With the ever increasing available computing power, alternative approaches that process the complete image-sequence are appearing. For example, the velocity Hough transform for conic sections [2] and its extension for arbitrary shapes [3] process a complete image sequence, overcoming the problems of image noise and occlusion by exploiting temporal correlation, treating the image sequence as a single entity rather than individual images. These approaches locate the perimeter of a moving shape by searching for a particular motion. However, a great deal of information can be held within a shape's perimeter—motivating techniques enabling holistic moving shape description.

Statistical moments, e.g. [4] describe a shape with respect to its axes, producing holistic descriptions encoding information including mass, centroid and variation across axes. Mukundan [5] provides descriptions of most of the current moment techniques, along with background information and applications. In general the different types of moments fall into two categories, orthogonal and non-orthogonal. Orthogonal moments produce features that are less correlated than their non-orthogonal counterparts. Further, the orthogonality property enables simple, accurate signal reconstruction from the generated moments. Moments that are non-orthogonal tend

* Corresponding author. Address: Remote Sensing Group, Plymouth Marine Laboratory, Prospect Place, Plymouth, PL1 3DH, UK.
Tel.: +44 1752 633417.
E-mail addresses: jams@pml.ac.uk, jamie@zepler.org (J.D. Shutler).

0262-8856/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2005.12.001

to be simpler to implement, computationally less expensive and include descriptors that have a range of useful properties, i.e. scale, translation and rotation invariance. Their highly correlated features (as a result of their non-orthogonal nature) make reconstruction more difficult. This correlation requires the need for high accuracy in the calculations when interested in the high frequency components of the image and/or when analysing large datasets.

There have been many studies using two-dimensional moments for image recognition purposes. However, to date, most applications use single images. Hoey [6] used Zernike polynomials to study facial motion by generating flow fields which provided input to hidden Markov models. Little [7] used moments to characterise optical flows between images for gait recognition. These techniques still only link adjacent images, and do not consider the complete sequence. Rosales [8] described motion by producing one image that contained information from a complete sequence, building on the work done by Davis [9]. Rosales's system was based on Hu [10] invariant moments and was used to recognise types of motion, e.g. sitting down or kicking; due to several images being compressed into one, subtle differences between subjects are lost due to self occlusion and overlapping of data.

For this work, we began by looking at a traditional statistical method of moments to describe the motion of a person through multiple images. Unfortunately, this does not provide a very detailed description of the motion, as there is no information linking the images of the sequence, since they are treated as separate entities. By using the general theory of moments a method has been developed that not only contains information about the pixel structure of the moving object, but also how its movement flows between images. Through analysing image sequences the temporal information can be exploited and the possibility of describing deforming shapes becomes apparent. Accordingly, we describe a new technique called velocity moments, enabling the holistic statistical description of temporal image sequences. We present this new technique to enable the application of statistical moments to image sequences. To aid its characterisation while demonstrating its beneficial attributes, we apply it to human gait recognition, an emergent biometric.

This paper is structured as follows. Firstly, Section 2 briefly reviews non-orthogonal and orthogonal statistical moments. Velocity moments are then introduced in Section 3. Section 4 uses human gait classification to illustrate their application. Section 5 details the performance attributes of the Zernike velocity moments analysing the effects of reduced resolution imagery and occlusion. Conclusions are then drawn.

2. Background theory

Statistical moments are applicable to many different aspects of image processing, ranging from invariant pattern recognition and image encoding to pose estimation. Moments of an image [10], describe the image content (or distribution) with respect to its axes. They are designed to capture both global and detailed geometric information about the image. In continuous form an image can be considered as a two-dimensional Cartesian density distribution function $f(x,y)$. With this assumption, the general form of a moment of order $(p+q)$, evaluated over the complete image plane ξ is:

$$M_{pq} = \iint_{\xi} \psi_{pq}(x,y) f(x,y)\,\mathrm{d}x\,\mathrm{d}y; \quad p,q = 0,1,2,\ldots,\infty \tag{1}$$

The *weighting kernel* or *basis* function is ψ_{pq}. This produces a weighted description of $f(x,y)$ over the entire plane ξ. The basis functions can have a range of useful properties that may be passed onto the moments, producing descriptions which can be invariant under rotation, scale, translation and orientation. For image analysis a discrete version is required, for this conversion we assume that ξ is divided into square pixels of dimensions $\Delta A = 1\times 1$, with constant intensity I over each pixel so $P_{xy} = I(x,y)\Delta A$.

2.1. Non-orthogonal moments

Early work by Hu [10] applied statistical moments to image analysis defining the Cartesian moments which in discrete form are:

$$m_{pq} = \sum_{x=1}^{M}\sum_{y=1}^{N} x^p y^q P_{xy} \tag{2}$$

Extending them to include translation invariance Hu defined the Centralised moments

$$\mu_{pq} = \sum_{x=1}^{M}\sum_{y=1}^{N} (x-\bar{x})^p (y-\bar{y})^q P_{xy} \tag{3}$$

where M and N are the image dimensions, $p+q$ is the order and P_{xy} is the pixel value at position (x,y). $\bar{x}$ and $\bar{y}$ are the x and y centres of mass (COMs)

$$\bar{x} = \frac{m_{10}}{m_{00}} \qquad \bar{y} = \frac{m_{01}}{m_{00}} \tag{4}$$

which describe a unique position within the field of view.

Cartesian moments, Eq. (2) are formed using a monomial basis set $x^p y^q$. This basis set is non-orthogonal and this property is passed onto the Cartesian moments. These monomials increase rapidly in range as the order increases, producing highly correlated descriptions. This can result in important descriptive information being contained within small differences between moments, which can lead to the need for high computational precision.

2.2. Orthogonal moments

Moments produced using orthogonal basis sets also exist. These orthogonal moments have the advantage of needing lower precision to represent differences to the same accuracy as the monomials. The orthogonality condition also simplifies the reconstruction of the original function from the generated moments as each descriptor (or moment) is independent (uncorrelated). Many orthogonal sets exist (Legendre, Zernike,