Structured-Image Retrieval invariant to rotation, scaling and translation

C. AVILÉS-CRUZ, A. FERREYRA-RAMIREZ, J. J. OCAMPO-HIDALGO AND I. VAZQUEZ-ALVAREZ
Universidad Autónoma Metropolitana,
Departamento de Electrónica,
Av. San. Pablo 180 Col. Reynosa, Delegación Azcapotzalco, 02200 Mexico D. F.
Mexico
{caviles, fra, jjoh, iva}@correo.azc.uam.mx

Abstract: In Content Based Image Retrieval (CBIR) domain a new methodology is proposed in this paper; this methodology is based on statistical features such as Hu invariant moments (invariant to scale and translation) and a correlation measure (between 2 images). In order to attack image rotation problem, we use principal components analysis. Image comparison is done by correlation. Proposed scheme was tested with hundred of structured images like: electronic circuits, cell phones, cars, trees, leaves, grass, glasses, tables, etc. Features characteristics are extracted from invariant moments, taken from window size estimation. From a query image, a set of features were estimated in order to compare to a set of images. Correlation function is applied to get image similarity, it is obtained a believe percentage value. As a methodology conclusion, we found that the invariant moments combined with principal component analysis gives excellent results in image retrieval task. An exhaustive study was performed with 1000 images. We also evaluated the impact of noise on the images testing additive Gaussian random noise.

Key words: Image retrieval, PCA, Invariant Moments, Pattern Recognition.

1 Introduction

Nowadays, due the availability of large storage spaces a huge number of images can be found in the Internet. With this huge distributed and heterogeneous image database, people want to search and make use of the images there contained. More and more, people want to search a specific image by their contains such as: electronic circuits, cell phones cars, trees, leaves, grass, glasses, tables, planes, maps, animals, buildings, cities etc. An image retrieval have to be in precise form. So search objective of an image in a image-database is that all the context returns to us (the finder) where is susceptible to exist this image, as well as, the greater number of sub images where it appears in. All commercial image searchers are based on finding images by the name of her in the file. Therefore, the results that throw these finders are not efficient when an objective in individual is required, for example: a type of person, an automobile, a tree, etc. The necessity to count on a searcher of images specialized in objects has been increased significantly in the last years in applications such as: military man, biomedicine, commerce, education, among others. At the moment a generic algorithm does not exists a servant far from it that helps to look for images us in Internet solely. Necessity to count on searcher of particularly is interesting if within image could to find for example hotel within map (taking as it bases the icon that represents a hotel), a seat of interest, a museum, etc. In the work presented/displayed by Wang [16] considers a search of images via "a semantics-sensitive" strategy, that in essence makes indexed of images and its attribute-properties yet, which makes the system is not but that a search in a data base of the properties of the image; the algorithm was proven in an application of the type digital bookstore. On the other hand, it exists another approach, Cordelia [17] defines a series of invariant attributes of the levels of grays for the search of images, the algorithm works with invariant moments [1][2][11] that on the basis of them it looks for that an image is equal to another one, we know that the invariant moments are insensible on the scale, rotation and transferring. One of the disadvantages is that it works solely for images in levels of grays and single to look for some type of structured image.

In the present work a methodology for the search of images in Internet sets out. The work includes/understands main phases, the schemes of comparison between two images[1][2][3]. Proposed scheme was tested with hundred of structured images: electronic circuits, phone cell, faces, trees, girls, cats, dogs, etc. The invariant moments technique is used in this study to extract image characteristics by window size estimation. From an entry image, a set a features were estimated in order to compare to a set of images, we obtain a fitness function based on correlation between two images. We found that the invariant moments combined with principal component analysis give excellent results in image identification task. An exhaustive study was performed with 100 images with an without additive Gaussian random noise.
The work images are in format JPG[13], which is the format but common of handling of images in Internet. The rest of the present work is organized as follow; in chapter 2 the methodology of comparison of images will be exposed. Chapter 3 contains the experimental results; finally the conclusions and perspective will be boarded in chapter 4.

2 Image retrieval methodology

Methodology proposed is shown in figure 1, it started from image query, following a image processing procedure (it will explained later), finally a fitness function is applied in order to compare two images, fitness function is based on correlation, as a results a set of images is shown (images seems to close to a query image)

In the image processing block we apply some technique of pattern recognition, in our case, sets out a sequence of stages that will allow obtain image from color to binary images. The used image processing chain appears in figure 2 and it is explained as follow:

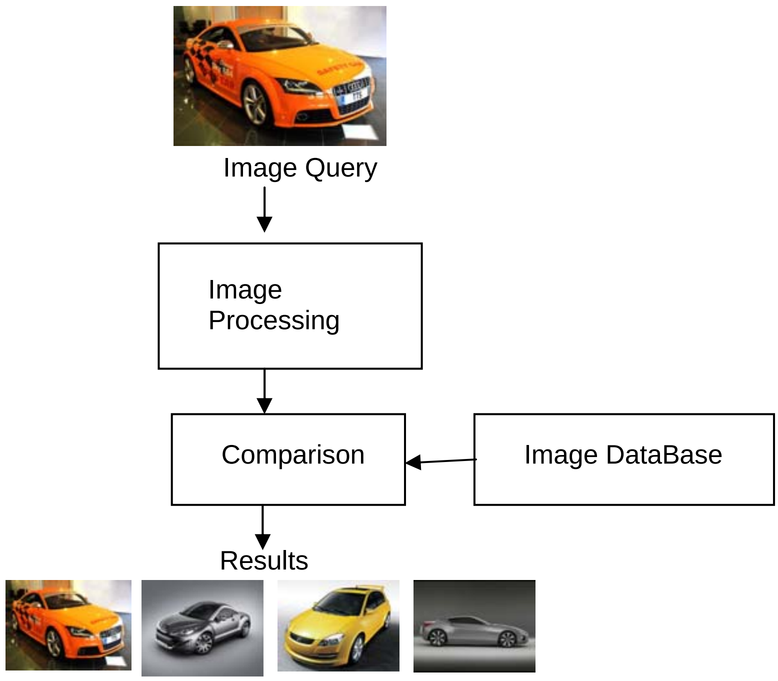


Fig. 1 Scheme of proposed methodology

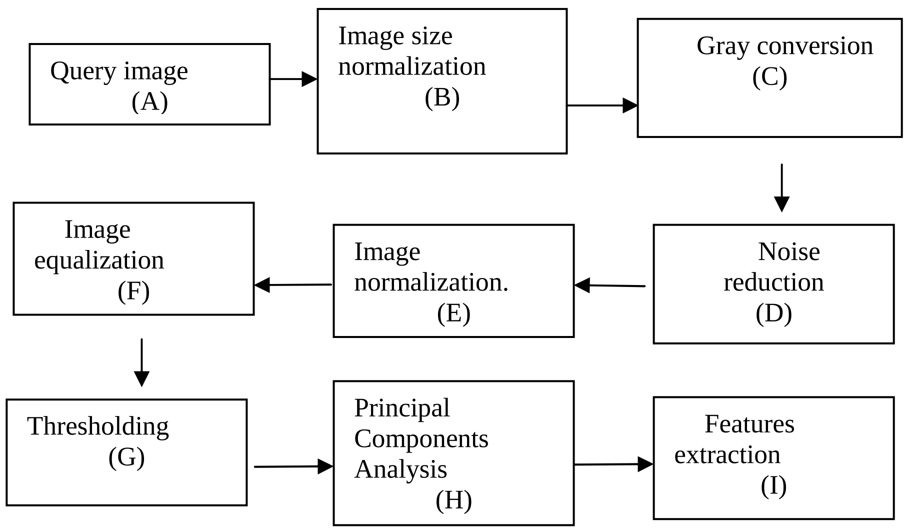


Fig. 2 Image processing sequence in order to compare to images of data base

- **(A) READ IMAGEN**.

The module "read image" is simply the reading of a disk file: opening of the file, reading image information and closing it. The file has been read in RGB format and it could be from any size.

- **(B) IMAGE SIZE NORMALIZATION**

The image can be of any size (NxM). So that we pruned to make the comparison of two images, in our method, these must of being of a same size. The images are scaled to a size of 160x160 pixels; although the size of the scale can be anyone, we chose this value arbitrarily, in the results chapter, it is explained the effect to take other sizes. The conformation of the size is made via a sampling of the information. This block does not introduce any additional color to the scaled image.

- **(C) GRAY CONVERSION**

From an image in color (format RGB), one goes to an image in gray levels, 8 bits quantized.

- **(D) NOISE REDUCTION**

In the present block, the convolution is performed between image and a Gaussian operator, the Gaussian operator has given radius and sigma parameters. The radius must be greater than sigma.

- **(E) IMAGE NORMALIZATION**

As the application of the Gaussian filter of the previous module, introduces pixel values different from rank 0-255, is used the module of image normalization to arrive the values at a rank of [0-255], according to the following equation 1.

$$output\ gray\ pixel = \left(\frac{255-0}{\max - \min}\right)(input\ gray\ pixel) + b \quad (1)$$

Where *max, min* are the maximum and minimum values of the image.

- **(F) IMAGE EQUALIZATION**

In order still more the important information of the image, we needed to emphasize an the contrast. By equalization, we obtain a uniform histogram, distributing the range of tones that appear more by all the histogram.

- **(G) THRESHOLDING**

The objective of this module is to transform the image of levels of grays to a binary image. Finally the thresholding is the one that obtains east result comparing each value with another call threshold to determine if the information is transformed to one logical or to zero logical. The equalization stage is fundamental to be able to make the thresholding stage, because when we do uniform gray values in the histogram, the value threshold can for example be placed to half of the interval from 0 to 255. Thus while the values below the threshold will be taken to zero, those that are over or equal to the value threshold will be taken to one

- **(H) PRINCIPAL COMPONENTS ANALYSIS (PCA)**

In order to characterize each font, principal components analysis is applied over each window estimation. The PCA principle is given in the follow expressions:

Fist we have to estimate C_x, the variance/covariance matrix by the equation 2.

$$C_x = E[(f(x,y) - \overline{X})(f(x,y) - \overline{X})^T] \qquad (2)$$

Where: is the windows estimation and $\overline{X}$ is the mean value of $f(x,y)$.

New maximal variance and decorreleted values are provided with Karhunen-Loeve Transform by (3)

$$C_Y = A \cdot C_X \cdot A^T = \sum \qquad (3)$$

Where: $C_Y = \sum$ is the diagonal eigenvalues matrix and A is the eigenvector matrix.

- **(I) FEATURES EXTRACCIÓN (INVARIANT MOMENTS)**

Moments were calculated for the random variable *X*, which was identified with the image block. In addition, *X* is a matrix of two coordinates *(x,y)*, obtained from the image matrix *f(x,y)*. The definition of *(p+q) order invariant moment* around the origin is given by:

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)dxdy \qquad (4)$$

For an image, equation (4) could be expressed as (5), as *(p+q)* is the order of the central invariants moments:

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\overline{x})^p (y-\overline{y})^q f(x,y)dxdy \qquad (5)$$

where *(x, y)* are the pixel coordinates, and $\overline{x}$, $\overline{y}$ are the average values. The third order *central moments* are eight estimated and no redundant *(p+q)* combinations $(\mu_{00}, \mu_{20}, \mu_{02}, \mu_{11}, \mu_{30}, \mu_{12}, \mu_{21}, \mu_{03})$. The normalized central moments, η_{pq} are given by:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \quad \text{where} \quad \gamma = \frac{p+q}{2} + 1 \qquad (6)$$

for $p + q = 2, 3, \ldots$

From the second and third moments we get a set of seven invariant moments [2]:

$$\phi_1 = \eta_{20} + \eta_{02}$$
$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$
$$\phi_3 = (\eta_{30} + 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$
$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$
$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$
$$\phi_6 = (\eta_{20} - \eta_{02})\left[(\eta_{30} + 3\eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$
$$\phi_7 = (\eta_{21} - \eta_{30})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] \qquad (7)$$

According to [2] this set of moments are invariant for translation, rotation or scaling, thus the computing of (7) for an image under transformations ensures that its moments will no change significantly. We will see later that these changes result from both the nature of the digital images and their images resulting after the transformation process.

- **(H) IMAGES COMPARISON**

Comparison is done by equation (8), where similarity function is taken between query image f(i,j) and database images g_k(i,j).

$$likeness(k) = \frac{\frac{1}{MxN}\sum_{i=1}^{N}\sum_{j=1}^{M} f(i,j) * g_k(i,j)}{\sqrt{f(i,j)f(i,j)} * \sqrt{g_k(i,j)g_k(i,j)}} * 100 \qquad (8)$$

So, one-dimension vector is the function ***likeness(k)***, for k-images from data base. ***likeness(k)*** vector is sorted in descendent form, finally it can be taken n-values near to 100 percent. N-value can be the nearest, the first 5 images, 10 images, 20 images, etc.

Noise Sensitivity

We also evaluated the impact of noise on the images. Random noise could arise from transmission errors. Noise power is given by:

$$P_N = \frac{\text{Image Power}}{10^{(\text{SNR}/10)}} \qquad (9)$$

where P_N is the power of noise and SNR is the signal-noise-rate. Once the signal to noise relationship was defined (in order to test over different SNR on image simulating level noise coming from the transmission process), the image power was calculated with:

$$\text{Image Power} = \frac{1}{N \times M} \sum_{x=1}^{N} \sum_{y=1}^{M} f\left(x, y\right)^2 \qquad (10)$$

Where N and M are the width and height dimensions of the image, respectively

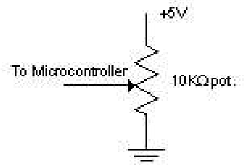
4 Experimental Results

The user through the program client selects an image and he sends it to the server. The server as well contains several images in his data base of images, which are compared with the received image.

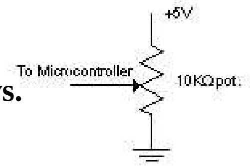
Figure 3 shows some examples of images to be identifier; our data base is conformed by 500 images from different sources: 3(a) TV, 3(b) Cell phone, 3(c) Tree, 3(d) Table, 3(e) Car, 3(f) Motorcycle, 3(g) Airplane 1, 3(h) Watch, 3(i) Airplane , 3(j) Tree, 3(k) flower and 3(l) Grass.

Fig. 3 Some examples of images to be identifier: (a) TV, (b) Cell phone, (c) Tree, (d) Table, (e) Car, (f) Motorcycle, (g) Airplane 1, (h) Watch, (i) Airplane , (j) Tree, (k) flower and (l) Grass

In order to give some examples of how works the proposed methodology and what results it gives. Different test were developed over simple images and complicated ones. As a first example, we test with a simple electrical circuit; result is shown in figure 4.



Compared vs.



100% seemed
(equal)

Fig. 4 Searching a microcontroller circuit and the image result, in percentage

An other example with a tree image:

(a) Original image to be searcher

(b) 100%

(c) 95.15%

(d) 92.81

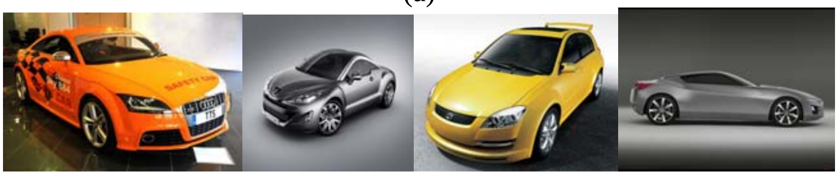
(e) 87.25 %

(f) 82.21 %

Fig. 5 Searching a tree and results given by our method: (a) original tree image, (b) original tree (the same), (c) tree 1, (d) tree 2, (e) tree 3 (f) tree 4.

If now we are interested in cars searching, we present the image that we are looking for in the data base, as a result, our method send us different cars with their seemed percentages, see figure 6.

(a)



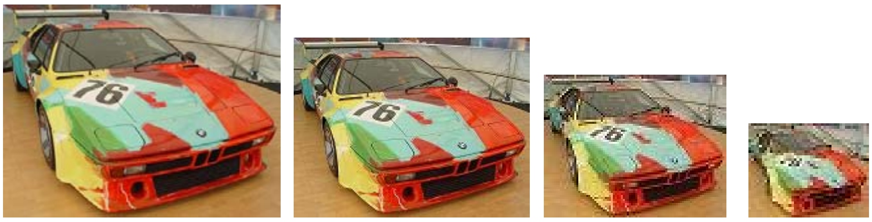
(b) Seemed 100% | (c) Seemed 96% | (d) Seemed 94% | (e) Seemed 88%

Fig. 6 Searching a car and results given by proposed method: (a) original car, (b) car 1(the same), (c) car 2, (d) car 3 and (e) car 4.

A threshold is fixed to the server in order to give as answers, only the images that have the higher seemed. In our test, the threshold was fixed to 80 percent. If the percentage is less than 80%, results are different images even those from other source (cell phones instead of cars).

As a sample of different image size recognition, figure 7 shows the image car sampled to different resolutions. 7(a) Original image at 1200x925 pixels, 7(b) 360x278 pixels, 7(c) 180x139 pixels, 7(d) 90x70 pixels and 7(e) 45x35 pixels. As we can see the percentage identification decreases from 100% until 88.68%, nevertheless, the recognition percentage still high (88.68%) for the smaller image (45x35 pixels).

(a)



(b) (c) (d) (e)

Fig. 7, Search results for car image: (a) Original image Car BMW (size 1200x925), (b)360x278 (98.41%), (c)180 x 139 (96.57%), (d)90x70 92.90% and (e) 45x35 88.68 %.

What it is observed with these results is that between greater it is the image size, less information of loses, and greater percentage of similarity in the comparison is obtained.

5 Conclusions and perspectives

When developing the program prototype for the search of images in Internet based on the analysis in main components and the calculation of its moments, we have reached satisfactory results on the basis of the propose methodology.

The images found with the calculation of the PCA fulfill the results waited for as far as invariant to the rotation, in addition to it thanks to the used method of comparison can be obtained images that have similar forms and they are rotated to any angle. The image or images found by the calculation of Hu moments, is exactly identical or too similar. It sets the standard to make searches very precise, that is to say, where nonsingle it is wanted images with seemed forms, but also with similar statistical properties.

Al to add type noise you leave and pepper to the image, we found that al to add up to 0,06 % of noise to the image, the thrown results are good by part of the PCA, the calculation of Moments is more sensible al noise and up to 0,03 % of noise it is possible to be found to the original image. If we sent to search a same image in different sizes, the PCA algorithm if it will be able to even give similar images making a downsizing of 10: 1.

If extensions to the image are made him look for, the algorithm will recognize without problems images, i.e. an extension of 1: 10 will recognize it the original one until in 96 %. the required computational effort in the calculation of the PCA and Moments he is considerable, an improvement to it would be to also have in the data base the moments already calculated for each images.

The development of this project can continue with future works in which any type of image can be introduced, or with the development of a system that allows to find some element of interest contained in an image. Then, this project does not finish here, and what it hopes it is that it can be retaken in those who is interested in the digital processing of images.

References:

[1] K. Jain, *Fundamentals of digital image processing*, Prentice Hall, Englewood Cliffs, New .Jersey, 1989.

[2] R. C. González and R. E. Woods, *Digital Image Processing*, Addison Wesley, 1999.

[3] D. Maravall G. Allende, *Artificial Vision and Pattern Recognition*, Addison Wesley Ibero american, 1993.

[4] Brian W. Kernighan y Dennis M. Ritchie, *The C programming language*, Prentice may, 1991.

[5] W. Richard Stevens, *Advanced Programming in the Unix Environment*, Addison Wesley, 2000.

[6] W. Richard Stevens, *Unix Network Programming*, Prentice Hall, 2000.

[7] Brian W. Kernighan, *El entorno de programación Uníx*. Prentice may

[8] David J. Kruglinski, *Programación Avanzada con Microsoft Visual C++*, Microsoft Press, 1998.

[9] Balena, *Advanced programming in Microsoft Visual Basic*, Microsoft Press, 1998.

[10] Andrews H.C. y B.R. Hunt. *Digital Image Restoration*, Prentice Hall, Inc., 1977

[11] Kenneth R. Castleman, Digital Image Processing, Prentice Hall.

[12] Domingo, A. *Tratamiento Digital de Imágenes. Madrid*. Anaya, 1993.

[13] Pennebaker, W.B.- Mitchell, J.L. JPEG, *Still Image Data, Compression* Standard. New York. Van Nostrand Reinhold, 1993

[14] Juan Humberto Sossa Azuela, *Un Algoritmo para el Etiquetado y el Cálculo de Características Geométricas de Objetos a través de su Contorno*, Centro de Investigación en Computación del I.P.N., 1999.

[15] Luis Hernández Lara, *Programación en Lenguaje C*, Centro de Investigación en Computación del I.P.N., 1999.

[16] J.Z. Wang, J. Li, G. Wiederhold, SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries, *Pattern Analysis and Machine Intelligent*, September 2001,Vol. 23, No. 9.

[17] Cordelia Schmid, Roger Mohr, Local Grayvalue Invariants for Image Retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence* pp. 530-535, May 1997.

[18] F. Long , H.J. Zhang and D.D. Feng Fundamentals of Content Image retrieval, *In Multimedia Information Retrieval and Management*, D Feng Eds, Springer 2003

[19] Alberto del Bimbo, " A Perspective View on Visual Information Retrieval Systems", Content Based Access of Image and Video Libraries, *Workshop on volume 21*, 108-109, IEEE 1998

[20] Ritendra Datta, Dhiraj Joshi, Jia Li and James Z. Wang.Image Retrieval: Ideas, Influences, and *Trends of the New Age ACM Computing Surveys*, Vol. 40, No. 2, Article 5, Publication date: April 2008.

[21] J. Vogel and B. Schiele. Semantic Modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision*, 72(2):133–157, 2007.

[22] Vogel, J., and Schiele, B. "Semantic Modelling of Natural Scenes for Content-Based Image Retrieval". *Int J of CV*, Springer , 10.1007/s 11263-006-8614-1, 2006.

[23] Vogel, J., A. Schwaninger, C. Wallraven and H. H. Bülthoff: Categorization of natural scenes: local vs. global information. *Proceedings of the Symposium on Applied Perception in Graphics and Visualization* (APGV06), 33-40. (Eds.) Spencer, S. N. ACM Press, New York, NY, USA (07 2006)

[24] Alain Cesar González Garcia. "*Image retrieval based on the contents*" Doctoral Thesis at the Center for Research in Computing (CIC), Instituto Politecnico Nacional, Mexico DF September 2007.

[25] Ying Liu, Dengsheng Zhang, et al. A survey of Content-Based Image Retrieval with high-level semantics, *Pattern Recognition* 40 (2007) 262-282

[26] Jia Li, *Senior Member*, *IEEE*, and James Z. Wang, *Senior Member*, *IEEE* Real-Time Computerized Annotation of Pictures, *Proceedings of the 14th annual ACM international conference on Multimedia* Pages: 911 - 920 Year of Publication: 2006 ISBN:1-59593-447-2

[27] P. S. Hiremath , Jagadeesh Pujari Content Based Image Retrieval using Color, Texture and Shape features. *15th International Conference on Advanced Computing and Communications* Pages 780-784. Year of Publication: 2007

[28] Dinesh Kumar V. P., Tessamma Thomas. Clustering of Invariance Improved Legendre Moment Descriptor for Content *IEEE-International Conference on Signal processing, Communications and Networking Madras Institute of Technology*, Anna University Chennai India, Jan 4-6, 2008. pages 323-327