HORIZONTAL FLIP-INVARIANT SKETCH RECOGNITION VIA LOCAL PATCH HASHING

Konstantinos Bozas and Ebroul Izquierdo

School of EECS, Queen Mary University of London
{k.bozas, e.iqzuierdo}@qmul.ac.uk

ABSTRACT

This paper introduces a flip aware patch matching framework that facilitates scalable sketch recognition. An overlapping spatial grid is utilized to generate an ensemble of patches for each sketch. We rank similarities between freely drawn sketches via a spatial voting process where similar patches in terms of shape and structure arbitrate for the result. Patch similarity is efficiently estimated via the min-hash algorithm. A novel spatial aware reverse index structure ensures the scalability of our scheme. We show the benefits of horizontal flip invariance and structural information in sketch recognition and demonstrate state-of-the-art results in two challenging sketch datasets.

1. INTRODUCTION

Machine understanding of everyday human activities and actions consists a fundamental challenge for computer vision. Sketching to express feelings or elaborate on a topic is a task dated to prehistoric times, yet it is still contemporary due to the outbreak of the touch screen technology. Sketch understanding requires little effort from humans. Furthermore, neuroscience studies [1, 2, 3] have shown that humans can decode complex natural scenes from simple line drawings. Evidently, sketching is an efficient and intuitive communication tool between humans. Human computer interaction could therefore benefit from this expression channel given successful machine interpretation of human sketches. Towards this direction a large database of 20,000 free hand drawn sketches [4] has motivated the study of how humans draw sketches. Computational recognition of line drawings is a challenging task due to the abstract nature of sketching and the inter and intra-class variations between drawings. Moreover, traditional object recognition techniques can not be directly applied to the sketch domain as it is characterized by lack of modalities such as color or texture. Recent research [4, 5, 6], identified shape and structure as key properties for robust sketch matching.

In this paper, we present a scalable sketch matching approach based on shape and structure similarity. Our algorithm generates a matching score between two sketches by counting their local region correspondences. We establish region correspondences based on similar patches in terms of shape that are located in nearby positions. Min-hash, a set similarity estimation technique originally applied to identify duplicate web pages [7] and later modified for near duplicate image search [8], provides scalability in our scheme. Our method differs from [8], where an image is described by a single set of min-hash values. We extract a sequence of min-hash values for each local patch and rely on a novel spatial aware index scheme to enforce holistic structure affinity and infer a ranking on the indexed sketches. The generated ranking can be exploited for robust sketch recognition. Furthermore, we propose a modification of our algorithm invariant to reflection symmetry across the vertical axis and we show that it can drastically improve recognition performance. We perform extensive experiments in two challenging sketch datasets with various appearance features and demonstrate state-of-the-art results in low computational time.

2. RELATED WORK

Machine understanding of human drawn sketches is an open issue for researchers and has been studied since the primal era of computer evolution [9]. Early approaches focused on sketch domains of structured nature, like diagram recognition [10, 11, 12]. These approaches extract simplistic stroke features and cannot cope with the complexity of freely drawn sketches.

Advances in sketch based image retrieval [13, 14, 15] identified histograms of oriented gradients as a pertinent feature for the sketch domain. Supervised learning methods used a bag-of-features (BoF) [16, 14] representation of these features for free hand sketch classification and have shown promising results. BoF has been successful in generic object recognition [17, 18]. One of its drawbacks is the lack of spatial information in vector encoding. Recent research [5, 6] demonstrated the benefits of structural information in sketch understanding. In [5], a star graph model is employed to establish appearance and

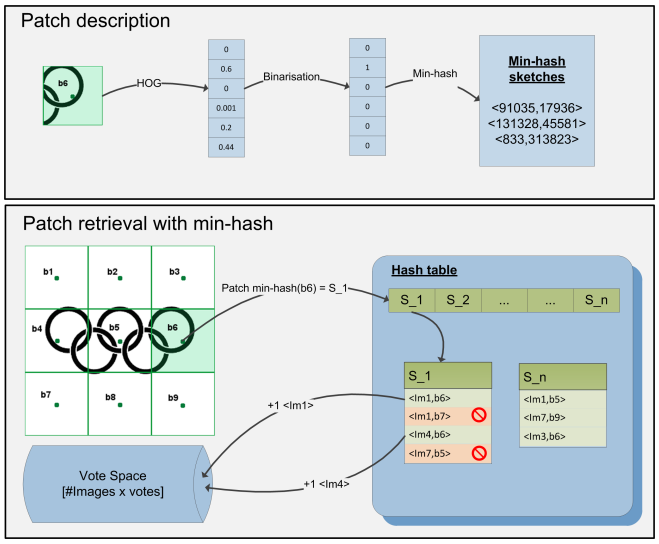

Fig. 1. Patch matching framework overview.

structure similarities between features. This approach is computationally expensive as several distance evaluations are carried out for each matched pair. In [6], a sketch to sketch retrieval algorithm is suggested using a spatial aware BoF variant to encode structure information.

3. SKETCH MATCHING BASED ON PATCH HASHING

In this section, we describe how we incorporate the unification of patch location and description in a scalable framework for efficient sketch matching. We match sketches based on accumulated shape similarities between local patches. Every sketch is divided into overlapping regions and for each region an appearance description is extracted. Min-hash is employed to estimate the similarity between two patches. An overview of the core modules of our method is presented in Figure 1. A reverse index is built on the unique min-hash value-location pairs pointing to the patches containing these values. A new input sketch query undergoes the same process and for each patch we look into the index to retrieve similar patches at nearby locations. Every index hit contributes a vote to the corresponding indexed sketch and the final ranking is generated by summing the votes for each sample. The ranking can be used to infer classification via the KNN algorithm.

3.1. Patch description

Feature extraction All sketches are scaled and centered inside a 256×256 canvas. An overlapping spatial grid is applied to finely describe an input drawing and feature vectors are extracted for every patch of the grid. Sketches contain sparse visual information, therefore local patches are adjusted to cover a large region of the image. The patch size and grid size parameters regulate how densely an image is sampled, controlling the detail of representation. The patch extraction process is visualized in the top part of Figure 1. Two patches are considered similar if they share shape characteristics, i.e. their strokes have similar orientation histogram and spatial arrangement. We suggest to quantify this similarity with the HOG descriptor known to perform well in general object detection problems. Moreover, descriptions relying on histograms of oriented gradients achieve superior performance in sketch based image retrieval, according to findings in literature [14, 15, 19, 20].

Binarization Descriptions extracted from the previous process return real valued histograms. In order to make the descriptor vector compatible for use with the min-hash algorithm, a binary representation is required. We modify the HOG vector to abide to this scheme. Without loss of crucial structure information we can binarize the descriptor by setting the $b\%$ highest values to 1 and the rest to 0. The binarization process is tailored to sketch matching as it highlights the strongest patch orientations corresponding to solid continuous contours while eliminating weak responses from noisy strokes. As we assess similarity between many local patch pairs there is no need for elaborate representations. This scheme captures the local shape of the sketches and by combining several local patch matches, it offers rich structure correspondences. The binarization step can be performed in linear time on the number of vector elements via a selection algorithm [21]. Finally, for each binarized descriptor we calculate k s-tuples of min-hash values which will be used to efficiently retrieve similar patches.

3.2. Location aware reverse index

To assess similarity between images one should count how many common min-hash values exist between the two patch collections. An appropriate data structure for this purpose, that allows constant-time look-ups is a reverse index hash table.

We would like to encode spatial information into our framework, hence we introduce spatial constraints in the matching scheme. The idea is to discard matches between distant regions. In other words, a successful match is defined between two patches that are visually similar and approximately located at nearby sketch regions. We propose a *key-location*-index structure index built on the collection of the database extracted min-hash sketches. An index key is defined for each unique min-hash tuple/location combination. For each key, we store the id

of the image a given patch originates from. The location information can be capitalized during the query process by rejecting non adjacent patches.

3.3. Horizontal flip matching through spatial voting

Images similar to a sketch query are returned based on a voting process (bottom of Figure 1). The pipeline of the query step is as follows: given a binary drawing, features are extracted according to the process illustrated in feature extraction paragraph. For every patch, k min-hash sketches are computed and for each sketch a look-up in the *key-location*-index is performed. If the key is found in the reverse index, we iterate through the entries and add a vote to the corresponding images. The locality constraint is enforced by discarding patches located further than a predefined distance radius r from the current examined patch. The location information is embedded in the hash key and a successful look-up in the table returns, in constant time, all the images that contain visually similar patches at the same location as the examined patch. In order to expand the spatial search radius r, we can generate *key-location* queries for each patch by fixing the key and inserting nearby location coordinates to check. An indexed image T is represented by a collection $\mathcal{T}$ of *key-location* values. A given *key-location* value v scores a hit on $\mathcal{T}$ if $v \in \mathcal{T}$.

$$s(\mathcal{Q},\mathcal{T}) = \sum_{v\in\mathcal{Q}} hit(v,\mathcal{T}). \quad hit(v,\mathcal{T}) = \begin{cases} 1, & \text{if } v \in \mathcal{T} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where v is a *key-location* hash value and $\mathcal{Q}$, $\mathcal{T}$ collections of *key-location* values. The final ranking is generated by counting the votes cast to each indexed sketch and sorting them in descending order. Classification can be achieved via the the KNN rule on the generated ranking. As in [5], category filtering can be applied via a learning algorithm before classification to narrow down complexity and improve recognition accuracy. In this paper, we apply category filtering via the SVM algorithm and keep the top N returned categories.

We observe in the large collection of hand drawn sketches of [4], the presence of reflection symmetry across the vertical axis between sketch pairs of the same category. We make our patching framework flip invariant across the vertical axis by generating a new horizontal flipped sketch $\mathcal{Q}_{flipped}$ for each new query $\mathcal{Q}$ and match both versions against the database. We keep the highest score among the two versions for each indexed exemplar.

$$s_{(f)}(\mathcal{Q},\mathcal{T}) = \max\{score(\mathcal{Q},\mathcal{T}), score(\mathcal{Q}_{flipped},\mathcal{T})\} \tag{2}$$

The suggested patch based matching scheme enhances flexibility since look-ups take place only for patches that have been drawn by the user, efficiently reducing query time and facilitating real time result updating when a new stroke is drawn. The ranking routine can be easily parallelized to enhance scalability even further.

4. EXPERIMENTS

4.1. Datasets and experimental setup

Datasets The evaluation is carried out in two challenging sketch datasets. The TU-Berlin [4] database consists of 20,000 hand-drawn sketches. It incorporates 250 object categories with each category being represented by 80 sketches. As the sketches are freely drawn by humans the dataset exhibits high variance over the categories. Humans recognize on average 73.1% of all sketches correctly. We also use the query set of the Flickr15k benchmark [15] as a second evaluation dataset. In total, there are 33 sketch categories describing shape, building landmarks, objects and scenes. Each category is represented by 10 sketches and some categories display high visual overlap.

Features and settings The HOG algorithm is selected to describe local patch appearances. We found that the min-hash parameters k and s have little effect on performance, hence we fix them to $k = 50$ and $s = 2$. We also globally fix the binarization threshold to top 20% of the vector values. At the voting stage we use the Manhattan distance and set the corresponding threshold to $r = 4$. For the KNN classification of the rankings we use K values $\{1, 3, 5, 7, 9\}$ and report the best score. Finally, category filtering is performed with SVM and we keep the top 5 categories in TU-Berlin and top 2 in Flickr15k. In the rest of this section, we denote as PHOG the patch hashing method with the corresponding descriptors. We prefix the category filtered results with the *SVM* keyword and suffix the flip invariant methods with the *flip* keyword.

Alternative methods We compare our algorithm against recent structured based techniques that demonstrated state-of-the-art results in the TU-Berlin dataset. Namely the star-graph model of Yi *et al.* [5] and the PHOG-A [6] algorithm. Additionally, we include comparisons against the baseline KNN and SVM methods in both datasets, built with the HOG features.

Metrics Following [5], we perform 4-fold cross validation on TU-Berlin dataset and 5-fold on Flickr15k. We measure the recognition accuracy on both datasets and additionally measure the Cumulative Matching Accuracy (CMA) and the Cumulative Best Matching Ac-

Table 1. Sketch recognition accuracy comparison.

	TU-Berlin		Flickr15k	
Method	**Unsupervised**	**Supervised**	**Unsupervised**	**Supervised**
KNN	45% [4]	N/A	$57.2\% \pm 3.7$	N/A
SVM	N/A	56% [4]	N/A	$76.9\% \pm 3.6$
Yi *et al.* [5]	53.3%	61.5%	N/A	N/A
PH-HOG	$56.2\% \pm 0.2$	$61.4\% \pm 0.3$	$74.2\% \pm 1.8$	$77.4\% \pm 3.6$
PH-HOG-flip	$\mathbf{58.5}\% \pm 0.2$	$\mathbf{62.8}\% \pm 0.2$	$\mathbf{75.7}\% \pm 2.8$	$\mathbf{77.8}\% \pm 4.7$

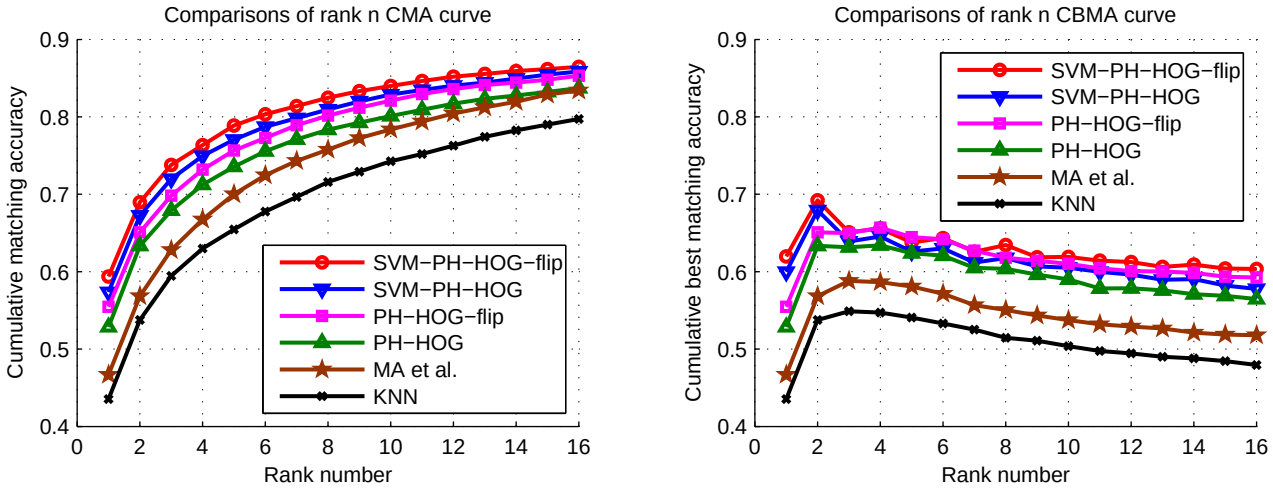

Fig. 2. Rank n CMA and CBMA curves in the TU-Berlin sketch data set. (Best viewed in color)

curacy (CBMA) in the TU-Berlin dataset. CMA shows how often the correct category appears in top n retrieved sketches, while CBMA measures the correctly retrieved sketches that account for the most of the top n retrieved sketches.

4.2. Discussion

Table 1 summarizes recognition accuracy over the two datasets. The category filtered (supervised) SVM-PH-HOG-flip algorithm achieves a new state-of-the-art score of 62.8% in the challenging TU-Berlin dataset. We also note that the unsupervised PH-HOG-flip outperforms the SVM and the unsupervised star-graph model by a large margin. Both our method and [5] are based on structured features. We attribute the superiority of patch hashing to the robust matching between local patches via the spatial voting and the binarization process that highlights the major patch orientations. Moreover, we verify that horizontal flip invariance improves the overall performance, as finer sketch matches are discovered. Results on the Flickr15k dataset are coherent with the findings on TU-Berlin, although the impact of flip invariance is less due to the low number of samples per category that leads to limited reflection variations within each class.

We further evaluate PH-HOG in the TU-Berlin dataset using the CMA and CBMA curves. The last 20 sketches of each category are used as queries. We compare patch hashing against Ma *et al.* [6] which has been especially developed for sketch retrieval. KNN classification [4] is also included in the evaluation as baseline. Figure 2 displays the curves. SVM-PH-HOG-flip achieves superior performance in both cases and maintains the edge over all ranks. Once more, flip invariance contributes to more robust results.

5. CONCLUSIONS

We presented a robust and scalable sketch recognition technique. Appearance and structure information is extracted from a sketch and captured in a spatial aware hash table. A binarization process further enhances strong continuous contours and facilitates the application of min-hash algorithm. We highlighted the significance of horizontal flip invariance in sketch recognition. State-of-the-art results were demonstrated in two challenging sketch datasets indicating the matching accuracy of our method and its benefits against competitive algorithms.

6. REFERENCES

[1] Alumit Ishai, Leslie G. Ungerleider, Alex Martin, and James V. Haxby, "The representation of objects in the human occipital and temporal cortex," *J. Cognitive Neuroscience*, vol. 12, no. Supplement 2, pp. 35–51, Nov. 2000.

[2] Dirk B. Walther, Barry Chai, Eamon Caddigan, Diane M. Beck, and Li Fei-Fei, "Simple line drawings suffice for functional mri decoding of natural scene categories," *Proceedings of the National Academy of Sciences*, vol. 108, no. 23, pp. 9661–9666, 2011.

[3] Bilge Sayim and Patrick Cavanagh, "What line drawings reveal about the visual brain," *Frontiers in Human Neuroscience*, vol. 5, no. 118, 2011.

[4] Mathias Eitz, James Hays, and Marc Alexa, "How do humans sketch objects?," *ACM Transactions on Graphics (Proceedings SIGGRAPH)*, vol. 31, no. 4, pp. 44:1–44:10, 2012.

[5] Yi Li, Yi-Zhe Song, and Shaogang Gong, "Sketch recognition by ensemble matching of structured features," in *In British Machine Vision Conference (BMVC)*, 2013.

[6] Chao Ma, Xiaokang Yang, Chongyang Zhang, Xiang Ruan, and Min-Hsuan Yang, "Sketch retrieval via dense stroke features," in *Proceedings of the British Machine Vision Conference*. 2013, BMVA Press.

[7] Andrei Z. Broder, Moses Charikar, Alan M. Frieze, and Michael Mitzenmacher, "Min-wise independent permutations," *Journal of Computer and System Sciences*, vol. 60, pp. 327–336, 1998.

[8] O. Chum, J. Philbin, and A. Zisserman, "Near duplicate image detection: min-hash and tf-idf weighting," in *Proceedings of the British Machine Vision Conference*, 2008.

[9] Nicholas Negroponte, "Recent advances in sketch recognition," in *Proceedings of the June 4-8, 1973, National Computer Conference and Exposition*, New York, NY, USA, 1973, AFIPS '73, pp. 663–675, ACM.

[10] Tevfik Metin Sezgin and Randall Davis, "Hmm-based efficient sketch recognition," in *Proceedings of the 10th International Conference on Intelligent User Interfaces*, New York, NY, USA, 2005, IUI '05, pp. 281–283, ACM.

[11] Tracy Hammond and Randall Davis, "Tahuti: A geometrical sketch recognition system for uml class diagrams," in *ACM SIGGRAPH 2006 Courses*, New York, NY, USA, 2006, SIGGRAPH '06, ACM.

[12] Christine Alvarado and Randall Davis, "Sketchread: A multidomain sketch recognition engine," in *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, New York, NY, USA, 2004, UIST '04, pp. 23–32, ACM.

[13] Konstantinos Bozas and Ebroul Izquierdo, "Large scale sketch based image retrieval using patch hashing," in *Advances in Visual Computing*. 2012, vol. 7431 of *Lecture Notes in Computer Science*, pp. 210–219, Springer Berlin Heidelberg.

[14] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, "Sketch-based image retrieval: Benchmark and bag-of-features descriptors," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 11, pp. 1624–1636, Nov 2011.

[15] Rui Hu and John Collomosse, "A performance evaluation of gradient field hog descriptor for sketch based image retrieval," *Comput. Vis. Image Underst.*, vol. 117, no. 7, pp. 790–806, July 2013.

[16] K. Bozas and E. Izquierdo, "Discriminant pairwise local embeddings," in *IEEE International Conference on Multimedia and Expo (Short Paper)*, July 2013, pp. 1–4.

[17] J. Sivic and A. Zisserman, "Video google: a text retrieval approach to object matching in videos," in *Proc. Ninth IEEE Int Computer Vision Conf*, 2003, pp. 1470–1477.

[18] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 4, pp. 509–522, Apr 2002.

[19] Rui Hu, M. Barnard, and J. Collomosse, "Gradient field descriptor for sketch based retrieval and localization," in *Proc. 17th IEEE Int Image Processing (ICIP) Conf*, 2010, pp. 1025–1028.

[20] Mathias Eitz, Kristian Hildebrand, Tamy Boubekeur, and Marc Alexa, "An evaluation of descriptors for large-scale image retrieval from sketched feature lines," *Computers & Graphics*, vol. 34, no. 5, pp. 482–498, 2010.

[21] Thomas H. Cormen, Clifford Stein, Ronald L. Rivest, and Charles E. Leiserson, *Introduction to Algorithms*, McGraw-Hill Higher Education, 2nd edition, 2001.