Proposed Image Similarity Metric with Multi Block Histogram used in Video Tracking

Dr. Alia K. Abdul Hassan
Computer Science Department, University of technology/Baghdad
Dr. Hasanen S. Abdullah
Computer Science Department, University of technology/Baghdad
Akbas E. Ali
Computer Science Department, University of technology/Baghdad
Email:akbas_ali@yahhoo.com

ABSTRACT

One of the important requirements in the object detection and tracking is the extracting of efficient features to track the target in video sequence. The feature of colour in image is one of the most visual features widely used. The using of colour histogram is the most popular method for representing color feature. One of the problems of using colour histogram to represent feature is its lack of spatial information where it is used to represents statistical distribution of the colours only. In this paper a new similarity metric with multi block colour histogram of image is proposed. This metric will be used by an object tracking method where the similarity will be applied to get a decision of choosing the correct solution (location) of the object from many candidate locations
Keywords: object tracking, feature extraction, similarity metric.

INTRODUCTION:

The main goal of computer vision is to make computers enable to replicate the main functions of human vision like: scene understanding and intelligent motion perception. To achieve the duty of motion perception, efforts has been focused on visual target tracking. It is a great computer vision mission which can be applied to many applications like human computer interaction, and visual surveillance.

Feature representation and extraction is an important stage for multimedia processing. The selection of ideal features that can reflect the fundamental content of the image is still a challenging issue in computer vision [2]. An important step in visual target tracking, after the object detection, is to extract effective and efficient visual features and build model for it to be able to track the target. The most common visual features include color, shape, and texture. The most significant advantages of color feature are simple extracting color information of images and high efficiency, power of representing visual content of images, relatively power in separating images from each other, independent of image size and orientation, and relatively robust to background complication [2].

Related Work

A review of the previous work that related to the subject of this research will be given below with a short explanation to each of them:
i- The paper of SZABOLCS SERGYA´ N, that titled "Precision Improvement of Content-based Image Retrieval Using Dominant Color Histogram Descriptor", 2013 [8].In this paper a novel color

feature is proposed, that takes into account the dominant colors in the color histogram, the HSV color space, and the spatial information of pixels.

ii- the paper of TUDOR BARBU, which titled "A Novel Image Similarity Metric using SIFT-based Characteristics", 2011 [10]. In this paper a powerful similarity measurement for images is proposed. First a SIFT image feature extraction is illustrated. Then, a metric which measures the distance between the resulted SIFT feature vectors is modeled. It utilizes the matching of the SIFT key points between the two compared images.

iii- The paper of Heebeom Bang · Sanghoon Lee and Dongjin Yu, that titled "Robust object recognition using a color co-occurrence histogram and the spatial relations of image patches", (2009), [9], a robust object recognition system is proposed where spatial relationships among patches and patch-based pyramid images are used for the image model. Both a color histogram and a color co-occurrence histogram are applied to obtain image features for each patch.

iv- The paper of Dengsheng Zhang, that titled "Improving Image Retrieval Performance by Using Both Color and Texture Features", 2003, [7], In this paper, a method combining both texture features and color of an image is proposed to improve the matching performance.

Object tracking

  Visual target tracking is a considerable computer vision function that may be applied to many applications like security, visual surveillance, video compression, human computer interaction, and traffic monitoring.

The essence of visual target tracking is to estimate the motion state (orientation, location, and size) of an object in every frame of video as time passes [4].

A typical visual target tracking system is consists of four logical modules:

-        Initialization. It can be automatic or manual. Automatic initialization is performed by target detectors. In contrast, manual initialization is achieved by users to mark object location with a bounding boxes.

-        Modeling of the object. This consists of two components: statistical modeling and visual representation. Statistical modeling focuses on how to build mathematical models to identify object using statistical methods. Visual representation concentrates on how to build strong target descriptors using different kinds of visual features.

-        Motion estimation. The function of motion estimation is achieved by using predictors such as Kalman Filters, linear regression techniques, or particle filters.

-        Localization of the object. This is performed by a maximum posterior estimation or greedy search based on motion estimation [4].

Harmony search algorithm is used in this paper to evaluate the proposed similarity measurement. The harmony search algorithm was developed in 2001 as a heuristic optimization algorithm for use within diverse optimization problems. It inspired by the improvisation procedure for music players. The musicians in the orchestra or band are represented by the various areas of the vector of solution. An ideal harmony happens when every musician plays the ideal note. In a similar trend an ideal solution vector is located once the worthiness of every part is optimal. The player improvises new tones and tests them for the harmony with the remaining portion of the band. If the newest improvisation is beneficial the improvised tone is remembered for future exploit, otherwise the player forgets the tone and plays an alternative improvisation.

The harmony search HS algorithm mimics this behavior by preserving a matrix of the greatest vectors of solution called the (HM) Harmony Memory. Generally, HSA is divided into five steps as follows [6]:

Step1: Initiate the Harmony Search Algorithm's parameters.

Step2: Initialize the memory of harmony (HM).
Step3: Improvise a new solution.
Step4: Update the (HM) Harmony Memory.
Step5: Repeat step3 and step4 until reaching the stopping condition.
Ste6: end
Steps (3 & 4) will repeat while the termination condition (the number of improvisations) is not reached.
The Harmony search can be used in object tracking to estimate prediction of target location. The Harmony memory (HM) is filled with improvised solutions by using motion model that depends on fixed velocity of the target in the video sequence. The optimal solution which has best similarity is chosen as a next location of the moving target. The proposed similarity measurement is used to choose the best solution which represents the location of the object in the next frame. This can be achieved by computing the similarity between the object in current frame with all the improvised locations in the next frame any location that contain the more similar object will represent the wanted location and will be labeled by a bounding box.

The proposed method
   The feature extraction is a method used to convert the tracking object image into a feature vector that represents its features in efficient and compact way.
  The choice of good features is the main problem in many multimedia applications and it is the backbone of the work that the success of the research will depends on it. Best features must be unique and invariant against geometric transformation.
The histogram represents statistical distribution of the colours in the object image only and ignores other properties like the spatial location of these colours, Therefore a new method to extract feature are proposed. This method can be achieved by dividing the image into ***nxn*** blocks or local regions as shown in figure 1,and take a local histogram for each region. By this method we combine each histogram with its location (spatial information) and make the histogram more robust and able to detect and track targets efficiently. The algorithm (1) is illustrating the multi block histogram feature vector extraction process.
One of the most important concerns is to make the system more consistencies against any variation in the direction and magnitude of the light to preserve the template model more robust. This can be done by converting the image from RGB color model to HSV color model and taking the *H* hue channels which contains the color information and discarding the *S* saturation and *V* brightness channels. By this approach the model becomes stronger against varying light conditions and illumination.

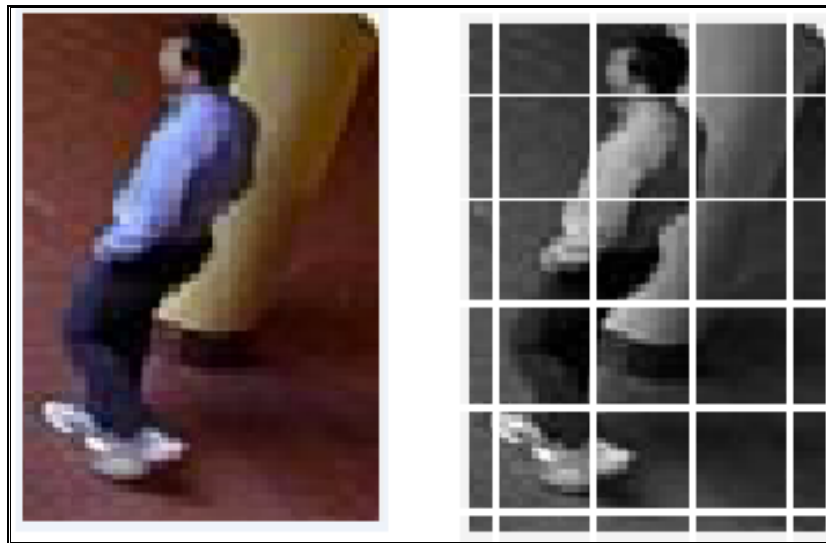| **Algorithm (1):** multi block histogram |
|---|
| **Input**   : image
Output: image-feature |
| Step1: start
Step2: convert the image from RGB color model into HSV
Step2: divide the image into mlocal blocks, each block with nxn pixels
Step3: for i = 1 to m   // blocks number//
block-hist(i) = histogram(the hue channel of block(i))
end   // i loop//
Step4: end |

Figure (1): the division of template image into blocks

To compute the degree of similarity of two images, a comparison between each corresponding blocks of the images are achieved to compute the distance between their local histograms. . If the distance is smaller than a threshold then these blocks will be marked as similar blocks. For all the blocks in these images the summation of similar blocks will be the similarity metric between the two images. The computation of similarity degree is shown in the proposed algorithm (2).

Algorithm (2): proposed of computing similarity of multi block histogram
Input: two images with their block-hist, one of them is the template image and the second is improvised image.   **Output:** similarity-measure

Step1: similarity counter = 0
 Step2: for i = 1 to m   //block number//
 distance = the Euclidean distance(Block-hist1(i),Block-hist2(i))
 If the distance <similarity threshold (like 0.2) then
 similarity_ counter = similarity_ counter + 1
 end   // i loop//
    similarity-measure = similarity_ counter
 Step3: end

The similarity-counter represents the number of similar blocks between the two images. This measurement will be good decision criteria to choose the best solution from the Harmony memory HM.

Object tracking evaluation metrics

The performance of the object tracking model can be evaluated empirically by focusing on detecting and tracking which means to measure the consistent labeling   of the target along time. The issue to address is specifying when the ground truths (the tracking targets) and estimates

(tracking outputs) are overlapping. This acquaintance is pivotal to determining if a target is detected and if it is tracked in correct way.

The first metric used to evaluate tracking is the Recall measure that computes how of the ground truth GTi is covered by tracking estimate Ei. Recall R is expressed as:

$$R(Ei,GTi) = \frac{|Ei \quad GTi|}{|GTi|} \quad \dots (1)$$

The second metric is thePrecision measures which compute how much of the tracking estimate Ei covers the ground truth. Precision P is expressed as:

$$P(Ei,GTi) = \frac{|Ei \quad GTi|}{|Ei|} \quad \dots (2)$$

The Recall and Precision can take values between 1 (fully overlapped) and 0 (with no overlapping).

Experiments

Power of any proposed algorithm arises when it is tested. This section presents the experimental results of implementing the proposed image similarity metric with multi block histogram testing by the Harmony Search tracking system to track object in a video sequences.

The experiment is performed on two videos to track objects. The first video contains a moving rigid ball in 40 frames and the second video contains a moving non-rigid object (person) in 55 frames. Some frames of the labeled output video are list in figure (2) for the first video and in figure (3) for the second video.
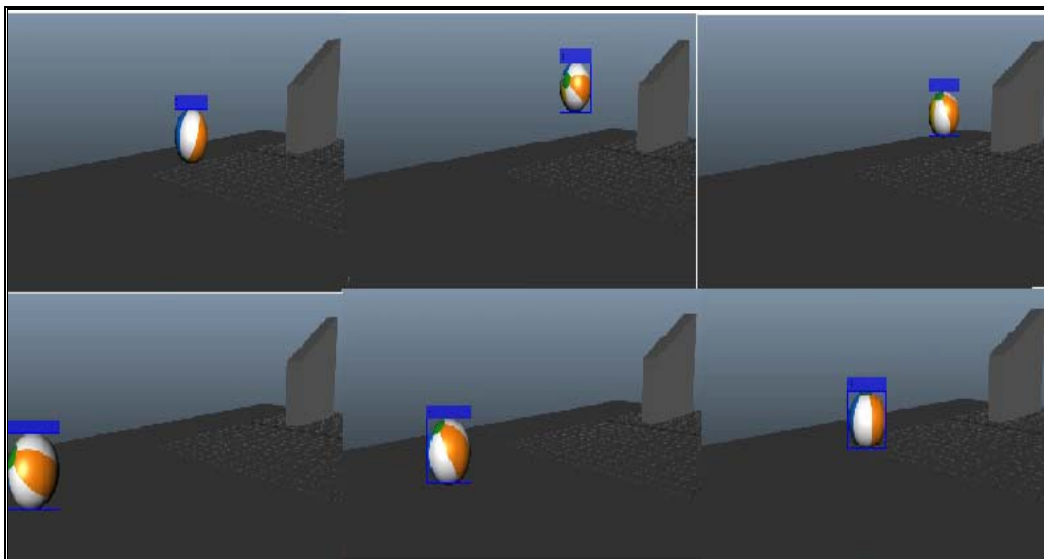

Figure (2) output frames from video-1

The results is evaluated by the tracking metrics and gives a high Recall and Precision which means it detect and track the target in efficient manner. The tracking achieve an average of Recall =

91.5%. The evaluation results are listed in the table-1 for the first video and table-2 for the second video.

Table-1 Recall and Precision of video-1

	Recall	Precession Average	Iteration Average(iterations)
HS filter	0.93	0.65	21.3

Figure (3): output frames from video-2

Table-2: Recall and Precision of video-2

	Recall	Precession Average	Iteration Average(iterations)
HS filter	0.90	0.60	24.3

The observation that can be obtained from the two tables is that the result of second table is less accurate than the result of first one, because it represents the tracking of the object in the video-2 which has a variable appearance due to its natural of the non-rigid, while the object of video-1 has a steady rigid appearance which can be recognized easily and gives more accurate tracking results.

CONCLUSIONS

The choice of good features is the main problem in object tracking applications. It is essential to choose a robust feature that must be unique and invariant against geometric transformation. The multi block histogram is a good feature used with the HSV color model to represent the tracking objects. The similarity of the template image and improvised image are evaluated by the proposed similarity metric which gives a good result by using the Harmony Search tracking method.

REFERENCES

[1] . Pallavi C. Hingane, S. A. Shirsat, "Object Tracking Using Joint Color-Texture Histogram", International Journal of Science and Research (IJSR), 2013

[2]. Dong ping Tian and Baoji, Shaanxi, " A Review on Image Feature Extraction and Representation Techniques", International Journal of Multimedia and Ubiquitous Engineering Vol. 8, No. 4, July, 2013

[3] Jaco Fourie, Steven Mills, and Richard Green," Visual Tracking Using the Harmony Search Algorithm", IEEE, 2008.

[4] Xi Li2;, Weiming Hu, Chunhua Shen1, Zhongfei Zhang, Anthony Dick, Anton van den Hengel1, "A Survey of Appearance Models in Visual Object Tracking", Appearing in ACM Transactions on Intelligent Systems and Technology, 2013.

[5] Fatemeh Alamdar1 and MohammadReza Keyvanpour, "2 A New Color Feature Extraction Method Based on Dynamic Color Distribution Entropy of Neighborhoods", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011.

[6] Osama Abdel-Raouf, and Mohamed Abdel-Baset Metwally, "A Survey of Harmony Search Algorithm", International Journal of Computer Applications (0975 – 8887) Volume 70– No.28, May 2013.

[7] Dengsheng Zhang, "Improving Image Retrieval Performance by Using Both Color and Texture Features", Australia, 2004.

[8] SZABOLCS SERGYA´ N, "Precision Improvement of Content-based Image Retrieval Using Dominant Color Histogram Descriptor", 2013.

[9] Heebeom Bang · Sanghoon Lee and Dongjin Yu, "Robust object recognition using a color co-occurrence histogram and the spatial relations of image patches", Artif Life Robotics (2009).

[10] TUDOR BARBU, "A Novel Image Similarity Metric using SIFT-based Characteristics", Mathematical Models in Engineering and Computer Science, 2011.

[11] Israa A. Alwan, "Object Tracking using Generalized Gradient Vector Flow", Eng. & Tech. Journal, Vol. 29, No. 7, 2011.

[12] Dr.Thair A. Salh & Mustafa Zuhaer Nayef,"Improvement of Face Recognition System Based on Linear Discrimination Analysis and Support Vector Machine", Eng. &Tech. Journal, Vol.31, Part (A), No.12, 2013.