IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 12, December 2015

Survey on Sketch Based Image Retrieval

Dipika R. Birari[1], Prof. J.V. Shinde[2]

M.E Student, Kalyani Charitable Trust's Late G.N. Sapkal College of Engineering, Nasik, India [1]

Assistant Professor, Kalyani Charitable Trust's Late G.N. Sapkal College of Engineering, Nasik, India [2]

Abstract: Sketch-based communication is the oldest form of writing. In which sketch depicts rough shape of object. Sketch-based image retrieval (SBIR) can therefore be a very valuable information search tool. Although sketch is good way to express people's thoughts, there is a large gap in the appearance of user sketches and photorealistic images, when people sketch, they usually focus on the main structure of an object and only draw the semantic contour boundary. In contrast, photo-realistic images contain the color, texture and detailed shape of an object, which makes it very difficult to directly match a sketch and the corresponding photo-realistic image. Therefore, this is fundamental challenge in SBIR. The existence of noisy edges on photo realistic image degrades retrieval performance and to bridge this gap there is framework consisting of line segment descriptor named and noise impact reduction algorithm. Descriptor extracts edges and captures the relationship between them. Object boundary selection algorithm used to reduce the impact of noisy edges. The hypothesis is used to maximize retrieval score, for which multiple hypotheses are generated. In scoring process there are sometimes false matches happens, to reduce such distraction, two constraints on spatial and coherent aspects are used.

Keywords: Descriptor, sketch retrieval, edge based, histogram, line relationship.

I. INTRODUCTION

A Sketch is swiftly accomplished freehand picture which serves various purposes, it might trace something that artist visualize, it might trace or increase an idea for later use or it might also be used as a rapid means of graphically representing an idea and an image. A sketch is rough or unfinished drawing, often made to assist in making a more finished picture. A style of painting that resembles photography in its meticulous attention to realistic detail. Although edge extraction can bridge the appearance gap between sketches and photo-realistic images to some extent, it is quite common for noisy edges from background clutter, object detail and texture to be extracted with the object shaping edges. These noisy edges usually widen the appearance gap and degrade retrieval performance. Therefore, retrieval performance can be enhanced if the impact of noisy edges is reduced. Retrieval performance of the human visual system is not sensitive to these noisy edges since humans are able to distinguish object boundaries or contours from noisy edges based on their inference ability. Using this fact, algorithm can select the object boundaries from all extracted edges, the appearance gap can be filled and the performance of SBIR can be improved. This motivation provides with a new pathway to improved performance, which imposes a new requirement, i.e., that sketches/extracted edges should be treated as a set of lines, and the descriptors should be able to capture line-level features. This is because line-based descriptors give the flexibility to achieve edge selection or removal by setting the corresponding parts of the feature vector to a certain value, which is critical for boundary selection. Beside the need to solve the noise problem, that an effective descriptor for SBIR should be designed to describe lines and their relationships, rather than describing image patches, since a sketch/object boundary is essentially composed of lines (strokes) and the shape is determined by the relationships between these lines. Following figure shows the edge extraction using canny detection.

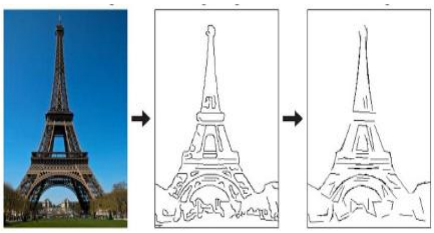

Fig 1: Edge Extraction

II. LITERATURE SURVEY

A. Canny Edge Detection

J. Canny [2] provided detection method is used to reduce amount of data in an image, by maintain its structural properties for further use in processing. This detection technique first detects the real edges by maximizing the signal-to-noise ratio. Then it applies localization and finds the number of responses. The source image and the thresholds can be chosen arbitrarily. Only a smoothing filter with a standard deviation of $\sigma = 1.4$ is supported. The implementation uses the correct Euclidean measure for the edge strengths. The different filter cannot be applied to edge pixels; this causes the output image to be 8 pixels smaller in each direction. The algorithm uses following steps:

- **Smoothing:** Blurring of the image to remove noise.
- **Finding gradients:** The edges should be marked where the gradients of the image has large magnitude.
- **Non maximum suppression:** Only local maxima should be marked as edges.

DOI 10.17148/IJARCCE.2015.412120

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 12, December 2015

- **Double Thresholding:** Potential edges are determined by thresholding.
- **Edge tracking by hysteresis:** Final edges are determined by suppressing all edges that are not connected to a very certain (strong) edges.

B. SBIR using patch hashing

CBIR is time consuming and user subjective. SBIR emerged as more expressive and interactive way to perform image search. K. Bozas[2] provides the technique which focused on both scalability and retrieval quality. In this image is represented by single set of visual words, and then extract local image patches represented with binary version of HOG descriptor which allows the utilization of min-hash algorithm. It provides more detailed image as well as flexibility during query time it omits searching of patches that not filled during drawing.

- **Min-hash algorithm:** It estimates similarity between two sets. The set overlap similarity between two sets D1 and D2 is defined as the ratio of their intersection and union and is a number between 0 and 1; it is 0 when the two sets are disjoint, 1 when they are equal, and between 0 and 1 otherwise. Min-hash has been successfully applied to text and image domains to detect near duplicate instances of a given set.

$$\text{sim}(D_1, D_2) = \frac{(D_1 \cap D_2)}{(D_1 \cup D_2)} \in [0,1] \qquad \ldots(1)$$

- **Preprocessing:** It bridges the gap between photos and sketches; this is achieved by extracting edge lines from the images. The well-known Canny algorithm requires in many cases manually tuning of the detection thresholds to return a desired edge map without many erroneous detected edges originating from background clutter.
- **Feature Extraction:** Grid is applied to finely describe the generated edge map and feature vectors are extracted for every patch of the grid. The size of spatial grid to 17 × 17 patches with each patch occupying 40 × 40 pixels. Two blocks are considered similar if they share similar shapes, i.e. their edges have similar orientation histogram and spatial arrangement, it shows this similarity with the HoG descriptor known to perform well in general object detection problems. The HoG algorithm further divides a patch in overlapping blocks and calculates an orientation histogram for each block.
- **Index construction and Patch retrieval:** The sketch collisions that will occur between similar images in a large database and then construct a hash table from the unique min-hash sketches. Every hash table entry can contain multiple pairs of identifier. Patches containing a small portion of edge are not taken into account during the index construction. Images similar to a sketch query are returned based on a voting process. The final ranking is generated by counting the votes for each image and sorting them in descending order. Patch queries can be executed independently in different machines and return a vote count for every image. An integration process will then merge all the votes to generate the final ranking.

C. Histogram of Oriented Gradients

The HOG descriptor is commonly applied in object recognition, and human detection tasks. HOG is a window based descriptor computed local to a detected interest point, which is provided by R. Hu and J. Collomosse[5]. Histogram of oriented gradients can be used as feature descriptors for the purpose of sketch retrieval, where the occurrences of gradient orientation in localized parts of a sketch image play important roles. The basic idea behind HOG is that the appearances and shapes of local regions within an image can be well described by the distribution of intensity gradients as the votes for dominant edge directions. Such feature descriptor can be obtained by first dividing the image into small contiguous regions of equal size, called cells, then collecting a histogram of gradient directions for the pixels within each cell, and finally combining all these histograms.

- Divide the image into small connected regions called cells, and for each cell compute a histogram of gradient directions or edge orientations for the pixels within the cell.
- Discretize each cell into angular bins according to the gradient orientation.
- Each cell's pixel contributes weighted gradient to its corresponding angular bin.
- Groups of adjacent cells are considered as spatial regions called blocks. The grouping of cells into a block is the basis for grouping and normalization of histograms.
- Normalized group of histograms represents the block histogram. The set of these block histograms represent the descriptor.

Gradient Field HOG

In order to encode the relative location and spatial orientation of sketches or canny edges of images, represent image structure using a dense gradient field interpolated from the sparse set of edge pixels. Begin with a mask of edge pixels, derived either from the mask of sketched strokes or from the Canny edge map of a photograph. A sparse orientation field is computed from the gradient of these edge pixels.

Bag of Visual Words (BoVW)

In a typical BoVW framework interest points are first detected and represented by descriptors. The GF-HOG features are extracted local to pixels of the canny edge map. Features from all images are clustered to form a single BoVW. A frequency histogramH^I is constructed foreach image, representing the distribution of GF-HOG derived visual words present in that image. The histogram is then normalized. At query time, a frequency histogram $H^S$is constructed from the query sketch by quantizing GF-HOG extracted from the sketch using the same codebook, and constructing a normalized frequency histogram of visual words present. Images are ranked according to histogram distance d (H^I, H^S) as defined.

D. SBIR on large scale database

One of the main challenges in image retrieval is to localize a region in an image which would be matched with the

DOI 10.17148/IJARCCE.2015.412120

query image in contour. To tackle this problem, they use the human perception mechanism to identify two types of regions in one image: the first type of region (the main region) is defined by a weighted centre of image features, suggesting that they could retrieve objects in images regardless of their sizes and positions. The second type of region, called region of interests (ROI), is to find the most salient part of an image, and is helpful to retrieve images with objects similar to the query in a complicated scene. For large scale databases, Eitz et al[3],presented an algorithm which divides an image into a fixed number of cells, and each cell corresponds to a tensor descriptor. Because of no database index structure, Eitz's algorithm must scan the whole database for each query.

- **The main region and region of interests (ROI):** The main region is defined to tackle the problem that one image only contains one scene (or object) similar to the query but different in size and position; ROI deals with one object similar to the query saliently appears in a complicated background.
- **Hierarchical Orientation Combination:** Human visual system processes images in a hierarchical structure. According to this mechanism, the hierarchical orientation combination is proposed first. H_iDenotes i^{th} level image resolution, H_{i+1} is higher than H_i. And O_j denotes j^{th}orientation, $C_q$denotes q^{th} RGB color component, D is difference image. The orientation information of an image is computed by:

$$D_{H_i O_j} = \max \{ D_{H_i O_j C_q} \} \quad \ldots(2)$$

Where, $D_{H_i O_j C_q}$ is the difference image at i^{th}level resolution, j^{th}orientation and q^{th} RGB color component. Max{}is the maximum value over three RGB components.

- **Candidate Region Estimation:** There are two candidate regions on one image. Main region estimation and ROI region estimation.

E. Visual Saliency Weighting and manifold ranking

T. Furuya and R. Ohbuchi[11],it employs Visual Saliency Weighting (VSW) to suppress background clutter in images. The features extracted from edge images processed by VSW are compared against the feature of a sketch query by using the Cross-Domain Manifold Ranking (CDMR), a distance metric learning algorithm adept at comparing heterogeneous feature domains.

- **Visual Saliency Weighting of edge image:** It converts (2D) images in a database into saliency-weighted edge images for comparison with sketches. The algorithm first computes Canny edge image from the database image. Then, edges due to background clutters are suppressed by using Visual Saliency Weighting (VSW). Visual saliency map is computed by using the MRSD algorithm. The "background-ness" is propagated from image periphery at four sides of the image toward the center. The "foreground-ness" is propagated from the foreground regions over the graph of super pixels.
- **Cross-Domain Manifold Ranking (CDMR):** The CDMR consists of two stages; Cross-Domain Manifold (CDM) generation stage and relevance diffusion stage. In the CDM generation

A CDM matrix **W** is generated. **W** is a graph whose vertices are the features from a sketch domain and an image domain. The similarity w(i, j) is computed by using the equation after normalizing the distance d(i, j) of features i and j to range [0,1].

$$W(i,j) = \begin{cases} exp(-d(i,j)/\sigma) & if\ i \neq j \\ 0 & otherwise \end{cases} \quad \ldots(3)$$

F. Descriptor for large scale image

Humans would probably describe different parts of the image and use different words depending on the cultural or professional background. M. Eitz[12] shows the task of comparing a rough sketch of feature lines to an image is natural yet difficult.

- **Edge Histogram Descriptor (EHD):** For each cell, compute gradient orientations and insert them into the corresponding histogram bin. Weigh each entry by its squared length based on the assumption that relatively stronger gradients are more likely to be sketched by the user. Let h_{ij} be the histogram of cellC_{ij} with d bins, Then for the computation of distances between histograms, first compute normalized histograms H_{ij} to account for the possibly different number of gradients in two corresponding cells:

$$H_{ij} = \frac{1}{\sum_k h_{ij}(k)} h_{ij} \quad \ldots(4)$$

Now let Hi j and ˜Hi j denote two normalized histograms. Let di j denote the L1 distance between Hi j and ˜Hi j :

$$d_{ij} = \sum_k |H_{ij}(k) - \tilde{H}_{ij}(k)| \ldots(5)$$

The distance between two edge histogram descriptors H and ˜H as:

$$\text{dist}(H, \tilde{H}) = \sum_i \sum_j d_{ij} \ldots(6)$$

- **Tensor Descriptor:** A compact representation of all information, yet not including the sign of the gradients, is given by the structure tensor G_{ij}. In order to detect similarly oriented image edges independently of the magnitude of the edges, store the structure tensor normalized by its Frobenius norm:

$$T_{ij} = \frac{G_{ij}}{||G_{ij}||F} \ldots(7)$$

The distance di j between two tensors an T_{ij} and ˜T_{ij} as the Frobenius Norm of the difference between the two tensors:

$$d_{ij} = ||T_{ij} - \tilde{T}_{ij}||F \quad \ldots(8)$$

The distance between two tensor descriptors as the sum over the tensor distances in their corresponding cells:

$$\text{dist}(T, \tilde{T}) = \sum_i \sum_j d_{ij} \ldots(9)$$

III.PROPOSED SYSTEM

This work contributes to the problem of multi-view object detection where same object should be detected from

different angle. This can be done using contour analysis on object. To do this, we need to generate Bag-of-Words using SURF features when we generate multiple hypotheses in Object Boundary Selection algorithm. Proposed system contains pre-processing of image so that we can get image with extracted edges. To remove noisy edges the descriptor designing is important and using that descriptor the boundary of object gets selected using hypotheses. In contour analysis, photo-realistic image and sketch image is analysed and points are calculated, and after comparison of that point, we can retrieve the matches. The goal of this paper is noise impact reduction and improves the retrieval performance. Therefore, we also attempt to address multi-view object detection in same. Following figure shows the architecture of proposed system.

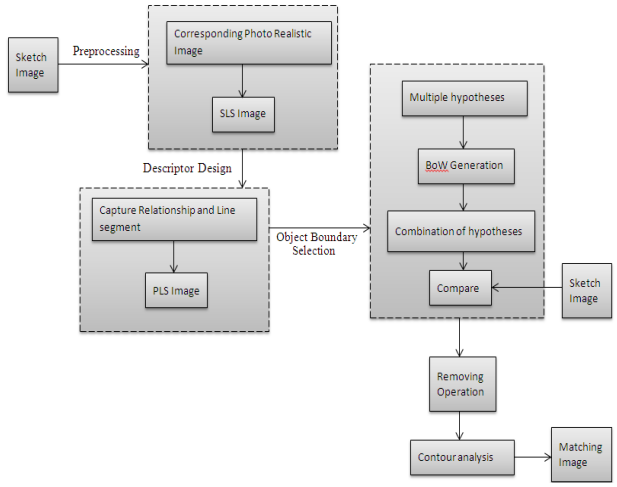

Fig2. Proposed System Architecture

IV. CONCLUSION

The proposed system extracts the edges using canny edge detector and then by applying descriptor, it enhance the performance of the image retrieval. A systematic approach that bridges the appearance gap for SBIR by considering sketches and extracted edges from a new angle, i.e., treating them as a set of line segments, laying the foundation for better sketch/extracted edge description and noise impact reduction using canny edge detector. Although this method achieves significant performance improvement in SBIR, this might be further improved to decrease the impact of quantization errors in the descriptor mapping procedure.

REFERENCES

[1] Shu Wang, Jian Zhang,Tony X. Han, and Zhenjiang Miao, "Sketch-Based Image Retrieval Through Hypothesis-Driven Object Boundary Selection With HLR Descriptor" in IEEE transaction on multimedia, Vol. 17, No. 7, July 2015.

[2] K. Bozas and E. Izquierdo, "Large scale sketch based image retrieval using patch hashing," Adv. Visual Comput., vol. 7431, pp. 210–219, 2012.

[3] R. Zhou, L. Chen, and L. Zhang, "Sketch-based image retrieval on a large scale database," in Proc. 20th ACM Int. Conf.Multimedia, 2012, pp. 973–976.

[4] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, "Sketch-based image retrieval: Benchmark and bag-of-features descriptors," IEEE Trans. Vis. Comput. Graph., vol. 17, no. 11, pp. 1624–1636, Nov. 2011.

[5] R. Hu and J. Collomosse, "A performance evaluation of gradient field hog descriptor for sketch based image retrieval," Comput. Vis. Image Understand., vol. 117, no. 7, pp. 790–806, 2013.

[6] M. Eitz, J. Hays, and M. Alexa, "How do humans sketch objects," ACM Trans. Graph., vol. 31, no. 4, pp. 44:1–44:10, Jul. 2012.

[7] R. Hu, T. Wang, and J. Collomosse, "A bag-of-regions approach to sketch-based image retrieval," in Proc. IEEE Int. Conf. Image Process., Sep. 2011, pp. 3661–3664.

[8] S. Salve and K. Jondhale, "Shape matching and object recognition using shape contexts," in Proc. 3rd IEEE Int. Conf. Comput. Sci. Inf. Technol., 2010, vol. 9, pp. 471–474.

[9] J. Yao, M. Li, Z. Li, L. Zhang, and W.-Y. Ma, "Natural image retrieval with sketches," in Proc. IEEE Int. Conf. Multimedia Expo, Jul. 2005, pp. 1198–1201.

[10] J.Canny, "A computational approach to edge detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[11] T. Furuya and R. Ohbuchi, "Visual saliency weighting and cross-domain manifold ranking for sketch-based image retrieval," in Proc. Int. Conf. Multimedia Modeling, 2014, pp. 37–49.

[12] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, "A descriptor for large scale image retrieval based on sketched feature lines," in Proc. 6th Eurograph. Symp. Sketch-Based Interfaces Modeling, 2009, pp. 29–36.

[13] P. Sousa and M. J. Fonseca, "Sketch-based retrieval of drawings using spatial proximity," J. Vis. Languages Comput., vol. 21, no. 2, pp. 69–80, 2010.

[14] C. L. Zitnick, "Binary coherent edge descriptors," in Proc. 11th Eur. Conf. Comput. Vis.: Part II, 2010, pp. 170–182.

[15] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," California Inst. Technol., Pasadena, CA, USA, Tech. Rep. CNS-TR-2007-001, 2007.

[16] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recogn., Jun. 2005, vol. 1, pp. 886–893.

[17] S. Salve and K. Jondhale, "Shape matching and object recognition using shape contexts," in Proc. 3rd IEEE Int. Conf. Comput. Sci. Inf. Technol., 2010, vol. 9, pp. 471–474.

[18] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2photo: Internet image montage," in Proc. ACM SIGGRAPH Asia, 2009, pp. 124:1–124:10.

[19] Y. J. Lee, C. L. Zitnick, and M. F. Cohen, "Shadowdraw: Real-time user guidance for freehand drawing," in Proc. ACM SIGGRAPH, 2011, pp. 27:1–27:10.