Diagram Image Retrieval and Analysis: Challenges and Opportunities

Liping Yang
University of New Mexico
Albuquerque, NM, USA
lipingyang@unm.edu

Ming Gong
University of Dayton
Dayton, OH, USA
gongm1@udayton.edu

Vijayan K. Asari
University of Dayton
Dayton, OH, USA
vasari1@udayton.edu

Abstract

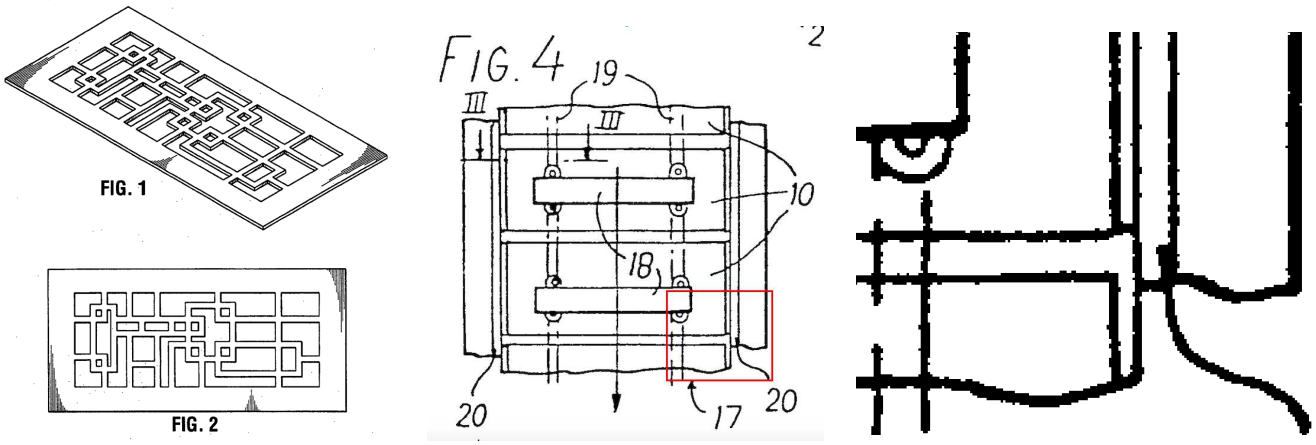
Deep learning has achieved significant advances for tasks such as image classification, segmentation, and retrieval; this advance has not yet been realized on scientific and technical drawing images. Research for technical diagram image analysis and retrieval retain much less well developed compared to natural images; one major reason is that the dominant features in scientific diagram images are shape and topology, no color and intensity features, which are essential in retrieval and analysis of natural images. One important purpose of this review, along with some challenges and opportunities, is to draw the attention of researchers and practitioners in the Computer Vision community to the strong needs of advancing research for diagram image retrieval and analysis, beyond the current focus on natural images, in order to move machine vision closer to artificial general intelligence. This paper investigates recent research on diagram image retrieval and analysis, with an emphasis on methods using content-based image retrieval (CBIR), textures, shapes, topology and geometry. Based on our systematic review of key research on diagram image retrieval and analysis, we then demonstrate and discuss some of the main technical challenges to be overcome for diagram image retrieval and analysis, and point out future research opportunities from technical and application perspectives.

1. Introduction and Motivation

Existing computer vision methods work well for natural images, but not for binary (black and white) technical drawing images (see [36, 20, 44] for recent evidence; and see Figure 1 for diagram image examples). Research on diagram images is much less well developed; major reasons are as follows: (1) natural images contain much more features (e.g., color, shape, intensity and texture), whereas technical diagram images (e.g., patent images) are usually binary with complex shapes, no color and little texture information [27, 29, 14, 12, 44, 36, 20]; (2) diagrams (e.g., patent images) were drawn by different people, the thickness of lines or the styles of drawings are varied. It will bring much difficulty in the process of contour extraction and accurate comparison [29]; (3) As these diagrams are usually from documents which are scanned to gray and become binary images by a threshold. Thus the localized zigzag noises are a bottle neck for diagram images analysis using existing methods [36, 20]; (4) diagram images often contain not only traditional visual parts in computer vision community but also annotation curves and arrows, along with labels such as text and numbers [31]. See diagram image examples in Figure 1 for an intuitive sense of the challenges.

Strong needs for advances in diagram image retrieval and analysis (DIRA) remain over two decades [28, 21, 11, 46, 29, 31, 17, 44, 40, 19]. For example, patent image retrieval plays an essential role in patent search and can further be combined with text-based image retrieval for accurate patent search, as images are an important element in patents and many experts use images to analyze a patent or to check differences between patents. Patent image search is one of the example domain for strong needs of advance in DIRA, many other domain applications can be enhanced by the advance in DIRA (detailed in Section 3.2.2). But up to now, this area of research is still lag much behind compared with those for natural images. One important goal of this review is to draw the attention of researchers and practitioners in the Computer Vision community to challenges and opportunities in diagram image domains, beyond current dominant focus on natural images, in order to move machine vision closer to artificial general intelligence.

Here, we provide a road map to the rest of the paper. Section 2 covers a systematic review on the state of the art methods for DIRA, specifically, CBIR-based (Section 2.1), texture-based (Section 2.2), shape-based (Section 2.3), topology and geometry-based (Section 2.4). Section 3 focuses on demonstrating some challenges (Section 3.1) and discussing potential opportunities (Section 3.2) in both technical (Section 3.2.1) and applications (Section 3.2.2) perspectives. The paper concludes in Section 4. For readability, we provide a list of abbreviations in Appendix A.



(a) An industrial designs diagram image from [4]

(b) A patent image from [2]

(c) A zoomed-in detail from (b).

Figure 1. Diagram image examples. Note the zigzag noise generated from scanning (see (c)) – commonly used computer vision techniques such as smoothing filters do not help to remove such type of noises because of the binary/grayscale nature of such types of images.

2. The State of the Art Methods for DIRA

In this section, we investigate the state of the art methods for DIRA, with a focus on approaches using CBIR (Section 2.1), texture (Section 2.2), shape (Section 2.3), and topology and geometry (Section 2.4).

2.1. CBIR-based methods

Color, shape, and texture are commonly used visual features in CBIR [43]. Traditional CBIR algorithms do not work well for diagram images (e.g., patent images), which are mostly binary with little texture and complex shape.[29], research on diagram image retrieval is not well developed.

Adaptive hierarchical density histogram (AHDH) [42] is one of the few effective methods proposed for patent diagram image retrieval. AHDH treats an image as a plane and the main idea is to calculate the distribution of black pixels on the white plane [11]. AHDH uses pyramid decomposition to extract both local and global distribution of density. The hierarchical decomposition of the image is generated by calculating the hierarchical geometric centroids. Specifically, AHDH calculates the centroid of the image plane, and then divides the image plane into four regions based on the centroid and calculates the distribution of black pixels in each region that serves as the local density estimation. This process is iterative, and the AHDH is finally obtained by concatenating the density features and the quantized relative density features [29, 11, 19]. AHDH uses the $L1$ distance to measure the similarity between an query image and images from a database [11]. AHDH is efficient and effective; it is able to deal with large binary image databases [19]. However, AHDH is not capable of sub-image retrieval and is not rotation invariant [11] (those challenges and opportunities are detailed in Section 3.1.1 and 3.2).

Similar to AHDH, a method using *hierarchical oriented gradient histogram (HOGH)* was proposed in [29]. HOGH extracts the local and global gradient distribution of an image as HOGH considers gradient distribution on different scales of image, whereas AHDH only focuses on the distribution of pixels. Specifically, HOGH division is based on the geometrical centroid of an image (i.e., the division varies according to the distribution of the black pixels of an image). The division process is illustrated in Figure 5. HOGH can be used for binary patent images containing very complex line drawings, which cannot be easily segmented into shapes. But like AHDH, HOGH is not rotation invariant.

2.2. Texture-based methods

To our knowledge, little research has been done using texture for binary and/or diagram image retrieval and analysis. Among many texture based methods for image retrieval [32, 7], *local binary patterns* (LBP) is a simple and powerful texture based method [30, 23], and most importantly it is computationally efficient. Thus, we have ran experiments using LBP for diagram images (detailed in Section 3.1.2).

2.3. Shape-based methods

The primary feature can be used for DIRA is shape, however, due to the challenges in diagram images, existing shape descriptors would often fail.

One of the most dominant descriptors for binary image retrieval is *shape context (SC)* [29]. SC is a descriptor for finding correspondences between point sets (i.e., binning of spatial relationships between points in a Polar coordinates system). It is efficient and effective matching algorithm, which proposed and further developed in [8, 9, 10]. SC has the following advantages: translation and scale invariant, and robust for small geometric distortion. However, SC is not robust to rotation [22]. SC was used in [22] for binary image retrieval with improved rotation invariant by looking for the direction where the most sampling points are included. Almost all existing SC-based methods ran experiments on data sets that each image only contains a simple and single object (e.g., hand, rabbit, see Figure 6), however, diagram images often contain very complex shapes formed by line drawings instead of one simple object. SC works well for "clean" shapes (e.g., handwriting, trademarks), but struggles with noise. Diagrams (e.g., patent images) often contains heavy zigzag noise (see (c) in Figure 1). Also, SC algorithm acquires high accuracy with the cost of high computational complexity both in descriptor extraction and image matching. For complex diagram binary images, numerous points are required to make a good representation of an image and this will lead to high dimensions of features [29]. SC does not work well for diagram images retrieval, see our experiments detailed in Section 3.1.3.

SC and LBP was combined for shape representation and classification in [41]. However, all the data sets the authors ran their experiments are simple and single-shape objects. For the complex shapes formed by line drawings in diagram images, this method would fail. Also, the method requires contour extraction from a given shape. It is challenging to get the clear contours from diagram images due to the complex shapes formed by line drawings and also due the zigzag noise in such images.

The *region-based shape descriptor* (RBSD) was proposed recently in [37] for binary image retrieval. RBSD requires some image pre-processing such as cropping object region, and image resize. RBSD combines the following four features: angular radial histogram feature (ARHist), vertical histogram feature, horizontal histogram feature, and Zernike moment. Using Zernike moment makes RBSD rotation-invariant. Euclidean distance (Edist) is used as the similarity measurement. Edist for each features between two images are computed. The finalEdist is the sum of the four Edist. Smaller finalEdist means more similar. RBSD experiments running MPEG-7 CE Shape-1 Part-B data set [38] (see Figure 7 for some example images), has demonstrated good retrieval results. However, RBSD does not work well for diagram images retrieval, see our preliminary experiments detailed in Section 3.1.3. This is not too surprising, as the shapes in MPEG-7 data set is much simpler compare with those complex shapes in diagram images.

A shape descriptor using salient keypoints detection is proposed in [14, 13] for binary image retrieval, named binary salient keypoints (BSK) descriptor. BSK descriptor requires contour and key points extraction, after which the most salient keypoints are automatically detected by filtering out redundant and sensitive keypoints. Finally, for each keypoint, a feature vector is computed using the distribution of contour points in its local area. The BSK descriptor is evaluated on several public data sets, including silhouette images, handwritten math expressions, hand-drawn diagram sketches, and noisy scanned logos. Experimental results demonstrated the effectiveness of the method. The authors pointed out BSK is reliable when applied on challenging images such as fluctuated handwriting and noisy scanned images. Among all of the reviewed shape descriptors in this section, BSK is the one that is promising to tackle DIRA. However, as the method requires contour and keypoint detection, challenges will remain when it is applied to complex diagram images.

2.4. Topology and geometry-based approaches

Topology and geometry based approaches have achieved success in binary image retrieval, but most of those methods would fail for diagram image retrieval; some recent image representation and methods are proposed for diagram image analysis. These are now reviewed.

A geometry and topology based image retrieval system is developed for multi-object images in [6], in which an object refers to a connected set of foreground or background pixels and a structured representation called curvature tree (CT) is used to model both shape and topology of image objects. The hierarchy of the CT reflects the inclusion relationships between the image objects. To facilitate measuring the similarity based on shape, the triangle-area representation (TAR) [5] of each closed boundary of an object formed by the boundary points is stored at the corresponding node in the CT [6]. TAR is invariant to position, scale, and rotation, robust against noise and moderate amounts of deformations, and computationally efficient [5].

The method proposed in [6] has demonstrated effectiveness on the two data sets: Shape Retrieval Test on MPEG-7 CE-Shape-1 database [24] and Medical Image Retrieval Test [34]. Their evaluation is based on human relevance judgements. However, the method would not work well for diagram image retrieval due to the following two major reasons: (1) TAR needs *a closed exterior contour* as a prerequisite. So this method cannot solve the retrieval of binary diagram images which usually have extricated structure or open contours [29] (see Figure 1 for examples). (2) the major limitation of the approach [6] is that it deals primarily with binary shapes, and segmentation of objects in an image is assumed already done, and the object boundary is well identified. In practice, noise and partial occlusion

can drastically change the topology of the CT. This is very challenging for diagram images due to the complex shapes formed by line drawings (see Figure 1 for examples).

In [33], a binary image is decomposed into a set of *closed contours*, and each contour is represented as a chain code. To measure the similarity between two images, the distances between each contour in one image and all contours in the other are computed using string matching. Then, a weighted sum of the average and maximum of these distances is the final similarity. Two major advantages of the methods: (1) encoding closed contours is insensitive to translation, rotation, and size variations, thus the similarity measures based on the closed contours are invariant. (2) it can retrieve partially similar trademark images. However, this method cannot solve DIRA for the following two reasons: (1) diagram images have extricated structure or open contours [29]. (2) the method relies on edge detection to get the borders of the shapes. The state of the art methods based on edge detection often fail for line detection in diagram images [20].

Diagram images pose multiple challenges that natural images do not have. To tackle zigzag noise (see Figure 1 (c)), which is very common in (digitized) diagram images, a novel image representation called *skeleton graph* (SG) is proposed in [48]. SG is a simple yet powerful image representation to deal with diagram images. A SG is a topological graph generated from the skeleton of an image (see Figure 8 in Appendix B for an illustration). Two major advantages of SG: (1) SG converts topological information in a diagram image effectively from raster to vector. Advanced applications such as diagram image retrieval can be built on top of the SG representation. (2) SG does not rely on edge or contour detection, which is an advantage over previous topology-based method for image analysis. Most existing patent image retrieval systems rely on edges extracted from images, the performance of which is often affected by the quality of edge detection [49]. SG was used in [36] for effective denoising of persistent zigzag noises (see Figure 1 (c) for such type of noise) from low quality binary diagram images. Based on SG, a robust straight line segment detector called *TGGLines* for low quality binary diagram images was proposed in [20].

2.5. Other methods

This section provides other methods that do not fall into the categories introduced above. See details in Appendix C.

3. DIRA Challenges and Opportunities

In this section we show some technical challenges (Section 3.1), and then discuss and provide opportunities from technical and application perspectives (Section 3.2).

3.1. Technical challenges

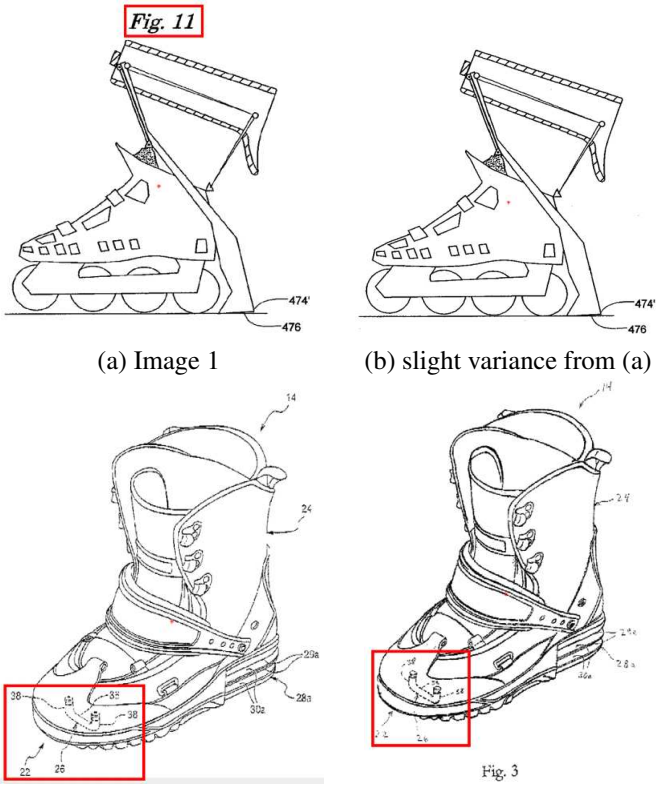
Color, texture, intensity, and shape are commonly used features in computer vision and image processing, but for many diagram images (e.g., patent and industrial design images), shape and topology are the most important features, due to binary/gray-scale nature of those images. We have experimented multiple methods for diagram images retrieval and analysis. Below we demonstrate some of the challenges we have met (Sections 3.1.1 to 3.1.3), followed by corresponding opportunities in Section 3.2.

3.1.1 AHDH related challenges

As introduced in Section 2.1, AHDH is not rotation invariant and cannot perform sub-image matching (due to space limit, see Appendix D.1 for detailed illustrations). One more downside of AHDH, beyond not rotation invariant and lack of sub-image matching capability, is that it is not robust. This is illustrated in Figure 2, where the centroid locations are marked as a red star in each image. Take a close look at Figure 2 (a) and Figure 2 (b), the centorid location is shifted, the only difference between the two images are that (a) has a text label "*Fig. 11*" (highlighted by the red box), where in (b) the text is removed. The shifting of the centroid will cause a different partition and thus will have an significant impact on the image retrieval results. Edist between Figure 2 (a) and Figure 2 (b) is 4.2243. Similarly, Figure 2 (c) and Figure 2 (d) are the same shoe but with different digits annotation labels (highlighted in red boxes), and image in (d) has the text label "*Fig.3*". Edist between Figure 2 (c) and Figure2 (d) is 12.0803. It is not a difficult task for humans to tell the two pairs of images are very visually similar but not so for machines. The results of AHDH imply the importance of pre-processing (e.g., removal of annotation text label and extraction of exterior boundary if any). From our preliminary experiments of using AHDH for diagram images, AHDH is not rotation invariant, lack of partial image matching power, and is very sensitive to text annotations, which are very common in diagram images.

3.1.2 Texture related challenges

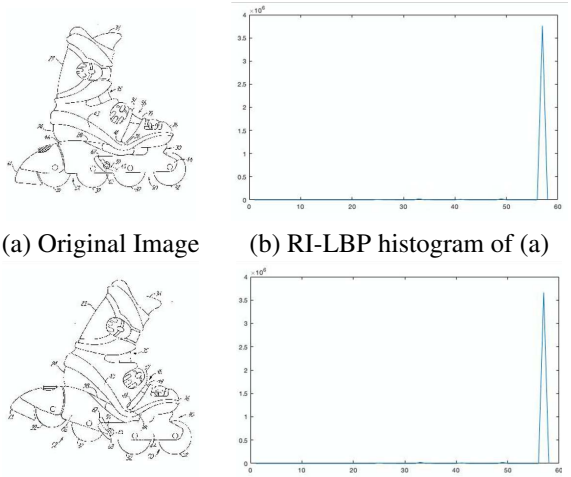
Local binary patterns (LBP) is a simple but powerful approach to capture local texture structures in images. Thus, we have applied an improved LBP[30] called rotation-invariant LBP (RI-LBP) [26] to the patent diagram image data set[45, 3]. Due to space limitation, see Appendix D.2 for detailed introduction to how LBP and RI-LBP works. Figure 3 provides the RI-LBP results for two diagram images from the data set [3]. RI-LBP is indeed rotation-invariant (see Figure 3 the RI-LBP histograms for the two images are the same). However, from the results shown in Figure 4, we can see that the peak occurs at the same po-



(a) Image 1 (b) slight variance from (a)

(c) Image 2 (d) slight variance from (c)

Figure 2. AHDH's robustness needs to be improved. Note that (b) is exactly the same as (a), but without the text "*Fig. 11*" highlighted in red box in (a); (d) is exactly the same shoe as (c), but (d) has the annotation text "*Fig. 3*" and the annotation number highlighted in the two red boxes are different.



(a) Original Image (b) RI-LBP histogram of (a)

(c) Rotated (a) (d) RI-LBP histogram of (c)

Figure 3. RI-LBP results for the (same) Images.

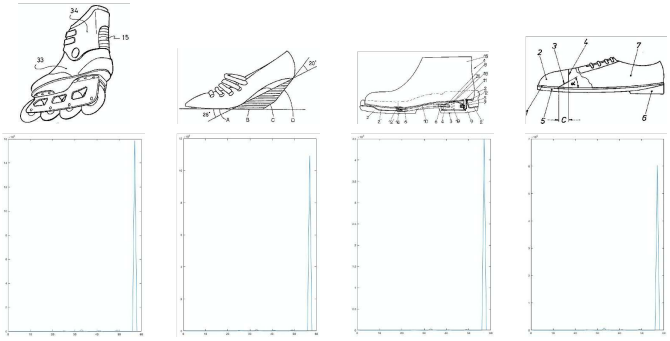


Figure 4. RI-LBP results for diagram images from different categories.

sition with different amplitude, which is caused by the binary nature of diagram images. Our experiments demonstrated that it is not effective to use texture feature alone for diagram image retrieval. LBP does not work well for diagram images that has little texture. Our experiment results align well with our brief review about texture-based method in Section 2.2: few binary/diagram images contain texture information – the reason why little work has been developed in the literature using texture-based methods for binary/diagram image retrieval.

3.1.3 Shape descriptor related challenges

We have done preliminary experiments for diagram images using two of the shape descriptors for binary image retrieval that we have reviewed in Section 2.3: (1) *shape context* (SC), as it is the dominant shape descriptor for binary image retrieval in the literature, and (2) *region-based shape descriptor* (RBSD), as it is a recent shape descriptor. The experiments results are provided below.

Using SC for diagram image retrieval SC is introduced in Section 2.3. Here, we only provide our preliminary experiment SC results (see Table 1) for diagram image matching. In Table 1, "Std diff" refers Pearson's chi-squared test, which is used to count cost matrix ; and "Cos diff" represents cosine distance. From the last two columns in Table 1, the computed shape difference cannot effectively and correctly measure the visual similarity for the three pairs of diagram images. The SC-based shape difference tells us the middle pair of images is much less similar compared with the bottom pair of images. This is not true. The visual similarity rank for the three pairs in Table 1 should be the following: (1) the middle pair of images are the most visually similar, as the only difference is the text label "Fig. 11" in image 1, and not in the other; (2) the top pair of images are visually similar, as the shoe shape is exact the same in the two images. The only difference is the annotation labels,

including the text notation "Fig. 3" in image 2; (3) the bottom pair of images are apparently not visually similar. We can see that SC descriptor is confused by text annotation in diagram images. Thus, SC doesn't work well for diagram images retrieval, it needs to be improved.

Using RBSD for diagram image retrieval As the RBSD shape descriptor (introduced in Section 2.3) is rotation invariant and it is for binary image retrieval, we ran some experiments on diagram images using the shape descriptor. (As the authors of RBSD do not provide code [37], we re-implemented it.) RBSD results for some images taken from the two data sets (patent diagram images [3] and MPEG shapes [38]) are shown in Table 2, where E_A, E_H, E_B, E_Z represents the Edist of angular radial histogram, horizontal histogram, vertical histogram and Zernike moment respectively; E_{RBSD} is the sum of E_A, E_H, E_B, E_Z. From Table 2, we can see that E_{RBSD} values cannot effectively and accurately tell the visual similarity among the three pairs of patent diagram images (i.e., the top three pairs in the table), although it does tell the 2nd pair of images is the most similar among the three pairs of patent images. RBSD fails to rank that the top image pair is more similar than the third pair. (Note that the bottom two pairs are for reference purpose. 4th pair shows the similar images from the MPEG-7 shape data set and the Edist is small, while 5th pair is opposite. This demonstrates the RBSD indeed works well for binary images that only contains a simple and single shape in each image.) In summary, RBSD shape descriptor cannot work for patent diagram images. See Appendix D.3 for further technical discussion about why RBSD works for shape images but not for patent diagram images.

3.2. Discussions and opportunities

From our systematic review (Section 2) and from the challenges demonstrated by our preliminary experiments (Section 3.1), we see that the methods and similarity measure metrics for diagram images are lagged far behind those for natural images. However, this also indicates there is a very large and novel research space for diagram images both from methodology and application perspectives. Among many research opportunities, below we provide some major ones we have identified through our systematic review (Section 2) and based on our preliminary experiments (Section 3.1). We group the opportunities into two sets: technical perspective (Section 3.2.1) and application perspective (Section 3.2.2).

3.2.1 Opportunities from technical perspective

See below for some potential research opportunities we have identified from technical perspective.

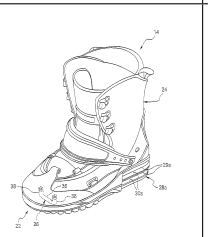
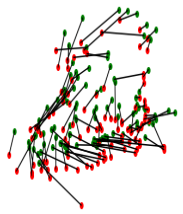
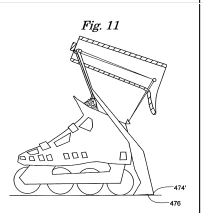
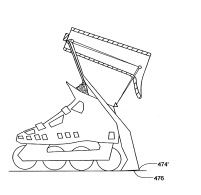
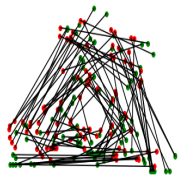
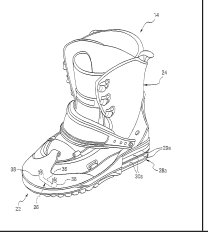
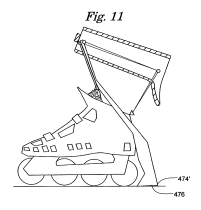
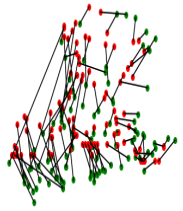
General directions: Potential general directions to advance research in diagram images retrieval and analysis includes but are not limited to: new image representations, methods and algorithms, and similarity metrics, as well as psychology inspired conceptual schemes and theories.

Large representative benchmark data sets: From the review in Section 2, we have seen existing methods are mainly tested on binary image data sets that often only contain simple and single object shape in each image (e.g., MPEG-7 CE-Shape-1 database [24], see Figure 6 for an example of such data set), without complex shapes formed by line drawings (see Figure 1). The research for diagram images are much lagged behind those for natural images. One of the main reasons and challenges are lack of large representative benchmark data sets. These are challenges and also opportunities for future research. However, it is not an easy task to prepare for large real (not synthetic) data sets for diagram image retrieval. Open a research portal platform that allows volunteers to "donate" visually similar diagram images in their domain (e.g., industrial/art design and patent images) would be a possible solution to generating large benchmark data sets for DIRA. Some data sets close to diagram images are provided in Appendix E.

Specific opportunities: Relating the challenges demonstrated in Section 3.1, we identify the following opportunities: (1) *RI-AHDH*: as illustrated in Section 3.1.1, AHDH is not rotation invariant, and it does not have the power to perform sub-image (i.e., partial image) retrieval. RI-AHDH is a promising direction, as the literature has shown the efficiency and effectiveness of AHDH for diagram patent image retrieval (Section 2.1). (2) *Integration of RI-LBP*: as shown in Section 3.1.2, RI-LBP loses its power when using directly for diagram images. But, as RI-LBP is a powerful texture-based method for image retrieval, and not all diagram images lack of texture information, we suggest not to use LBP alone for DIRA, but integrate it with other methods, such as SC. (3) *Extension of SC*: SC is the dominant shape descriptor in the literature for binary image retrieval (Section 2.3). However, as demonstrated in Section 3.1.3, SC cannot be directly used for DIRA, it needs to be improved. Some research opportunities include: (1) make SC less sensitive to annotation labels (e.g., "Fig.3") that often appear in diagram images, and (2) improve SC's computational efficiency, as the number of key points in diagram images formed by complex line drawings can be very large (see the 3rd and 4th and 5th columns in Table 1). (4) *Extension of RBSD*: RBSD is a recent shape descriptor for binary image retrieval (Section 2.3). However, as demonstrated in Section 3.1.3, it does not work for DIRA. The features used in RBSD (especially the angular radial histogram, see our further discussion in Appendix D.3) can be improved to increase its capability for DIRA.

Topology related opportunities: The image representation called *skeleton graph* (SG) proposed in [48] (introduced in Section 2.4) has been demonstrated useful and ef-

Table 1. Diagram image matching results using shape context (point-to-point matching).

Image 1	Image 2	Shape extracted from image 1	Shape extracted from image 2	Matching	Std diff	Cos diff
					20.01	0.40
					36.93	0.77
					22.04	0.48

fective for diagram image analysis (see [36, 20] for recent successful usage of the image representation for diagram image analysis). It is promising to develop efficient DIRA methods that are able to both exact and partial diagram image matching, based on the SG image representation extracted from diagram images, and then by using matching based on spatial similarity.

3.2.2 Opportunities from application perspective

Visual similarity based image retrieval for diagram images has many potential applications, not just for patent image retrieval. We highlight just a few of these below.

Patent image retrieval: It is a direct and the most complex and challenging DIRA application domain. As patent images involve in diverse disciplines and different types of technical drawings that contain not only visual part but also mixed with text annotation such as figure labels.

OCR and text recognition: Character and diagram images share some common properties (e.g., characters can be viewed as line drawings, and are often in binary/grayscale). Advances in DIRA will improve research and applications in optical character recognition (OCR) and text recognition.

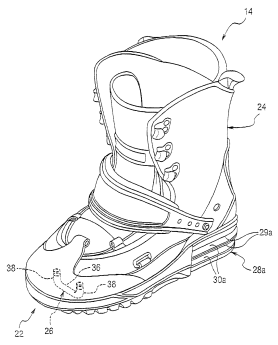
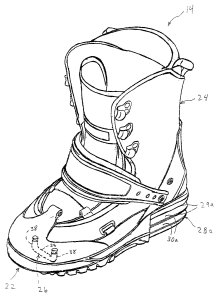
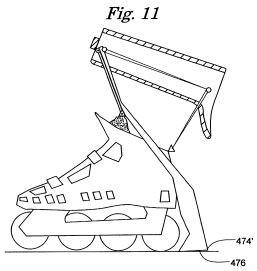
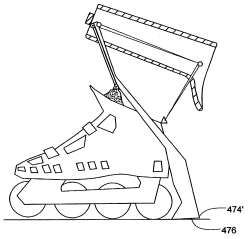
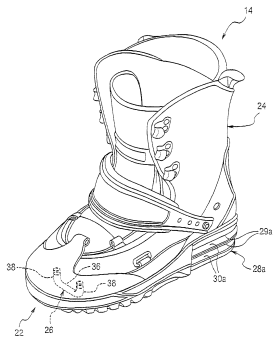
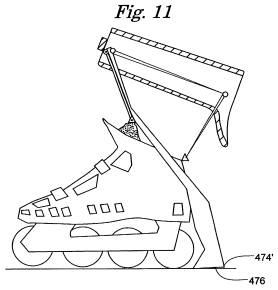
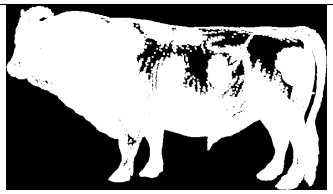
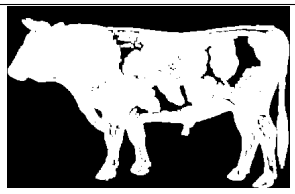
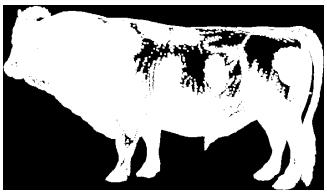
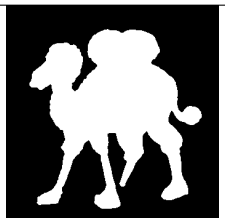
Industrial and art design: When a designer have a design draft, the designer will be inspired by visually similar existing designs (could be designs from different domains) if a visually similar image retrieval interface available, and thus better designs can be generated.

Autonomous driving: Advance in DIRA will advance road Lane line detection in real world complex scenarios, including dashed lane line. The complexities of road conditions increase in real world situations. All existing autonomous vehicle systems assume that line markings exist, are clear and, more importantly, are visibly distinct. But in reality, most roads are in poor condition in different severity level (e.g., worn road markings, lane marking covered with dirt, falling leaves). Autonomous vehicles will have considerable difficulty driving on such roads.

4. Conclusion

Research for DIRA are much less well developed compared with those for natural images. To draw the attention of researchers and practitioners in the computer vision community to advance DIRA, we have provided a systematic review for DIRA. We have seen most existing methods, even those methods designed and demonstrated effective for binary image retrieval (which is closer to diagram images compared with natural images) do not work well for DIRA, due to the complex shapes formed by line drawings in diagram images. We also provide some chal-

Table 2. RBSD results for patent diagram images and shape images.

Image 1	**Image2**	E_A	E_V	E_H	E_Z	E_{RBSD}
	 Fig. 3	0.27	1.26	1.26	0.29	3.08
 *Fig. 11*		0.07	0.65	0.65	0.18	1.55
	 *Fig. 11*	0.30	0.93	0.93	0.19	2.35
		0.88	3.78	3.78	0.21	8.66
		1.31	11.83	11.86	0.20	25.17

lenges and thus some identified opportunities, specifically, CBIR-based method AHDH, texture-based method RI-LBP, SC and RBSD shape descriptors for binary image retrieval. Again, through our systematic survey on DIRA, little research has been done for this area, and many challenges thus opportunities await us ahead.

Acknowledgments

The authors are grateful to Dr. Brendt Wohlberg, Dr. Diane Oyen, Catherine Potts, and Manish Bhattarai for discussions relating to this work. We also thank the three anonymous reviewers for their helpful comments and suggestions.

References

[1] Line drawings of 3D shapes (artists' drawings), 2008. Available online: https://gfx.cs.princeton.edu/proj/ld3d/ (accessed on March 27, 2020). 14

[2] Multimedia knowledge and social media analytics laboratory (MKLab). 2000 Binary Patent Images Database, 2010. Available online: https://mklab.iti.gr/results/patent-image-databases/ (accessed on March 27, 2020). 2, 14

[3] Multimedia knowledge and social media analytics laboratory (MKLab). Concept Patent Image Database (8 concepts, 1042 images), 2010. Available online: https://mklab.iti.gr/results/patent-image-databases/ (accessed on March 27, 2020). 4, 6, 14

[4] Canadian industrial designs database, 2020. Available online: https://www.ic.gc.ca/app/opic-cipo/id/ (accessed on March 27, 2020). 2, 14

[5] Naif Alajlan, Ibrahim El Rube, Mohamed S Kamel, and George Freeman. Shape retrieval using triangle-area representation and dynamic space warping. *Pattern recognition*, 40(7):1911–1920, 2007. 3

[6] Naif Alajlan, Mohamed S Kamel, and George H Freeman. Geometry-based image retrieval in binary image databases. *IEEE transactions on pattern analysis and machine intelligence*, 30(6):1003–1013, 2008. 3

[7] Prithaj Banerjee, Ayan Kumar Bhunia, Avirup Bhattacharyya, Partha Pratim Roy, and Subrahmanyam Murala. Local neighborhood intensity pattern–a new texture feature descriptor for image retrieval. *Expert Systems with Applications*, 113:100–115, 2018. 2

[8] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape context: A new descriptor for shape matching and object recognition. In *Advances in neural information processing systems*, pages 831–837, 2001. 3

[9] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *IEEE transactions on pattern analysis and machine intelligence*, 24(4):509–522, 2002. 3

[10] Serge Belongie, Greg Mori, and Jitendra Malik. Matching with shape contexts. In *Statistics and Analysis of Shapes*, pages 81–105. Springer, 2006. 3

[11] Naeem Bhatti and Allan Hanbury. Image search in patents: a review. *International journal on document analysis and recognition (IJDAR)*, 16(4):309–329, 2013. 1, 2

[12] Naeem Bhatti, Allan Hanbury, and Julian Stottinger. Contextual local primitives for binary patent image retrieval. *Multimedia Tools and Applications*, 77(7):9111–9151, 2018. 1, 12

[13] Houssem Chatbri, Kenny Davila, Keisuke Kameyama, and Richard Zanibbi. Shape matching using keypoints extracted from both the foreground and the background of binary images. In *2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 205–210. IEEE, 2015. 3

[14] Houssem Chatbri, Keisuke Kameyama, Paul Kwan, Suzanne Little, and Noel E O'Connor. A novel shape descriptor based on salient keypoints detection for binary image matching and retrieval. *Multimedia Tools and Applications*, 77(21):28925–28948, 2018. 1, 3

[15] Forrester Cole, Aleksey Golovinskiy, Alex Limpaecher, Heather Stoddart Barros, Adam Finkelstein, Thomas Funkhouser, and Szymon Rusinkiewicz. Where do people draw lines? In *ACM SIGGRAPH 2008 papers*, pages 1–11. 2008. 14

[16] Forrester Cole, Kevin Sanik, Doug DeCarlo, Adam Finkelstein, Thomas Funkhouser, Szymon Rusinkiewicz, and Manish Singh. How well do line drawings depict shape? In *ACM SIGGRAPH 2009 papers*, pages 1–9. 2009. 14

[17] Gabriela Csurka. Document image classification, with a specific view on applications of patent images. In *Current Challenges in Patent Information Retrieval*, pages 325–350. Springer, 2017. 1

[18] Gabriela Csurka, Jean-Michel Renders, and Guillaume Jacquet. Xrce's participation at patent image classification and image-based patent retrieval tasks of the clef-ip 2011. In *CLEF (Notebook Papers/Labs/Workshop)*, volume 2, 2011. 14

[19] Ilias Gialampoukidis, Anastasia Moumtzidou, Stefanos Vrochidis, and Ioannis Kompatsiaris. Exploiting images for patent search. In *Springer Handbook of Science and Technology Indicators*, pages 889–906. Springer, 2019. 1, 2

[20] Ming Gong, Liping Yang, Catherine Potts, Vijayan K Asari, Diane Oyen, and Brendt Wohlberg. TGGLines: A robust topological graph guided line segment detector for low quality binary images. *arXiv preprint arXiv:2002.12428*, 2020. 1, 4, 7, 11

[21] Allan Hanbury, Naeem Bhatti, Mihai Lupu, and Roland Mörzinger. Patent image retrieval: a survey. In *Proceedings of the 4th workshop on Patent information retrieval*, pages 3–8, 2011. 1

[22] Yan He, Lei Yang, Yichun Zhang, Xiaoyu Wu, and Yun Zhang. The binary image retrieval based on the improved shape context. In *2014 7th International Congress on Image and Signal Processing*, pages 452–456. IEEE, 2014. 3, 11

[23] Nazgol Hor and Shervan Fekri-Ershad. Image retrieval approach based on local texture information derived from predefined patterns and spatial domain information. *arXiv preprint arXiv:1912.12978*, 2019. 2

[24] Longin Jan Latecki, Rolf Lakamper, and T Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, volume 1, pages 424–429. IEEE, 2000. 3, 6

[25] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 11

[26] Zhi Li, Guizhong Liu, Yang Yang, and Junyong You. Scale-and rotation-invariant local binary pattern using scale-adaptive texton and subuniform-based circular shift. *IEEE Transactions on Image Processing*, 21(4):2130–2140, 2011. 4, 13

[27] Shuang Liang and Zhengxing Sun. Sketch retrieval and relevance feedback with biased svm classification. *Pattern Recognition Letters*, 29(12):1733–1741, 2008. 1

[28] Doug Love and Jeff Barton. Aspects of design retrieval performance using automatic gt coding of 2d engineering drawings. In *4th Int. Conf. on Integrated Design and Manufacture in Mechanical Engineering*. Citeseer, 2004. 1

[29] Hui Ni, Zhenhua Guo, and Biqing Huang. Binary patent image retrieval using the hierarchical oriented gradient histogram. In *2015 International Conference on Service Science (ICSS)*, pages 23–27. IEEE, 2015. 1, 2, 3, 4, 11

[30] Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59, 1996. 2, 4

[31] Jeong Beom Park, Thomas Mandl, et al. Patent document similarity based on image analysis using the sift-algorithm and ocr-text. *International Journal of Contents*, 13(4), 2017. 1

[32] Jigisha M Patel and Nikunj C Gamit. A review on feature extraction techniques in content based image retrieval. In *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pages 2259–2263. IEEE, 2016. 2

[33] Hsiao-Lin Peng and Shu-Yuan Chen. Trademark shape recognition using closed contours. *Pattern Recognition Letters*, 18(8):791–803, 1997. 4

[34] Euripides G. M. Petrakis and A Faloutsos. Similarity searching in medical image databases. *IEEE transactions on knowledge and data engineering*, 9(3):435–447, 1997. 3

[35] Florina Piroi, Mihai Lupu, Allan Hanbury, and Veronika Zenz. Clef-ip 2011: Retrieval in the intellectual property domain. In *CLEF*, 2011. 14

[36] Catherine Potts, Liping Yang, Diane Oyen, and Brendt Wohlberg. A topological graph-based representation for denoising low quality binary images. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 0–0, 2019. 1, 4, 7

[37] Moe Zet Pwint, Thi Thi Zin, Mitsuhiro Yokota, and Mie Mie Tin. Shape descriptor for binary image retrieval. In *2016 IEEE 5th Global Conference on Consumer Electronics*, pages 1–2. IEEE, 2016. 3, 6

[38] Richard Ralph. Mpeg-7 core experiment ce-shape-1 [tar.gz]. 1999. 3, 6, 11

[39] Christoph Riedl, Richard Zanibbi, Marti A Hearst, Siyu Zhu, Michael Menietti, Jason Crusan, Ivan Metelsky, and Karim R Lakhani. Detecting figures and part labels in patents: competition-based development of graphics recognition algorithms. *International Journal on Document Analysis and Recognition (IJDAR)*, 19(2):155–172, 2016. 12

[40] Walid Shalaby and Wlodek Zadrozny. Patent retrieval: a literature review. *Knowledge and Information Systems*, pages 1–30, 2019. 1, 14

[41] BH Shekar, Bharathi Pilar, and Josef Kittler. An unification of inner distance shape context and local binary pattern for shape representation and classification. In *Proceedings of the 2nd international conference on perception and machine intelligence*, pages 46–55, 2015. 3

[42] Panagiotis Sidiropoulos, Stefanos Vrochidis, and Ioannis Kompatsiaris. Content-based binary image retrieval using the adaptive hierarchical density histogram. *Pattern Recognition*, 44(4):739–750, 2011. 2, 12

[43] Arnold WM Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on pattern analysis and machine intelligence*, 22(12):1349–1380, 2000. 2

[44] Wiebke Thode. Integration of images into the patent retrieval process. In *European Conference on Information Retrieval*, pages 359–363. Springer, 2019. 1

[45] Stefanos Vrochidis, Anastasia Moumtzidou, and Ioannis Kompatsiaris. Concept-based patent image retrieval. *World Patent Information*, 34(4):292–303, 2012. 4

[46] Stefanos Vrochidis, Anastasia Moumtzidou, and Ioannis Kompatsiaris. Enhancing patent search with content-based image retrieval. In *Professional Search in the Modern World*, pages 250–273. Springer, 2014. 1

[47] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017. 14

[48] Liping Yang, Diane Oyen, and Brendt Wohlberg. Image classification using topological features automatically extracted from graph representation of images. In *Proceedings of the 15th International Workshop on Mining and Learning with Graphs (MLG)*, 2019. 4, 6

[49] Ju-Hee Yoo and Kyoung-Mi Lee. Patent image retrieval using surf direction histograms. *Journal of KIISE*, 42(1):33–43, 2015. 4

A. Abbreviations

In this appendix, we provide the abbreviations (ordered alphabetically) of terms we used in the paper.

AHDH	Adaptive hierarchical density histogram
AR-Hist	Angular radial histogram feature
BSK	Binary salient keypoints
CBIR	Content-based image retrieval
CLP	Contextual local primitives
CT	Curvature tree
DIRA	Diagram image retrieval and analysis
Edist	Euclidean distance
EPO	European Patent Office
HOGH	hierarchical oriented gradient histogram
LBP	Local binary pattern
OCR	Optical character recognition
SC	Shape context
SG	Skeleton graph
SVM	Support vector machine
TGGLines	Topological graph guided lines
TAR	Triangle-area representation
RBSD	Region-based shape descriptor
RI-AHDH	Rotation invariant-AHDH
RI-LBP	Rotation-invariant LBP
USPTO	United States Patent and Trademark Office

B. Additional illustration figures for Section 2

In this appendix, we provide some additional illustration figures to help our readers understand the reviewed methods in Section 2 that cannot fit in the main paper due to space limitation.

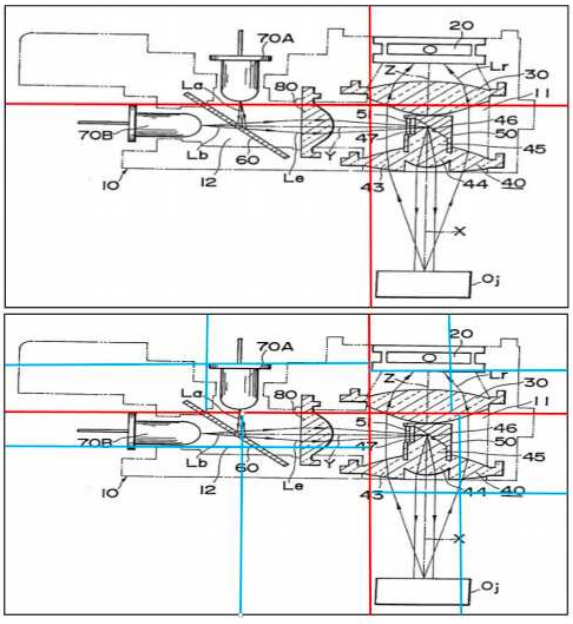


Figure 5. A patent image and its first (in red line) and second division (in blue line) based on its hierarchical geometry centroid, from HOGH – an improved AHDH method (Figure from [29]).

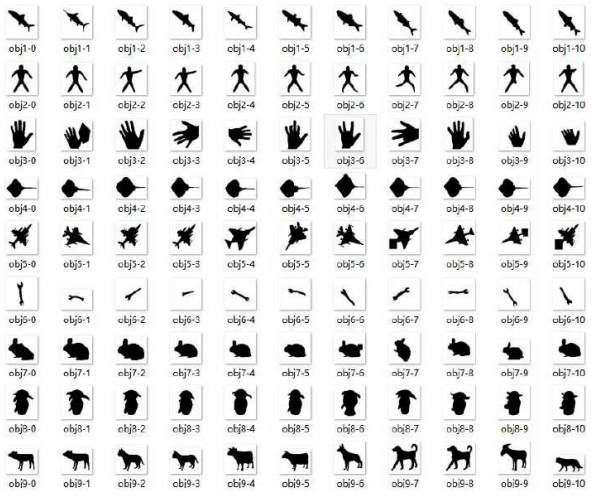
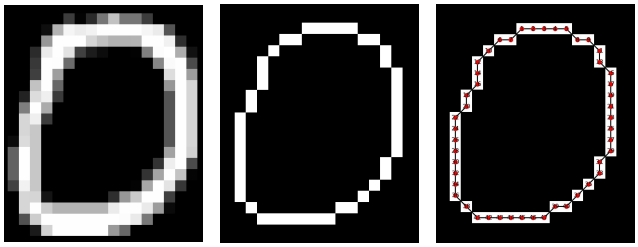


Figure 6. The Kimia-99 shape database for binary image retrieval (Figure from [22]).

(a) Camel (b) Cattle

Figure 7. MPEG data set examples [38] .



(a) Input image (b) Image skeleton (c) Skeleton graph

Figure 8. An example of skeleton graph image representation . Figure 8 (a) is the input image. Figure 8 (b) shows the image skeleton extracted from the input image. Figure 8 (c) provides the skeleton graph corresponding to the skeleton present in (b). In the skeleton graph, each node represents a pixel in the image skeleton, and each edge indicates that the two pixels it connects are neighbors. (Figure from [20]. The handwritten digit image used in the figure is taken from the MNIST data set[25].)

C. Other methods for Section 2.5

Local features and descriptors that perform well for natural images are often unable to capture the content of binary technical drawings. A local feature representation, called *contextual local primitives* (CLP), is proposed in [12]. CLP is based on detecting junction and end points in diagram images, then classifying the local primitives to local primitive words, and finally establishing the geodesic connections of the local primitives. The authors of CLP exploited the granulometric information of the binary patent images they ran experiments on, in order to set all the necessary parameters of the involved mathematical morphology operators and window size for the local primitive extraction. This makes the whole framework parameter free. However, this also indicates the CLP will not work well for other technical drawings beyond binary patent images. CLP is scale invariant and to a certain extent rotation invariant, but lacks affine invariance.

United States Patent and Trademark Office (USPTO) hosted a-month-long online competition in which participants developed algorithms to detect figures and diagram part labels. Competition-based graphics recognition algorithms for detecting figures and part labels that are commonly appear in patent images (e.g., "FIG. 2"), are provided in [39]. This is very important to improve research for DIRA, as those labels pose a big challenge for most of existing methods (see Section 3.1).

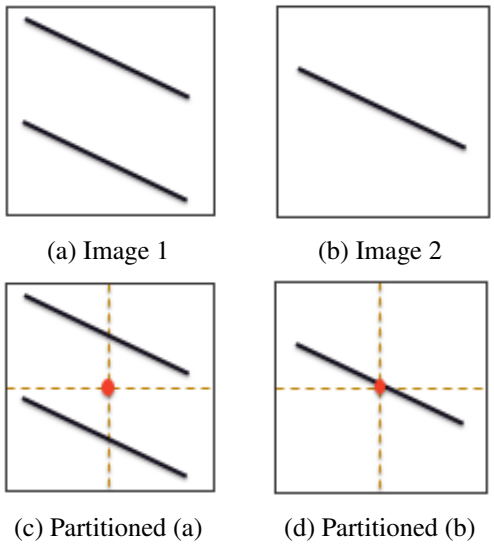
D. Additional technical details

In this appendix, we provide some additional technical details beyond the main body page limit to help our readers understand well the challenges for the methods we have experimented for DIRA (i.e., those demonstrated in Section 3.1).

D.1. AHDH technical details

AHDH uses the visual feature vectors that consider the geometry and the pixel distribution of patent diagram images[42]. AHDH requires pre-processing (e.g., noise reduction, coordinate calculation, normalization). Next, the adaptive geometric centroid is computed from the pre-processed images. The image is then partitioned into four sub-regions based on the whole image centroid. For each partitioned sub-region, extracted density features, including density, relative density and quantized relative density, is concatenated as a feature vector.

Figure 9 illustrates how a 1-level partition of AHDH works to distinguish images (a) and (b) provided in the top row. Images (c) and (d) are the partitioned results corresponding to (a) and (b) respectively. We can see the two images share the same centroid location for the 1-level partition. The sub-regions are marked as $sub_1, sub_2, sub_3, sub_4$ in clockwise starting from the top-left sub-region.



(a) Image 1 (b) Image 2

(c) Partitioned (a) (d) Partitioned (b)

Figure 9. An illustration of AHDH (1-level partition and geometric centroid calculation.)

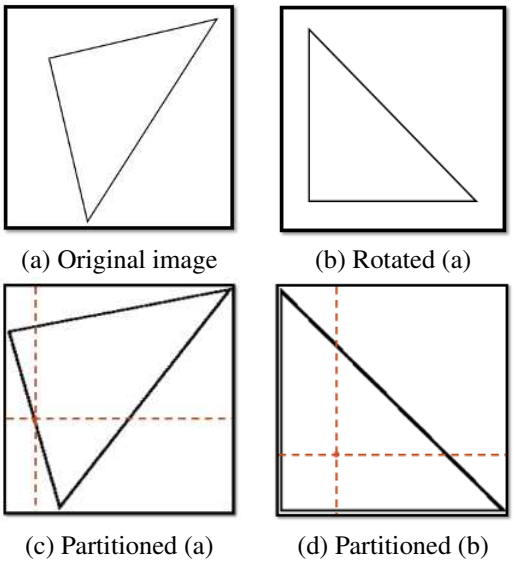
Density: The distribution of the black pixels for each sub-region. For example, density for Figure 9 (a) is $0.25, 0.25, 0.25, 0.25$ and $0.5, 0, 0.5, 0$ for Figure 9 (b).

Relative density: The ratio of the density of each sub-region to the percentage of that sub-region's area. For example, the relative density feature for Figure 9 (a) is $1, 1, 1, 1$ (note that for each sub-region, 0.25/0.25 = 1) and $2, 0, 2, 0$ for Figure 9 (b), note that 0.5/0.25 = 2 for sub_1, and 0/0.25 = 0 for sub_2.

Quantized relative density: It is treated as a higher-level binary classifier. The two class names are"F" and "E", depending on the relative density. A sub-region is classified as "E" if the relative density is < 1, otherwise, the sub-region is marked as "F". For example, the quantized relative density feature vector for Figure 9 (a) is $FFFF$ and $FEFE$ for Figure 9 (b). The concatenation of quantized relative density is then converted to a decimal value ranging from 0 to 15.

The final concatenated density feature vector is $[0.25, 0.25, 0.25, 0.25, 1, 1, 1, 1, 15]$ for Figure 9 (a) and $[0.5, 0, 0.5, 0, 2, 0, 2, 0, 10]$ for Figure 9 (b), which can be fed into a machine learning classier such as support vector machine (SVM). Edist measure (for 1-level AHDH feature vector) was used to calculate the similarity between the two images. Edist for the two images in the top row of Figure 9 is 5.4083.

AHDH is not rotation invariant, as extracted feature vectors of one image and its rotated version can



(a) Original image (b) Rotated (a)

(c) Partitioned (a) (d) Partitioned (b)

Figure 10. An illustration of AHDH not rotation invariant. We can see the location of the geometric centroid for the original image and for its rotated varies significantly.

vary substantially (see Figure 10 for an illustration). The centroid location shifts when an image is rotated, and this results in the first level partition will be completely different. In Figure 10, for 1-level AHDH, the extracted feature vector of Figure 10 (a) is $[0.3755, 0, 0.249, 0.3755, 2.0131, 0, 3.9231, 1.9718, 13]$ and $[0.5, 0, 0, 0.5, 7.0002, 0, 0, 1.387, 9]$ for Figure 10 (b). Edist for the top row image pair in Figure 10 is 7.5297. Also, AHDH cannot be used for partial image matching. For example, the image shown in Figure 9 (b) is a sub-image of the image in Figure 9 (a), but AHDH cannot tell this.

D.2. LBP and RI-LBP technical details

LBP considers intensity value of each pixel and its 8 neighbors' (window size: 3*3). Based on the intensity difference, a binary value (0 or 1) is re-assigned to each of the surrounding pixels. Taking current pixel p as an example, let v_p be intensity value of p, n be one of p's surrounding neighbors with intensity value as v_n, if $v_n > v_p$, v_n is re-assigned to 1, otherwise v_n is re-assigned to 0. After re-assigning the values of p's 8 neighbors, the re-assigned intensity values of p's neighbors are concatenated, and then converted to a decimal value v'_p, where v'_p ranges from 0 to 255. v_p is updated to v'_p. Note that the concatenation order does not matter, as long as it keeps consistent for all pixel value calculation.

However, LBP is not rotation invariant. RI-LBP was proposed in [26]. Uniform patterns are used to reduce the LBP dimension from 256 to 59. A uniform is a local binary pattern that contains at most two 0 to 1 or 1 to 0 transitions.

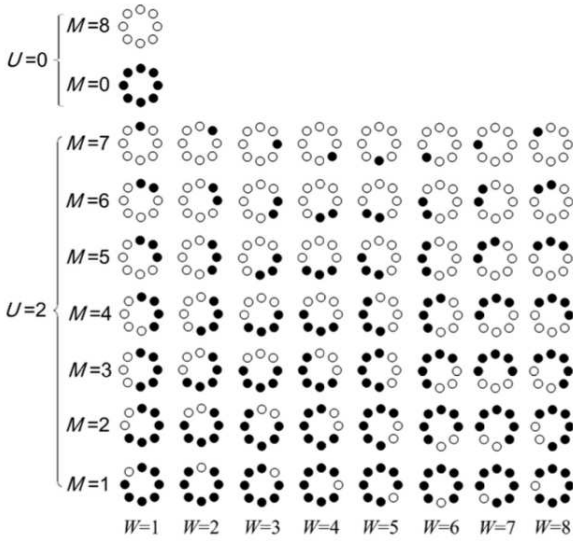


Figure 11. Fifty-eight defined sub-uniform patterns. Bit values of 0 (black circle) and 1 (white circle) in the output of the operator (from [26]). Labeled positions of a pixel's eight neighbors are in clockwise order and the top-left neighbor is labeled as 1.

Each uniform pattern has different sub-uniform patterns. Figure 11 provides all the possible sub-uniform patterns, where W denotes the position of 1 to 0 transition, M denotes the number of neighboring pixels with value as 1, and U represents uniform pattern. After the sub-uniform pattern of each pixel is determined, a histogram of the sub-uniform patterns is generated. In order to be rotation invariant, for each of the uniform patterns (M = 1, 2, .., 7), the sub-uniform pattern with the maximum statistical value from the histogram (i.e., the dominant-orientation sub-uniform) is moved to the first column, and the other bins are circularly shifted, after which 9 histograms of sub-uniforms are concatenated (in the order of $M1, M2, ...M7, M8, M0$) to generate RI-LBP features.

D.3. RBSD technical details

The main difference between patent diagram images and MPEG shapes is that patent images consists of complex line drawings and MPEG shapes has more details than lines. The Figure 12 shows an example of binary mask operation on the image for AR-Hist feature extraction, where Result 1 to Result 3 correspond to the operation with binary mask (r_1, θ_1) to (r_3, θ_3). AR-Hist features doesn't work for patent image because one type patent images (such as shoes in this case) from the same angle share similar shapes. The AR-Hist is similar for 2 different types of shoes as shown in Figure 12 (b) and (c).

(a) Examples of binary masks

(b) A binary mask operation example for Figure 2 (a)

(c) A binary mask operation example for Figure 2 (d)

(d) A binary mask operation example for Figure 7 (b)

Figure 12. Illustrations of binary mask Operation for patent diagram images and shape images from MPEG-7 data set.

E. DIRA related data sets

To advance research in DIRA, below we list some existing data sets that are close to diagram images:

- *Canadian industrial designs database* [4]: the database contains industrial design diagrams, the database is updated daily.
- *CLEF-IP 2011 collection* [35, 18]: a data set created as a test collection for four tasks: prior art search, patent classification, image-based prior art search, and image classification [40], where the last two tasks are relevant to diagram images.
- *Two data sets from the Multimedia Knowledge and social media analytics laboratory (MKLab)*: one contains 2000 binary images from 2000 patent images extracted from patent documents provided by the European Patent Office (EPO)[2]; and the other contains 1042 patent images (with different image size) extracted from around 300 patents from EPO and USPTO. This data set was manually annotated with 8 concepts of different types of shoes' designs [3].
- *Line Drawings of 3D Shapes* [15, 16, 1]: the data set contains the initial and registered drawings from artists.
- *Fashion-MNIST* [47]: a data set contains 28x28 grayscale images of 70,000 fashion products from 10 categories (7,000 images per category). The training set has 60,000 images and the test set 10,000 images.