International Journal of Emerging Engineering Research and Technology
Volume 3, Issue 1, January 2015, PP 98-104
ISSN 2349-4395 (Print) & ISSN 2349-4409 (Online)

Feature Extraction Technique for Robust and Fast Visual Tracking: A Typical Review

Kavya R
M. Tech Student, Department of ECE
HKBK College of Engineering
Bangalore, India

Harisha
Assoc. Professor, Department of ECE
HKBK College of Engineering
Bangalore, India

Abstract: *Visual tracking is a fundamental task in many computer vision applications and it remains a huge challenge, although numerous approaches have been proposed. Feature extraction is one of the major research fields in video processing. Feature Extraction is a method of capturing visual content of images for indexing and retrieval. It is a crucial step for multimedia processing. In this paper, we focus our review on different feature extraction technique. In particular, we first analyze the state-of-the-art feature descriptors which are used to represent the appearance of tracked objects. The image descriptors include texture, color, and shape of the object inside an image. Several feature-extraction techniques viz., Average RGB, Color Moments, Co-occurrence, Local Color Histogram, Global Color Histogram and Geometric Moment have been typically reviewed in this paper. The effectiveness of the fusion of global and local features in automatic image annotation and content based image retrieval system.*

Keywords: *Visual tracking, feature extraction, image retrieval system.*

1. Introduction

Visual tracking is an important research field in computer vision applications, such as security surveillance, access control, vehicle navigation, and human- computer interaction. Visual tracking, in general, is a very challenging problem due to the loss of information caused by the projection of the 3D world on a 2D image, noise in images, cluttered-background, complex object motion, partial or full occlusions, illumination changes as well as real-time processing requirements, etc. In the early years, almost all visual tracking methods assumed that the object motion was smooth and no abrupt appearance change. However, tremendous progress has been made in recent years. Some algorithms can deal with the problems of abrupt appearance change, leaving out from scenes and drifting, etc. To build a robust tracking system, some requirements should be considered such as robustness, adaptive and real-time processing.

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples.

2. Feature Extraction

Feature extraction is a general term for methods of constructing combinations of the variables to get around the problems while still describing the data with sufficient accuracy. Feature extraction involves the image features to a distinguishable extent. Average RGB, Color Moments, Co-occurrence, Local Color Histogram, Global Color Histogram and Geometric Moments are used to extract features from the test image.

Average RGB: The objective of use this feature is to filter out images with larger distance at first stage when multiple feature queries are involved. Another reason of choosing this feature is the fact that it uses a small number of data to represent the feature vector and it also uses less computation as compared to others.

Color Moments: Color Moments are measures that can be differentiate images based on their feature of color, however, the basic concept behind color moments lays in the assumption that the distribution of color in an image can be interpreted as a probability distribution. The three color moments can be defined as:

Mean: Mean can be understood as the average color value in the image

$$E_i = \sum_{N}^{j=1} \frac{1}{N} P_{ij}.$$

Standard Deviation: The standard deviation is the square root of the variance of the distribution

$$\sigma_i = \sqrt{\left(\frac{1}{N}\sum_{N}^{j=1}(P_{ij} - E_i)^2\right)}$$

Skewness: Skewness can be understood as a measure of the degree of asymmetry in the distribution.

$$S_i = \sqrt[3]{\left(\frac{1}{N}\sum_{N}^{j=1}(P_{ij} - E_i)^3\right)}$$

Many other image properties have been introduced for better performance of the system, such as Auto Correlation, Contrast, Energy, Entropy, Homogeneity, Sum Variance, Sum Average, Difference Entropy, Maximum Probability, Dissimilarity, Cluster Prominence etc

Gray Level Co-occurrence Matrices (GLCM): GLCM is a popular representation for the texture in images. They contain a count of the number of times a given feature (e.g., a given gray level) occurs in a particular spatial relation to another given feature.

Color Histogram: Color is the most widely used “feature” owing to its intuitiveness compared with other features and most importantly, it is easy to extract from the image. The color histogram depicts color distribution using a set of bins.

Geometric Moments: In image processing, simple properties of the image which are found via image moments include area (or total intensity), its centroid, and information about its orientation. The pros of using this feature combine with other features such co-occurrence, which can provide a better result to user.

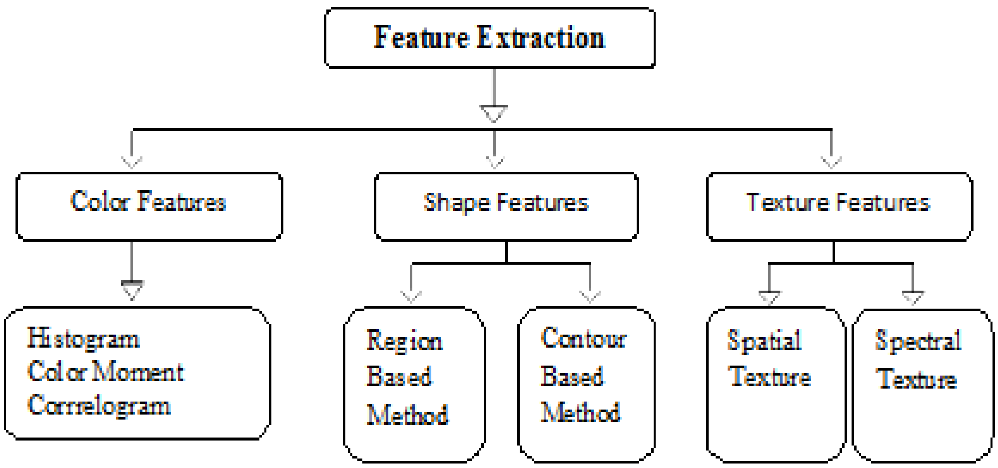

Fig1. *Block Diagram of Feature Extraction*

The first step for a computer program in semantic understanding, however, is to extract efficient and effective visual features and build models from them rather than human background knowledge. So we can see that how to extract image low-level visual features and what kind of features will be extracted play a crucial role in various tasks of image processing. As we known, the most common visual features include color, texture and shape, etc and most image annotation and retrieval systems have been constructed based on these features. However, their performance is heavily dependent on the use of image features. In general, there are three feature representation methods, which are global, block-based, and region-based features. Chow et al., [1] present an image classification approach through a tree-structured feature set, in which the root node denotes the whole image features while the child nodes represent the local region-based features. Tsai and Lin [2] compare various combinations of image feature representation involving the global, local block-based and region-based features for image database categorization. In addition, a block-based image feature representation is proposed by Lu [3] in order to reflect the spatial features for a specific concept. However, little attention has been paid to image feature extraction compared to a significant amount of research on annotation /retrieval model itself construction. Table 1 summarizes the global and local features employed in these references as below.

Table1. *Contrast of global and local feature extraction*

Sources	Global features adopted	Local features adopted
Chow et al.[1]	Color histogram in HSV space	Color moments, Gabor texture, shape and size
Tsai et al.[2]	Color moment in HSV space, four levels of Daubechies-4 wavelet decomposition	Color moment in HSV space, four levels of Daubechies-4 wavelet decomposition
Zhu et al.[4]	Grid color moment, LBP, Gabor wavelets texture and edge orientation histogram	SURF
Tian et al.[5]	Color histogram in HSV space	Color moments in HSV space, Gabor wavelets texture and sobel shape
Lisin et al.[6]	LBP and shape index	SIFT
Zhao et al.[7]	Pseudo Zemike moments	SIFT

2.1. Color features

Color is one of the most important features of images. Color features are defined subject to a particular color space or model. A number of color spaces have been used in literature, such as RGB, LUV, HSV and HMMD. Once the color space is specified, color feature can be extracted from images or regions. A number of important color features have been proposed in the literatures, including color histogram, color moments (CM), color coherence vector (CCV) and color Correlogram, *etc.* To increase the discriminative power, color descriptors have been proposed, which are robust against certain photometric changes. The apparent color of an object is influenced primarily by two physical factors, (1) the spectral power distribution of the illuminant and (2) the surface reflectance properties of the object. Recent advances in color descriptors can be categorized into novel histogram-based color descriptors and SIFT-based color descriptors.

In the HSV color space, it is known that the hue becomes unstable near the grey axis. Van de Weijer et al. [8] applied an error propagation analysis to the hue transformation. The analysis shows that the certainty of the hue is inversely proportional to the saturation. Therefore, the hue histogram is made more robust by weighing each sample of the hue by its saturation. The H color model is therefore scale-invariant and shift-invariant with respect to light intensity. The SIFT descriptor is not invariant to light color changes, because the intensity channel is a combination of the R, G and B channels. Van de Weijer et al. [8] introduced a concatenation of the hue histogram with the SIFT descriptor, which is scale- invariant and shift-invariant. In [10], color invariants had been first used as an input to the SIFT descriptor, which leads to a CSIFT descriptor that is scale-invariant with respect to light intensity.

Table 2 provides a summary of different color methods excerpted from the literature [11], including their strengths and weaknesses. DCD, CSD and SCD denote the dominant color descriptor, color structure descriptor and scalable color descriptor respectively. More detailed information of color descriptors can be found in reference [11].

Table2. *Contrast of different color descriptors*

Color method	Advantages	Disadvantages
Histogram	Simple to compute, intuitive	High dimension, no spatial info, sensitive to noise
CM	Compact, robust	enough to describe all colors, no spatial info
CCV	Spatial info	High dimension, high computational cost
Correlogram	Spatial info	Very high computational cost, sensitive to noise, rotation and scale
DCD	Compact, robust, perceptual meaning	Need post-processing for spatial info
CSD	Spatial info	Sensitive to noise, rotation and scale
SCD	Compact on need, scalability	No spatial info, less accurate if compact

2.2. Shape Features

Shape is very important visual feature to identify and recognize the real-world objects, whose purpose is to encode simple geometrical forms such as straight lines in different directions. Shape feature extraction techniques can be broadly classified into two groups, viz., contour based and region based methods. Region-based methods use the whole area of an object for shape description, while contour-based methods use only the information present in the contour of an object. Local space-time features have recently become a popular representation for action recognition and visual detection. Local space-time features capture characteristic salient and motion patterns in video and provide relatively independent representation of events with respect to their spatio-temporal shifts and scales as well as background clutter and multiple motions in the scene.

Several methods for feature localization and description have been proposed in the literature and promising results were demonstrated for action classification and various detection tasks [12]. Ke et al. [13] studied the use of volumetric features for event detection in video sequences. They generalized the notion of 2D box features to 3D spatio-temporal volumetric features, which is an extension of the Haar-like features [14]. Liu et al. [15] proposed a contour-motion feature descriptor for robust pedestrian detection. The space-time contours are used as the low level representation of the pedestrian. A 3D distance transform is then applied to extend the one-dimensional contour into three-dimensional space.

2.3. Texture Features

Texture is another important property of images. It is generally believed that human visual systems use texture for recognition and interpretation. Texture is a measure of the intensity variation of a surface which quantifies properties such as smoothness and regularity. Texture, on its own does not have the capability of finding similar images, but it can be used to classify textured images from non-textured ones and then be combined with another visual attribute like color to make the retrieval more effective. Basically, texture representation methods can be classified into two categories: structural; and statistical. Statistical methods, including Fourier power spectra, co-occurrence matrices, shift-invariant principal component analysis (SPCA), Tamura features, Wold decomposition, Markov random field, fractal model, and multi-resolution filtering techniques such as Gabor and wavelet transform, characterize texture by the statistical distribution of the image intensity. A large number of techniques have been proposed to extract texture features. Based on the domain from which the texture feature is extracted, they can be broadly classified into spatial texture feature extraction methods and spectral texture feature extraction methods. For the former approach, texture features are extracted by computing the pixel statistics or finding the local pixel structures in original image domain, whereas the latter transforms an image into frequency domain and then calculates feature from the transformed image. Both spatial and spectral features have advantage and disadvantages. Table 3 summarizes their advantages and disadvantages.

Table3. *Contrast of texture features*

Texture method	Advantages	Disadvantages
Spatial texture	Meaningful, easy to understand, can be extracted from any shape without losing info.	Sensitive to noise and distortions
Spectral texture	Robust, need less computation	No semantic meaning, need square image regions with sufficient size

Gabor wavelet [16] is probably the most studied texture feature. The Gabor filters can be considered as orientation and scale tunable edge and line detectors, and the statistics of these micro-features in a given region are often used to characterize the underlying texture information. To be specific, Gabor filter is designed to sample the entire frequency domain of an image by characterizing the center frequency and orientation parameters. The image is filtered with a bank of Gabor filters or Gabor wavelets of different preferred spatial frequencies and orientations. Each wavelet captures energy at a specific frequency and direction which provide a localized frequency as a feature vector. Thus, texture features can be extracted from this group of energy distributions [17]. Given an input image $I(x, y)$, Gabor wavelet transform convolves $I(x, y)$ with a set of Gabor filters of different spatial frequencies and orientations. A two-dimensional Gabor function $g(x, y)$ can be defined as follows

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} exp\left[\frac{-1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\rho_y^2}\right) + 2\pi j w_x\right]$$

Where σ_x and σ_y are the scaling parameters of the filter (the standard deviations of the Gaussian envelopes), W is the center frequency, and θ determines the orientation of the filter.

3. Image Retrieval System

An image retrieval system can be defined as searching, browsing, and retrieving images from massive databases consisting of digital images. Although Conventional and common techniques of retrieving images make use of adding metadata namely captioning keywords so as to perform annotation of words. However image search can be described by dedicated technique of search which is mostly used to find images. There are three techniques for image retrieval: text-based method, content-based method and hybrid method. Image retrieval system can be classified as:

- Text based Image retrieval system
- Content Based Image retrieval system

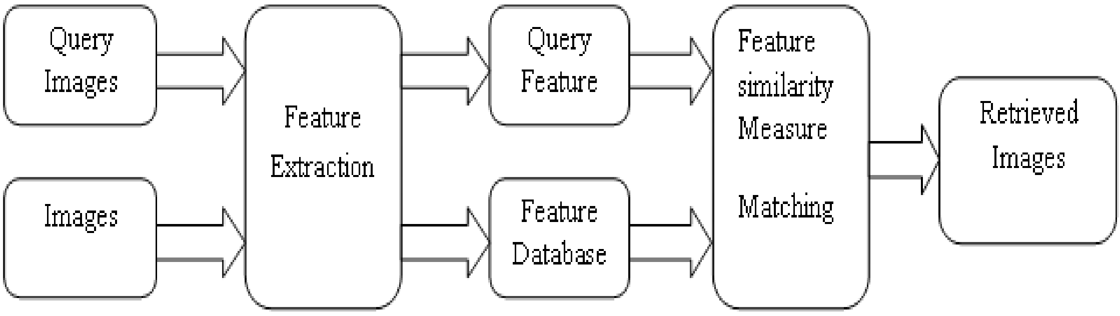

Fig2. *Diagram of the Image Retrieval System*

3.1. Text Based Image Retrieval (TBIR*)*

Text Based Image Retrieval is presently used in almost all general-purpose web image retrieval systems today. This approach uses the text connected with an image to determine what the image contains. This text can be text surrounding the image, the image's filename, a hyperlink important to the image, an explanation to the image, or any other part of text that can be associated with the image. Google, Yahoo Image investigate engines are instance of systems using this approach.

A textual-based image retrieval system always suffers from two problems: high-priced manual annotation and inaccurate and inconsistent automated annotation. On one hand, the cost associated with manual annotation is prohibitive with regards to a large-scale data set. On the other hand, inappropriate automated annotation yields distorted results for semantic image retrieval. As a result, a number of powerful image retrieval algorithms have been proposed to deal with such problems over the past few years. CBIR is the mainstay of current image retrieval systems.

3.2. Content Based Image Retrieval (CBIR)

Content Based Image Retrieval is a set of techniques for retrieving semantically-relevant Images from an image database based on automatically-derived image features. This aims at avoiding the use of textual descriptions and instead retrieves images based on their visual similarity to a user-supplied query image or user-specified image features. The main goal of CBIR is efficiency during image indexing and retrieval, thereby reducing the need for human intervention in the indexing process. The computer must be able to retrieve images from a database without any human assumption on specific domain (such as texture vs. non texture). CBIR combines high-tech elements such as: multimedia, signal and image processing, pattern recognition, human-computer interaction, and human perception information sciences.

The algorithms used in these systems are commonly divided into three tasks:

- Extraction,
- Selection, and
- Classification.

The extraction task transforms rich content of images into various content features. Feature extraction is the process of generating features to be used in the selection and classification tasks. Feature selection reduces the number of features provided to the classification task. Those features which are likely to assist in discrimination are selected and used in the classification task. CBIR can be divided into two stages:

- *Preprocessing:* The image is first processed in order to extract the features, which describe its contents. The processing involves filtering, normalization, segmentation, and object identification. The output of this stage is a set of significant regions and objects.
- *Feature extraction:* Features such as shape, texture, color, etc. are used to describe the content of the image. Image features can be classified into primitives.

In content-based image retrieval, images are automatically indexed by generating a feature vector (stored as an index in feature databases) describing the content of the image. The similarity of the feature vectors of the query and database images is measured to retrieve the image.

Let $\{F(x, y); x = 1, 2, \ldots, X, y = 1, 2, \ldots, Y\}$ be a two-dimensional image pixel array. For color images $F(x, y)$ denotes the color value at pixel (x, y) i.e., $F(x, y) = \{FR(x, y), FG(x, y), FB(x, y)\}$. For black and white images, $F(x, y)$ denotes the grayscale intensity value of pixel (x, y). The problem of retrieval is following: For a query image Q , we find image T from the image database, such that distance between corresponding feature vectors is less than specified threshold, i.e.,

$$D\ (\text{Feature}\ (Q), \text{Feature}\ (T) \leq t$$

The content-based approach can be summarized as follows:

- Computer vision and image processing techniques in are used to extract content features from the image.
- Images are represented as collections of their prominent features. For a given feature, an appropriate representation of the feature and a notion of similarity are determined.
- Image retrieval is performed based on computing similarity or Dissimilarity in the feature space, and results are ranked based on the similarity measure.

4. Conclusions

This paper provides a typical review of different feature extraction technique and image retrieval system. Feature extraction plays a crucial role in multimedia processing. In this paper, we survey the recent progress on feature extraction techniques where selecting the right feature plays a critical role in visual tracking. We have presented recent advances in feature descriptors motivated from recent innovation in visual object detection area which uses single type of feature. To improve the detecting performance, multiple features can be integrated. We have also presented review on image retrieval techniques. The retrieval performance can be further improved by using a ‘Text and Image’ query system as compared to a text-only or image-only query system, which can take advantage of the keyword annotations. The results from the keyword annotations and image retrieval can then be matched using the feature extraction techniques to present an optimized set of results to the user. A number of systems using image content feature extraction technologies proved reliable enough for professional applications in industrial automation, biomedicine, social security, biometric authentication and crime prevention.

Acknowledgement

I am very grateful and would like to thank my guide Assoc. Prof. Harisha for his advice and continued support. I would like to thank my friend for the thoughtful and mind stimulating discussion we had, which prompted us to think beyond the obvious.

References

[1] T. W. S. Chow and M. K. M. Rahman, “A new image classification technique using tree-structured regional features”, Neurocomputing, vol. 70, no. 4-6, (2007), pp. 1040-1050.

[2] C. F. Tsai and W. C. Lin, "A comparative study of global and local feature representations in image database categorization", In Proc. 5th International Joint Conference on INC, IMS & IDC, (2009), pp. 1563-1566.

[3] H. Lu, Y. B. Zheng, X. Xue, *et al.*, "Content and context-based multi-label image annotation", In Proc. Workshop of CVPR, (2009), pp. 61-68.

[4] J. Zhu, S. Hoi, M. Lyu, *et al.*, "Near-duplicate key frame retrieval by nonrigid image matching", In Proc. ACM MM, (2008), pp. 41-50.

[5] D. Tian, X. F. Zhao and Z. Shi, "Support vector machine with mixture of kernels for image classification", In Proc. ICIIP, (2012), pp. 67-75.

[6] D. A. Lisin, M. A. Mattar, M. B. Blaschko, *et al.*, "Combining local and global image features for object class recognition", In Proc. Workshop on CVPR, (2005).

[7] J. Zhao, Y. Fan and W. Fan, "Fusion of global and local feature using KCCA for automatic target recognition", In Proc. ICIG, (2009), pp. 958-962.

[8] J. Van de Weijer, T. Gevers, A.D. Bagdanov, Boosting color saliency in image feature detection, IEEE Trans. Pattern Anal. Mach. Intell. 28 (1) (2006) 150–156.

[9] T. Gevers, J. van de Weijer, H. Stokman, Color Image Processing: Methods and Applications: Color Feature Detection: An Overview, CRC Press, 2006.

[10] A. Abdel-Hakim, A. Farag, CSIFT: A SIFT descriptor with color invariant characteristics, in: CVPR, 2006.

[11] D. S. Zhang, Md. M. Islam and G. J. Lu, "A review on automatic image annotation techniques", Pattern Recognition, vol. 45, no. 1, (2012), pp. 346-362.

[12] H. Wang, M. Ullah, A. Klaser, I. Laptev, C. Schmid, Evaluation of local spatio- temporal features for action recognition, in: BMVC, 2009.

[13] Y. Ke, R. Sukthankar, M. Hebert, Efficient visual event detection using volumetric features, in: ICCV, 2005.

[14] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: CVPR, 2001.

[15] Yazhou Liu, Xilin Chen, Contour-motion feature (CMF): a space–time approach for robust pedestrian detection, Pattern Recognition Lett. 30 (2) (2009) 148–156.

[16] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of large image data", IEEE PAMI, vol. 18, no. 8, (1996), pp. 837-842.

[17] S. E. Grigorescu, N. Petkov and P. Kruizinga, "Comparison of texture features based on Gabor filters", IEEE TIP, vol. 11, no. 10, (2002), pp. 1160-1167.

[18] Ryszard S. Chora´s , "Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems", 2007.

[19] T. K. Shih, J. Y. Huang, C. S. Wang, *et al.*, "An intelligent content-based image retrieval system based on color, shape and spatial relations", In Proc. National Science Council, R. O.C., Part A: Physical Science and Engineering, vol. 25, no. 4, (2001), pp. 232-243.