International Journal of Electronics Communication and Computer Engineering
Volume 5, Issue (4) July, Technovision-2014, ISSN 2249–071X

Technovision-2014: 1st International Conference at SITS, Narhe, Pune on April 5-6, 2014

Coherent Theme Discovery in Videos and Images based on Visual Patterns

Ms. Vanita P. Punjabi
M.E. (VLSI And Embedded Systems), Pune University
SND College Of Engineering And Research Center
Yeola, India
Email: vinnipunjabi2111@gmail.com

Prof. Dabhade Ravindra
Dept. of Electronics And Telecommunication Engg.
SND College Of Engineering And Research Center
Yeola, India
Email: rgdabhade@gmail.com

Abstract **– In this speedy development of digital span, the most important thing is to acknowledge the existence frequently appearing object in the given collection of videos or images. This frequently appearing object is called as theme object. There are two significant provocations for discovering theme objects, 1) We don't have existing skills about it, like where and when the object will come into sight, 2) The theme object of heed can be under drastic variations in aspect due to changes in luminance, scale variation, point of view, and color. A novel bottom-up approach is proposed to moderately prune unfamiliar local invariant features. A multilayer candidate pruning procedure is sketched to precede rapidly the image data mining process by assigning the commonness score to visual features. By searching the sub image of the highest commonness score in each image, we can locate and crop the theme object required.**

Keywords **– Image Data Mining, Visual Patterns, Theme Object Discovery.**

I. Introduction

A. Overview of data mining

We are the Elizabethan age an age of computers and fax machines in an age commonly allude to as the intelligence age. In this intelligence age, because we conclude that intelligence conduces to potential and victory, and owing to knowledgeable technologies such as computers, satellites, etc., we have been gathering huge quantity of intelligence. Data mining, also universally known as Knowledge Discovery in Databases (KDD), introduce to the significant removal of implied, previously unknown and potentially useful intelligence from data in databases. The following figure 1.1 shows data mining as a knowledge discovery process. The Knowledge Discovery in Databases process consists of a small steps superior from raw data groups to some configuration of recent knowledge. Once the discovered knowledge is introduced to the user, the judgment count can be elevated, the mining can be additionally concentrated, new data can be further converted. The computationally process is composed of the following evaluation.

Data cleaning is also known as data cleansing, it is a stage in which noise data and unrelated data are discarded from the collection. Data integration at this instant, several data sources, frequently varied, may be combined in a recurrent source. Data selection at this stage, the data applicable to the examination is pronounced on and recovered from the data collection. Data transformation is a phase in which the determined data is transformed into forms suitable for the mining procedure. Data mining is the critical move in which intelligent techniques are applied to extract patterns. Pattern evaluation in this phase, are recognized established on given measures. Knowledge representation is the endmost step in which the discovered knowledge is visually constituted to the user. This required step uses visualization techniques to support users illuminate and understand the data mining outcome.

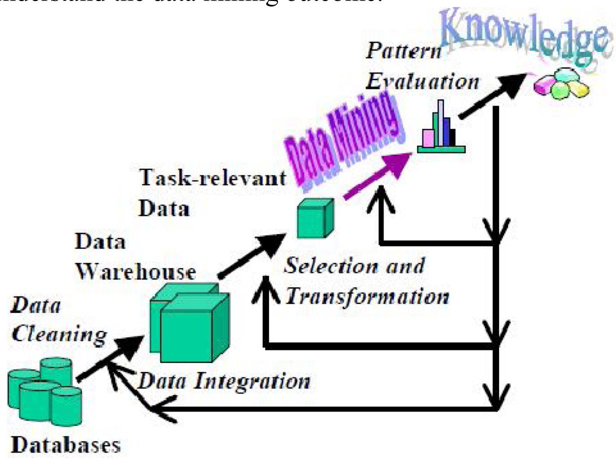

Fig.1.1. Knowledge Discovery in Databases Process

B. Visual Pattern

Visual patterns are formation of visual primitives that appear commonly in video and image datasets. As shown in figure 1.2, the multiple visual pattern in video or image data can be a texton that represents the frequently appearances of image texture, e.g., a face pattern collected of two eyes, a nose, and a mouth; a bedroom including a bed, a lamp, a vase, and so on; or a human being action

Fig.1.2. Multiple visual patterns

International Journal of Electronics Communication and Computer Engineering
Volume 5, Issue (4) July, Technovision-2014, ISSN 2249–071X

that narrate postures and movements of body, e.g., a bent-leg layover spin motion [8]. Before mining visual patterns, it is essential to extract invariant visual primitives from video or image data.

C. Bottom-Up Mining Approach

Bottom-up approach begins from the local layout of visual primitives to detect common visual patterns in video and image data. There is multiple profit of bottom-up approach. First of all, bottom-up approach can be widely registered for their data driven feature. Second, bottom-up techniques can clearly include varieties of contexts such as spatial cooccurrence of multiple visual primitives and correlation between pairs of visual primitives. Third, bottom-up methods are simple to execute.

D. Theme Object Discovery

We are given with a troupe of video sequence or a troupe of images to distinguish commonly appearing objects. Each sequence contains many occurrences of the same theme object but not every frame contains an occurrence [4]. The discovered theme object is localized by the red bounding box. Figure 1.3 illustrates the examples of theme object discovery, in that we need to recognize or discover the frequently appearing objects that are illustrative of the visual contents. This frequently appearing of objects is called as theme or thematic objects [3]. Discovering of theme objects in a video sequence and in a collection of images is a challenging task [3]. There are two significant provocation for discovering thematic objects 1) we don't have an existing information or skills about it, that is where and when the object will come into sight; 2) the total number of thematic objects; and 3) the thematic object of heed can be under drastic variations in aspect to changes in luminance, scale variation, point of view and color. It is not important to handle its changes and exactly discover its occurrences. Moreover, with the help of local invariant features, it can handle object variations such as scale and slight point-of-view changes, color and lighting condition variation [5].

For discovering thematic objects we proposed a novel bottom up approach. Without being affecting the background clutters the more special importance is on the accurate localization of the objects. The aim of a system is to produce the list of thematic objects that are applicable to image query. Initially each image must be extracted to a quantitative description of

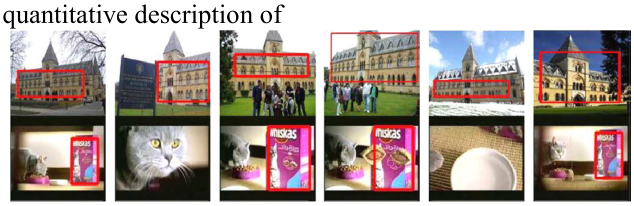

Fig.1.3. Examples of Theme object discovery

To discover the complete theme objects these matched visual primitives are moderately expanded. In the previous phase, uncommon visual features are removed as they do not belong to any similar pattern that is, theme object. Further each individual primitive are granted as its local spatial, nearest neighbors as a big visual group. Multilayer candidate pruning is used to check the commonness score of this spatial pattern. Due to assigning a commonness score to individual primitives, it point out its probability of belonging to a theme object. The commonness score of any sub image is the instance of the scores of its visual features. By searching the sub image of largest commonness score in each image, where it can locate and crop the theme objects using Branch and Bound search Algorithm. Efficient theme discovery only needs matching of visual primitives. It can directly locate and discover respective theme objects without requiring any skills of the total amount of such objects in previous. Local invariant features provides an excellent features, it can manipulate object variations such as lighting, point of view, scale, color, and it is not sensitive to limited occlusion. Feature must be invariant to illumination changes, point of view, scale changes and changes in angles. We match the visual primitives to identify the presence theme object in a video. The match function is applied to the patch descriptor to find there is any match between the given image and the frames of the video. Sector localization of the regions is supposed to be containing the theme object. The branch and bound algorithm can be used for the extraction of the required sector.

E. K-Spatial Nearest Neighbors

Nearest neighbor search, also known as nearness search, parallel search or nearest point search, is an optimization problem for searching closest points. Nearest is typically conveyed in phrase of a distinction function. Knearest neighbor algorithm identifies the upper "k" nearest neighbors to the query. In pattern recognition, the k-Nearest Neighbors algorithm is a non-parametric method used for organization and regression. In k-NN classification, the output result is a class associateship. An object is classified by a larger vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.

When the input data to an algorithm is too vast to be handled and it is mistrusted to be infamous dismissed then the input data will be transformed into a reduced representation set of features vectors. Converting the input data into the set of features vectors is called local feature extraction. If the features extracted are carefully chosen it is imagined that the features set will extract the applicable skills from the input data in order to represent the desired assignment using this diminished representation instead of the full size input. Feature extraction is performed previously on raw data to apply *k-SNN* algorithm on the transformed data in feature spatial.

International Journal of Electronics Communication and Computer Engineering
Volume 5, Issue (4) July, Technovision-2014, ISSN 2249–071X

II. Literature Survey

To discover similar visual primitives in videos and images, some existing task denote an image as a grid composed of visual primitives , such as corners, interest points, and image segments [3], [5].

Hongliang Li and King Nga Ngan [1] introduced a procedure to indentify the existence of co-saliency from an image pair that may have identical objects in common. Here the co-saliency is described as the multi-image saliency map and linear combinations of the single –image saliency map. SISM is described to design local attention. And in MISM, the image pairs are divided into a spatial pyramid to construct a co-multilayer graph. Each node represented in the graph consists of two types of visual descriptors like color and texture properties. A normalized single pair Sim-Rank algorithm is used to calculate the similarity score.

Jun-Bin Yeh, Chung-Hsien Wu, and Sheng-Xiong Chang [2] presents a visual language model characterizes the temporal relation among the frames in a visual stroke. Detecting the object in a visual stroke is difficult and a textual term may generally match to respective word. The sentence based alignment corresponds to the language model-based temporal relation. The visual patterns are extracted from the bag-of-words represented from the main objects in the key frames of a visual stroke. The How Net knowledge base drives the textual terms to the textual concepts. The IBM model-1 is used to process the textual concepts and the visual patterns. David liu and Tsuhan chen [3] presents a situation or model where there are multiple possible outcomes framework for discovering the thematic objects in video. The background can be covered with undity collection of things, the video can vary between different strokes, and the unrevealed objects can enter or leave the position at multiple times. They had used maximally stable extremal regions as observations in model and hence provide robustness to object variations. These frameworks can be applied to a large-scale range of different objects and videos types. They had demonstrated excellent performances to method that drives global image statistics and instant item set data mining techniques.

Shalini N. and Sharada K. [5] proposed a paper where they are discovering thematic object in a video. This frequently appearances of object in a video are needful for object search and summarization of that object. Harris corner detection algorithm is referred to find the corner points of that object to be discovered. Video data mining process is used to recognize the common pattern that appears in that video.

Jianbo Shi and Jitendra Malik [6] proposed a novel approach for solving the grouping problem in vision. The theme of this paper is exactly to extract the global impression of an image. The complete dissimilarity between the different groups as well as the complete similarity within the groups is measured by normalized cut criterion. An efficient computational technique is referred to minimize this criterion. Zhengwei Yang and Fernand S. Cohen [7] proposed affine transformation for object detection. An object is detected by affine never changing to create the correlation between the vertices of a test image with a query image. The algorithm used in this will detect an object that must be constituted as polygonal outliners and also as a group of scattered feature points. It uses point similarity approach for detection. However, if the objects have unique shape then they have uniform curved framework.

III. System Model

The below figure 2, explains the block diagram to identify the theme objects in videos and images [3], [5]. Initially the images in the database are divided into sectors which may represent an object defined by R. Features in the form of visual primitives are extracted for the formation of the feature vector. Further is the Feature matching, where the features extracted from the input image are matched with the stored template or reference model and a recognition decision is made? The similarity among the similarly themed objects should be maximized; hence the uncommon features are to be pruned (removed).

For the purpose of pruning, *k-SSN* is which estimates the similarity co-efficient among the similarly themed objects. To discover theme object in a video, we characterize video as a collection of video frames, each frame as an image sequence. Each image is characterized by collection of local visual primitives.

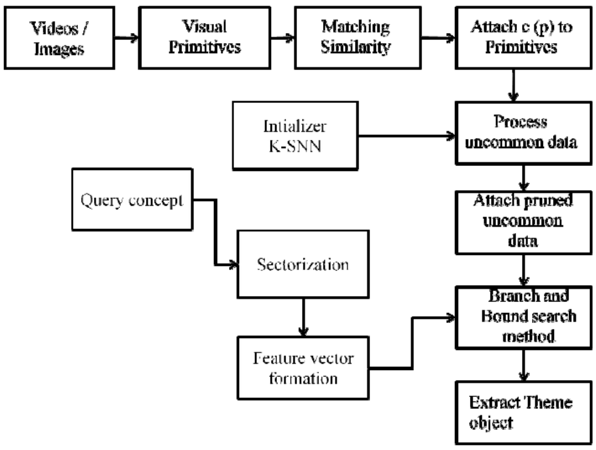

Fig.2. System block diagram

A visual primitive is a property of an image located on a single point or small region. The local features of an object are color or gray value of a pixel. For object recognition, the local feature must be invariant to illumination changes, point of view scale changes and changes in angles. We match the visual primitives to identify the presence theme object in a video. The match function is applied to the patch descriptor to find there is any match between the given image and the frames of the video. Sector

Published by IJECCE (www.ijecce.org)

International Journal of Electronics Communication and Computer Engineering
Volume 5, Issue (4) July, Technovision-2014, ISSN 2249–071X

localization of the regions is supposed to be containing the theme object. The branch and bound algorithm can be used for the extraction o the required sector.

IV. Implementation Detail

4.1. Overview

To discover theme object from given a collection of T images $\mathcal{D} = \{\boldsymbol{I}i\}$, it describes every image $\boldsymbol{I}i = \{p_1 . . . . p_m\}$ by a number of visual primitives. To gradually reject uncommon visual primitives $p \in I$ to recover R^*. In the previous step, we discard uncommon primitive's p that find few matches among the rest of images in $\mathcal{D}$. The remained set of visual primitives is described as $\mathcal{D}^1 \subseteq \mathcal{D}$.

For the further verification $p \in \mathcal{D}^1$ each, it spatially expands and forms a larger pattern. For each $p \in \mathcal{D}$, its spatial neighbors form a visual group $Gp = \left\{p, p_1^{NN} ....p_k^{NN}\right\}$, where p_1^{NN} is one of the nearest neighbors of p in the image. A commonness score $C(p)$ will be assigned to each visual primitive p. Visual primitive p has a positive commonness score if its visual group Gp frequently appears among the data set $\mathcal{D}$, and vice versa, whereas it has a negative commonness score if it repeats rarely. The thematic object can be located as the sub image region $R \subset I$ that contains the most common primitives.

4.2. Multilayer Candidate Pruning

Here we can examine the commonness of each primitive by treating its k-SNN_5. The best performances for primitives p depends on size of k as it increases. The number of SNN is k for groups in $\mathcal{D}^1$, the similarity between two sets G_q and G_p, $Sim\left(G_p, G_q\right)$, can be defined as a matching problem.

$$\mathbf{Sim}\ (\mathbf{G_p}\ ,\ \mathbf{G_q}) \triangleq \max_{f \in F} \sum_{i=1}^{|Gp|} s(pi, f(pi)) \quad (1)$$

Here f denotes a matching between two point sets G_p and G_q, F is the complete set of all possible matching. Given a group G_p, its supportive set consists of the groups in the rest of images that match G_p.

$$\mathbf{S}_p = \{\mathbf{G_q}: \mathbf{Sim}\ (\mathbf{G_p}\ ,\ \mathbf{G_q}) > \theta\} \quad (2)$$

After rejecting uncommon groups, an even smaller candidate set is obtained as $\mathcal{D}^2$. For the total L layers denoted as $\mathcal{D}^L$ the final set, we obtain a filtration,

$\mathcal{D}^L \subseteq . . . \mathcal{D}^l \subseteq \mathcal{D}$ and spatial neighborhood size $k^L >> ..... > k^1 > 0$.

As compare to $\mathcal{D}^l$, a visual primitive $p \in \mathcal{D}^l$ $(2 \le l \le L)$ responds to a larger spatial neighborhood which belongs to be a part of a common theme object. Multilayer checking assigns a commonness score for each p. For particular primitive $p \in \{\mathcal{D}^L . . . . \mathcal{D}^l\}$, its commonness score is assigned with a positive value. Whereas for the primitives in that is, $\mathcal{D} / \mathcal{D}^l = p\{\ p\!: p \in \mathcal{D},\ p \notin \mathcal{D}^l\}$. Its commonness score is designated as s negative value. The commonness score designated to each p as given below. Here τ is assigned as the negative vote value.

$$C(p) = \begin{cases} k^l & if\ p \in \{\mathcal{D}^l \setminus \mathcal{D}^{l+1}\}, 1 \le l \le L \\ \tau & \text{if } p \in (\mathcal{D} \setminus \mathcal{D}^1) \end{cases} \quad (3)$$

4.3. Detecting Theme Object

As the commonness score $C(p)$ is obtained, then it can detect the theme object in each image using a bounding box. $\boldsymbol{I}i$, we search for the bounding box R^* with the maximum commonness score.

$$\mathcal{R}^* = arg\max_{R \subseteq Ii} \sum_{p \in R} C(p) = \arg\max_{R \in \wedge} F(R) \quad (4)$$

Where $F(R) = \sum_{p \in R} C(p)$ is the objective function and ^ denotes the candidate set of all valid sub images in $\boldsymbol{I}i$. To speed up this localization process, we apply the branch and bound search. The target bounding box R^* is determined by four parameters, i.e., top, bottom, left, and right positions in the image.

4.4. Differentiation of Multiple Thematic Objects

The branch-and-bound search can recognize all theme objects. We need to additionally differentiate them if there are multiple theme objects in the same image set. We thus conduct object clustering to differentiate theme objects. As each recognized object is distinguished by a subimage region R, we estimate the affinity relationship between two subimages R_i^* and R_j^*. First, for each subimage, we count the number of primitives whose k-SNN is supported by other subimages. For layer this number is estimated as,

$$N_i^l = |\{p(x,y) \colon p \in \mathcal{D}^l \wedge (x,y) \in \mathcal{R}_i^*\}| \quad (5)$$

Where (x, y) represent p's spatial location. For subimage R_i^*, we count the number of primitives whose k-SNN is carried by the other R_j^* and express this number as S_{i-j}^l. After counting these numbers for all layers l, the pair wise similarity is defined as,

$$A_{i,j} = \sum_{l=2}^{L} \frac{k^l \times S_{i-j}^l}{N_i^l} \quad (6)$$

If no visual primitive is similar between R_i^* and R_j^*, their affinity value is set to be a negative value, i.e., 1. Once affinity matrix A is gained, we start to group one common object by finding a dense subgraph $\Omega \subseteq \wedge^*$ as

$$\Omega^* = arg\max_{\Omega \subseteq \wedge^*} W(\Omega) \quad (7)$$

Where

$$W(\Omega) = \sum_{\mathcal{R}_i^*, \mathcal{R}_j^* \in \Omega} A_{i,j}$$

We employ a fixed-point iteration procedure in to solve "Equation 7", and to extract the heavy subgraphs one by one. For example, after finding the heavy subgraph, We can remove all of its nodes and continue to find the second heavy one, until no qualified subgraph can be found, i.e., the obtained subgraph only contains a single node. Multilayer pruning algorithm is summarized in Algorithm 1.

Algorithm 1: Discovering Of Theme Objects

Input: T unlabeled images with extracted local primitives $\mathcal{D} = I_1 \cup I_2 \cup \ldots \cup\ I_T$, and threshold λ
Output: thematic objects $\mathcal{R}_i^* \ \in \ \Omega^*$ in each image I_i
/ * Multilayer Candidate Pruning
foreach $p \in \ \mathcal{D}$ **do**
if M$p \neq \emptyset$ **then**
add p to $\mathcal{D}$
for $2 \leq \boldsymbol{l} \leq \boldsymbol{L}$ **do**
for $p, \boldsymbol{q} \in \mathcal{D}^{l-1}$ **do**
$\boldsymbol{Sp} = \{q\text{: Sim (Np, Nq)} \ > 0.5k^{l-1}\}$
if $|\boldsymbol{Sp}| > \lambda T$ **then**
add p to $\mathcal{D}^l$
for $2 \leq l \leq L$ **do**
if $\boldsymbol{p} \in \mathcal{D}^{\boldsymbol{l}} \ \setminus \mathcal{D}^{l+1}$ **then**
$C(p) = k^l$
if $p \in \mathcal{D} \setminus \mathcal{D}^{l+1}$ **then**
$C(p) = \tau$
/ * Thematic Object Localization
foreach $\boldsymbol{Ii}$ **do**
Obtain $\mathcal{R}_i^* = arg\max_{R\in\Lambda} F(R)$
add $\mathcal{R}_i^*$ *to a set* Λ^*

The below figures shows the sample results as figure 3.1 is the input selected video. Figure 3.2 shows the identification of theme object in video that is tracking of the movement of the object in the video.

The frames extracted from the video are identified first, features of each object in the image are extracted, pattern matching is done on the consecutive frames having the desired features in the hand, the motion vectors are calculated and mask is moved accordingly. The x and y directional flow components are calculated from the gradient of the image. The secondary differences in the x and y components of the image are used to find the final flow of intensities respectively. The laplacian gives the visualization of the edges for the proper tracking of the movement of the object in the video. This data are used for the extraction of visual primitives which will be further used by the classification algorithm. The moving object tracking in video pictures has attracted a great deal of interest in computer vision.

Fig.3.1. Selected input video

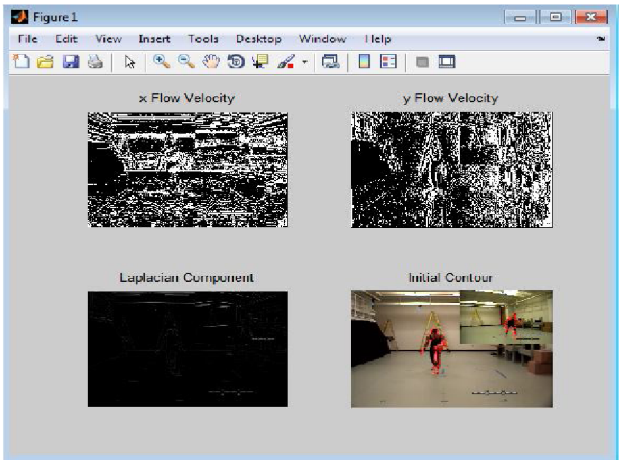

Fig.3.2. Tracking of the object in the video

V. Conclusion and Future Work

In this paper, the commonly appearing object in a video is recognized as by tracking of the movement of the object. There are three key steps in video surveillance: detection of interesting moving objects, tracking of such objects from frame to frame, and analysis of object tracks to recognize their behaviour. It is skilful to handle theme object variations due to lighting, scale, color, and view of point. Here we are identifying only one theme object at a time. Our future work covers demonstrations on both video sequences and image collections. Also covers identifying the multiple theme objects from the given collection of videos and images under the different theme object changes.

Acknowledgement

I am very glad to thanks my respected guide Prof. Dabhade R. Electronics and Telecommunications Department, for furnishing his creative ideas. Special thanks to SND College of Engineering for helping and for motivation.

REFERENCES

[1] Hongliang Li and King Ngi Ngan, "A co-saliency model of image pairs," IEEE Trans. Image Process., vol. 20, no. 12, pp. 3365–3375, December 2011.

[2] Jun-Bin Yeh, Chung-Hsien Wu, *Senior Member, IEEE*, and Sheng-Xiong Chang, "Unsupervised Alignment of News Video and Text Using Visual Patterns and Textual Concepts," IEEE Transaction on multimedia, vol.13, no.2, pp. 206-215, April 2011.

[3] Davide Liu, Tshun Chen: "A Framework for Discovering Objects in Video," IEEE Transactions on Multimedia, vol. 10, no.2, pp. 200-208, February 2008.

[4] Junsong Yuan, Gangqiang Zhao, Yun Fu, Zhu Li, Aggelos K. Katsaggelos and Ying Wu, "Discovering Thematic Objects in Image Collections and Videos," IEEE Transactions on Image Processing vol.21, no. 4, pp. 2207-2219, April 2012.

[5] Shalini N. And Sharada A. "Discovering Thematic Object in a Video," IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727 vol. 13, pp.33-37, Issue 2 July-August 2013.

[6] Jianbo Shi and Jitendra Malik, "Normalized cuts and image segmentation," IEEE Transaction Pattern Anal. Mach. Intell., vol. 22, no. 8, pp. 888–905, August 2000.

[7] Zhengwei Yang and Fernand S. Cohen, "Image Registration and Object Recognition using Affine Invariant and Convex Hulls," IEEE Transaction on Image Processing, vol. 8, no. 7, pp. 934-946, July 1999.

[8] Hongxing Wang, Gangqiang Zhao and Junsong Yuan, "Visual pattern discovery in image and video data: a brief survey," WIREs Data Mining Knowedgel Discovery, vol. 4, pp. 24-37, January/February 2014