Stereo Image Analysis: A New Approach Using Orthogonal Moments

R. Mukundan
Department of Computer Science
University of Canterbury
Christchurch, New Zealand.
(mukund@cosc.canterbury.ac.nz)

Angeline Pang Nyuk Khee
Faculty of Information Technology
Multimedia University,
Cyberjaya, Malaysia
(angeline@mmu.edu.my)

Abstract

This paper presents the mathematical framework of discrete orthogonal moment functions based on Tchebichef polynomials, and provides the conceptual ideas on applying such moments to the problem of stereopsis. Stereo vision algorithms generally use a set of feature descriptors characterizing the image intensity distribution within a small window centered at the pixel being analyzed, to determine the disparity value, and thereby to estimate the depth of the image point at that pixel. Discrete orthogonal moments have several favorable properties making them good candidates for such applications. Some preliminary results obtained using a window based matching algorithm are presented. Apart from being able to represent independent image features, the capability of discrete orthogonal moments to reconstruct the intensity distribution from a moment set could perhaps be utilized in developing a coarse-to-fine disparity estimation algorithm.

1 Introduction

Estimating the depth information from a pair of stereo images is an important problem in computer vision and photogrammetry. The depth information can be translated into a closely related information of pixel disparities, based on the stereo camera geometry, and noting that corresponding pixels in the stereo image are projections of the same point in the three-dimensional scene. Stereo vision (or stereopsis) thus generally means the correspondence problem of finding for each point in one image, the matching point in the other.

Even though depth estimation is the primary aim of stereopsis in many computer vision and robot vision applications, stereo image analysis is also useful in a wide range of other areas such as the estimation of surface orientation [1,2], 3D motion estimation [3,4], 3D surface description[5,6], and information fusion. Stereopsis is also used as a tool in view synthesis which consists of processes for synthesizing new views from existing images of a scene [7,8].

The matching of corresponding areas in a stereo image pair can be done in many ways. Feature based methods establish the correspondence by using a set of extracted features such as edges or image discontinuities [9]. Area based algorithms compare the intensity values within a square window centered at a point on one image with the corresponding values in an identical square centered at points on the epipolar line of the other image. Researchers have used different types of cost functions to obtain an optimal matching set of windows. The squared sum differences (SSD) is a classic example [8], where the sum of squared intensity differences across the window is used as a measure of dissimilarity. The cross correlation of intensity distribution is another example. Marr and Poggio [10] introduced a cooperative algorithm to computer stereo disparity by applying a set of global constraints.

Moment functions represent global shape characteristics in an image, and have been used in many image analysis applications such as pattern recognition, pose estimation, and image classification [11]. A set of moment functions computed on the intensity values within a square window of one image can be compared with the corresponding function values on windows of the other image to establish a correspondence between regions of matching intensity distributions. A moment based sliding window algorithm is given in [12]. Most of such moment based algorithms use geometric moments for fast computation of the image features. However, geometric moments are highly sensitive to image noise, and also exhibit large variation in the dynamic range of values, which can yield wrong results in a matching algorithm.

Discrete orthogonal moments using Tchebichef polynomials as basis functions were recently introduced [13] as a feature descriptor that eliminates many problems associated with geometric as well as continuous moment functions like Zernike and Legendre moments. Two important characteristics of Tchebichef moments, *viz.*, (1) a discrete domain of definition which matches exactly with the image coordinate space, and (2) absence of numerical approximation terms, allow a more accurate representation of image features than otherwise possible using conventional moments. As a result, images can be exactly reconstructed from a complete set of discrete moments.

This paper details a few of the conceptual ideas related to the application of discrete orthogonal moments to stereopsis. An area based algorithm similar to that of the SSD, where the Euclidean distance between two moment vectors is used as a similarity measure, is presented. The preliminary results are encouraging, and could be further improved by imposing additional constraints on disparity gradients, such as those employed by Marr and Poggio[10]. The motivation for this work is derived from the capability of discrete orthogonal moments to represent independent feature characteristics of the image. It is therefore expected that a matching algorithm using a suitably defined

feature vector would perform better than conventional SSD or SAD (Sum of Absolute Differences) algorithms. As described in the next section, Tchebichef moments could also be used to reconstruct the intensity function from a moment set. This feature could also be used in developing robust sterepsis techniques based on a coarse-to-fine disparity estimation strategy. Some of the conceptual ideas towards such a goal are also given in this paper.

2   Tchebichef Moments

The Tchebichef polynomials of degree n are defined over a discrete coordinate space $x \in [0,1,2...N{-}1]$ as follows [14,15]:

$$t_n(x) = (1-N)_n \, {}_3F_2(-n, -x, 1+n;\, 1, 1-N;\; 1),\qquad n, x = 0,1,2,..N{-}1.$$

(1)

where $(a)_k$ is the Pochhammer symbol given by

$$(a)_k = a\,(a{+}1)\,(a{+}2)\,...(a{+}k{-}1),$$

(2)

and ${}_3F_2(.)$ is the generalized hypergeometric function,

$$
{}_3F_2(a_1, a_2, a_3; b_1, b_2; z) = \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k (a_3)_k}{(b_1)_k (b_2)_k} \frac{z^k}{k!}.
$$

(3)

Given a set of discrete image intensity values $f(x,y)$, the Tchebichef moments of order $p{+}q$ are defined as

$$
T_{pq} = \frac{1}{\tilde{\rho}(p,N)\tilde{\rho}(q,N)} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \tilde{t}_p(x)\tilde{t}_q(y) f(x,y),\quad p, q, x, y = 0, 1, 2, ..N{-}1.
$$

(4)

where $\tilde{t}_n(x)$ is a scaled version of $t_n(x)$ given by

$$\tilde{t}_n(x) = t_n(x) / N^n,$$

(5)

and

$$
\tilde{\rho}(n, N) = \frac{(N^2 - 1)(N^2 - 2^2)...(N^2 - n^2)}{N^{2n-1}\; 2n+1}.
$$

(6)

The following inverse moment transform allows us to reconstruct the image intensity distribution from a moment set:

$$f(x, y) = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} T_{mn} \tilde{t}_m(x) \tilde{t}_n(y), \quad x, y = 0, 1, \ldots N-1.$$

(7)

A detailed description of the properties of Tchebichef polynomials and moments, together with their implementation aspects can be found in [13].

3 Stereo Geometry

A simple stereo geometry used in many vision algorithms consists of two identical cameras whose image planes coincide, and whose lateral axes are parallel to the line connecting their centers of projection (Fig. 1).
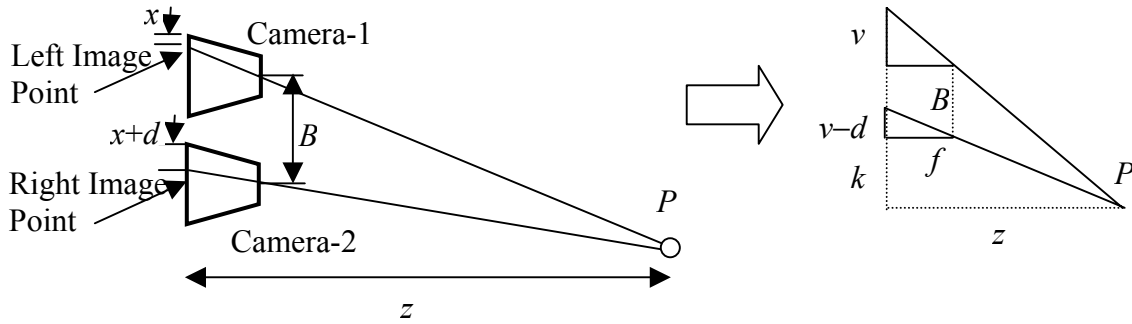

Figure 1. A simple stereo camera geometry

If z denotes the depth of the imaged point P with respect to the image planes, f the focal length, and B the lateral separation of the optical centers of the two cameras, then we can derive the equation relating the depth of the point to the disparity in the image at the corresponding pixel, as follows:

$$\frac{v-d}{f} = \frac{v-d+k}{z}; \qquad \frac{v}{f} = \frac{v+B+k}{z}$$

and therefore,

$$z = \frac{(B+d)f}{d}.$$

(8)

The epipolar line of a pixel on the left image is a line where the matching pixel lies on the right image. The above camera configuration also ensures that the epipolar lines for all pixels on a row is the same row of pixels (*i.e.,* the pixels have the same *y*-coordinate) on

the right image. The search space for the matching pixels can thus be limited to the same row on the right image, without having to explicitly compute the epipolar line for every pixel.

4 The Correspondence Problem

The stereo correspondence problem in area based algorithms is solved by considering a square window centered at pixels on epipolar rows on each image, and by using a similarity measure that compares the intensity values/distribution in both the windows to determine the matching positions (Fig. 2).
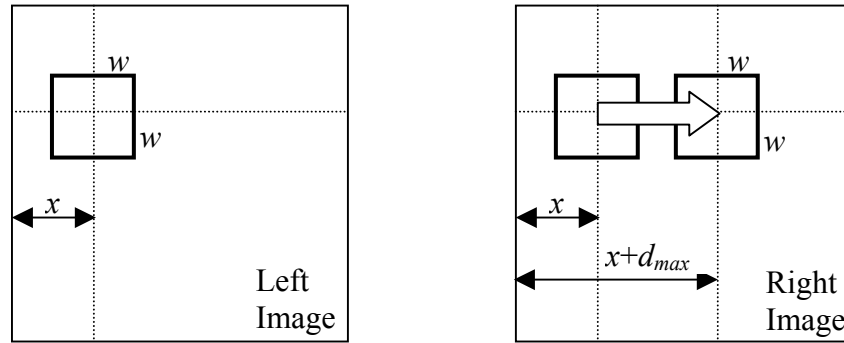

Figure 2. Area based stereo algorithm using local intensity matching.

The window size (w x w) is chosen to be small so that a fairly accurate matching of the local intensity distribution around the pixel currently being analyzed, can be obtained. For every pixel on the left image at a distance x (from the left boundary of the image), window positions from x through $x+d_{max}$ are considered, where d_{max} is the maximum possible disparity value in the whole image. The correspondence between the matching positions of pixels is established by using a similarity measure derived from the intensity values within the windows. If $I(x, y)$ and $I'(x, y)$ denote the intensity values at the pixel (x, y) on the left and the right image respectively, then the sum of squared difference (SSD) algorithm matches the pixel (x, y) on the left image with the pixel $(x+d, y)$ on the right image, where the term

$$\sum_{i=-w/2}^{i=w/2} \sum_{j=-w/2}^{j=w/2} \left(I(x+i, y+j) - I'(x+i+d, y+j) \right)^2$$

(9)

attains a minimum value for $0 \le d \le d_{max}$.

In a moment based stereopsis algorithm, instead of comparing the intensity values directly as in (9), we compare the shape characteristics of the image within the square window. Moments are weighted sum of intensity values, and therefore can provide a

measure, which is more robust than the method using direct intensity matching/correlation, in the presence of image noise.

5 Stereo Correspondence Using a Moment Vector

The basic idea here is create a moment feature vector of size k, $\mathbf{v} = \{T_1, T_2, …,T_k\}$ where each T_i is a moment computed from the intensity values within the window (centered at (x,y)) on the left image, and a similar vector $\mathbf{v_d'} = \{T_1', T_2', …,T_k'\}$ from the intensity values within the window (centered at $(x+d, y)$) on the right image, and then define a similarity measure using $\mathbf{v}$ and $\mathbf{v_d'}$. A similarity measure using the Euclidean distance between the two vectors is given by

$$S_{d,k} = \| \mathbf{v} - \mathbf{v_d'} \| = \sqrt{\sum_{i=1}^{k} \left(T_i - T_i' \right)^2 } \ .$$

(10)

The following are the main considerations in choosing the components of the feature vector:

(i)   The moments should represent some shape characteristic of the global intensity distribution within the window.

(ii)   The moments should not be sensitive to image noise.

(iii)   The components within the feature vector should have uniform dynamic range of values, so that the values of a few components will not dominate over the others in the decision criteria.

(iv)   The feature vector should have sufficient amount of information redundancy.

Geometric moments are the simplest type of moments, but they do not satisfy the requirements (ii), (iii) and (iv) above. On the other hand, discrete orthogonal moments satisfy all the requirements (except possibly (iii)), and in addition allows us to reconstruct the intensity distribution from a sufficiently small set of moments. The condition (iii) can be satisfied by a scale normalization as given in (5).

This paper explores the feasibility aspects related to the application of Tchebichef moments as feature descriptors, since lower order Tchebichef moments have sufficient information content required for a correspondence problem, and each component of a different order would represent a statistically different characteristic of the intensity distribution. It may be noted that two important parameters to be decided here are the size of the window w, and the length of the feature vector k. A large window size increases the complexity of moment computation, looks for near-global features than near-local features, and leaves a large region around the image margins where a proper correspondence could not be established. A window size which is too small will make the feature vector more sensitive to image noise, and will fail to capture essential information

regarding the nature of intensity distribution in the neighborhood of the pixel. A window size which is roughly 10% of the image size is usually found acceptable. The size of the feature vector k depends on how much information the vector should store to ensure a fairly accurate match. Moments of the first few orders (normally up to a maximum order of four) are stored in the feature vector, so that the amount of computation is kept reasonably low, while capturing sufficient information about the intensity distribution within the window.

6 Experimental Results

A set of preliminary experiments were carried out with both binary and gray-level images to gauge the performance of Tchebichef moments. Fig. 3 shows a random-dot stereo image pair with three layers of disparities. The images which have a size of 64x64 pixels, are used as inputs to the Marr&Poggio's cooperative stereopsis algorithm, the conventional SSD algorithm, and the moment based algorithm. A window size of 6x6 pixels was chosen form both SSD and moment algorithms. The gray coded disparity maps obtained are also shown in Fig.3.
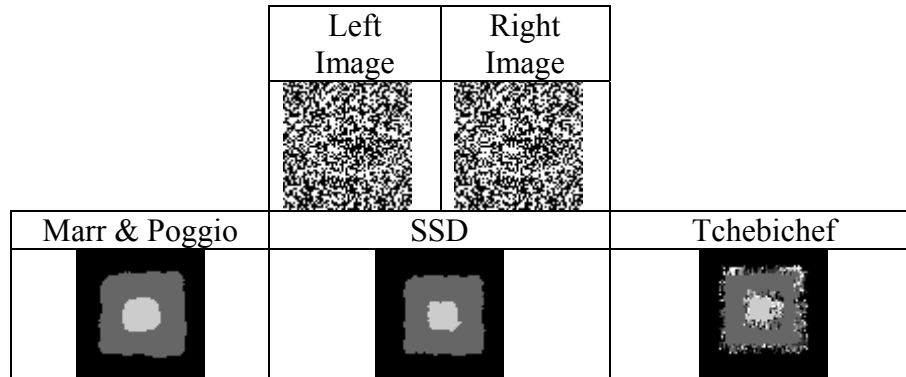

Figure 3. Random-dot stereo images and the disparity maps.

For the moment algorithm, Tchebichef moments up to the second order were used in the feature vector. That is,

$$\boldsymbol{v} = \{T_{00}, T_{10}, T_{01}, T_{20}, T_{02}, T_{11}\}.$$

(11)

The discontinuities seen on the boundaries of the disparity layers obtained from the Tchebichef moment algorithm are acceptable, since those are the regions where a random-dot stereogram will have undefined region correspondence (known as occlusions). The disparity map can be further improved by using well-known smoothing functions, or by introducing global constraints on disparity gradients.

A set of synthetic gray-level images shown in Fig.4 was also used in the analysis, with higher order moment terms. Since Marr&Poggio's cooperative algorithm is not suitable for gray-level images, only the SSD algorithm was used for the comparison of the

outputs. Since the images are of size 128x128 pixels, a window of size 10x10 pixels was used in both the methods.
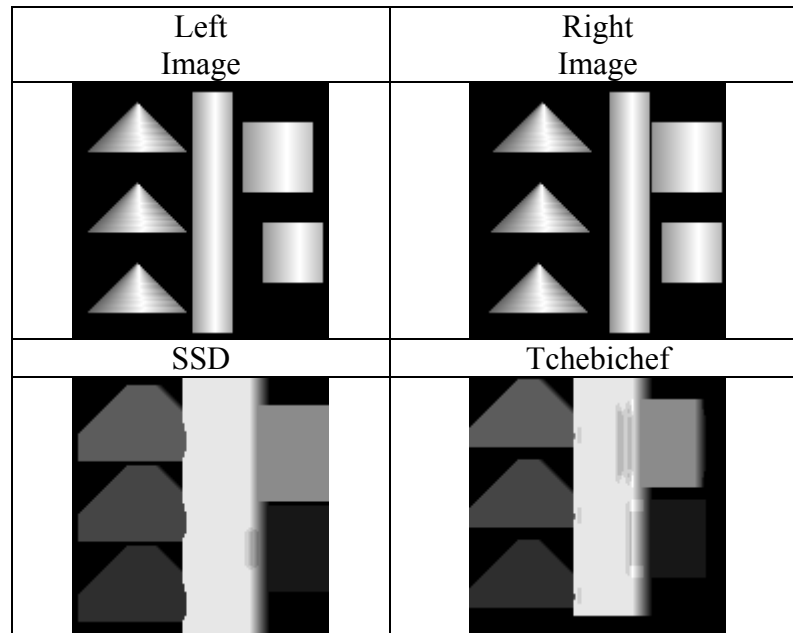

Figure 4. Gray-level stereo images and their disparity maps

The Tchebichef algorithm for the above set of images used moments of order up to 4, in the feature vector. The SSD algorithm generates larger disparity regions around the true values, while the Tchebichef algorithm gives a significantly more accurate result. As mentioned earlier, the results could be further improved by adding continuity constraints on the disparity values.

7 Future Work

The experimental results shown in the previous section show that discrete orthogonal moments could be used for feature correspondence in a stereo vision algorithm. However the true potential of orthogonal moments lies in their capability to capture independent image characteristics, and to reconstruct the image from a moment vector. Possible future extensions of the work presented in this paper, are outlined below.

   i.   Addition of continuity constraints on disparity values will limit the number of mismatched points, especially at image boundaries where large disparity gradients are present.

   ii.   A detailed analysis on the feature representation capability of individual moments and their relative sensitivity to image noise will help in deciding which moment terms need be included in the feature vector, and what should be the optimal size of the feature vector for a given image.

iii.   An set of analytical relations between the moments of the left and the right images in terms of the unknown disparity values at each pixel can be derived and, then solved for the disparity values in closed-form.

iv.   A coarse-to-fine disparity estimation could be attempted using a progressive image reconstruction strategy. Images reconstructed using low order moments would contain coarse global level disparities. As we add higher order moments to improve the quality of the reconstructed images, the disparity values can also be simultaneously propagated to finer pixel levels.

8 Conclusions

This paper has presented a new application area for discrete orthogonal moments, namely, stereo image processing. Being robust feature descriptors representing independent image characteristics, Tchebichef moments also have minimal information redundancy in a feature vector. These qualities were the main motivation factors for choosing a moment vector based similarity measure for a stereo algorithm. A few preliminary results have also been presented, which showed the feasibility of applying moment techniques to stereopsis. Possible extensions of the work using additional capabilities of orthogonal moments in image reconstruction, have been outlined.

References

1.   Kanatani K, Detection of Surface Orientation and Motion From Texture by a Stereological Technique, Artificial Intelligence, Vol. 23, No. 2 (1984), pp. 213-237.

2.   R.Mukundan, Estimation of Orientation Parameters of a Planar Object Using Moment Functions, Proc. Intnl. Symposium on Intelligent Robotics (Bangalore), ISIR-93, pp. 866-875.

3.   Young S.J and Chellappa R, 3D Motion Estimation Using a Sequence of Noisy Stereo Images: Models, Estimation and Uniqueness of Results, IEEE Trans. on Patt Anal and Mach Intell, Vol. 12, No. 8 (1990), pp 735- 759.

4.   Weng J, Motion and Structure Estimation From Stereo Image Sequences, Journal of Robotics Research, Vol. 8, No. 3 (1992), pp. 362- 382.

5.   Steven D. Cochran, 3D Surface Description from Binocular Stereo, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 14, No. 10 (Oct 1992), pp. 981-994.

6.   Dalmia A.K and Trivedi M, High Speed Extraction of 3D Structure of Selectable Quality Using a Translating Camera, Computer Vision and Image Understanding, Vol. 64, No. 1 (1996), pp. 97-110.

7.   Daniel Scharstein, View Synthesis Using Stereo Vision, Springer - Lecture Notes in Computer Science (1999).

8.   S.E.Chen and L.Williams, View Interpoloation for Image Synthesis, Computer Graphics (SIGGRAPH-95), pp. 29-38.

9.   H. H. Baker and T.O. Binford, Depth From Edge and Intensity Based Stereo, Proc. 7th Intnl. Joint Conf. on Artificial Intelligence, Vancouver (1981), pp. 631-636.

10. Marr D and T. Poggio, Cooperative Computation of Stereo Disparity, Science, Vol. 194 (1976), pp. 209-236.

11. R. Mukundan and K.R. Ramakrishnan, Moment Functions in Image Analysis – Theory and Applications, World Scientific (1998).

12. K.R. Ramakrishnan, R. Mukundan, "Estimation of Stereo Disparity Using Moment Functions", The Third Intnl. Conf. on Automation, Robotics and Computer Vision (ICARCV-94) Singapore, Vol. 1, WA2.3, (Nov 1994), pp. 30-34.

13. R. Mukundan, Image Analysis by Tchebichef Moments, IEEE Trans. on Image Processing, 10(9), Sep 2001, 1357-1364.

14. Erdelyi, A. et al , Higher Transcendental Functions, Vol. 2, McGraw Hill, (New York: 1953).

15. Arnold F. Nikiforov, Special Functions of Mathematical Physics, Birkhauser Verlag (1988).