

Chemical Symbol Feature Set for Handwritten Chemical Symbol Recognition

Peng Tang, Siu Cheung Hui, and Chi-Wing Fu

School of Computer Engineering,
Nanyang Technological University, Nanyang Avenue, Singapore 639798
{ptang1, asschui, cwfu}@ntu.edu.sg

Abstract. There are two main types of approaches for handwritten chemical symbol recognition: image-based approaches and trajectory-based approaches. The current image-based approaches consider mainly the geometrical and statistical information from the captured images of users' handwritten strokes, while the current trajectory-based recognition approaches only extract temporal symbol features on users' writing styles. To recognize chemical symbols accurately, however, it is important to identify an effective set of important chemical features by considering the writer dependent features, writer independent features as well as context environment features. In this paper, we propose a novel CF44 chemical feature set based on the trajectory-based recognition approach. The performance of the proposed chemical features is also evaluated with promising results using a chemical formula recognition system.

1 Introduction

There are two main types of approaches for handwritten chemical symbol recognition, namely *image-based approaches* and *trajectory-based approaches*. The image-based approaches aim at recognizing chemical symbols based on the image input of the pen strokes. The trajectory-based recognition approaches recognize chemical symbols when they are written with a pen-based device, where a sensor picks up the pen-tip movements as well as pen-up/pen-down switching. The actual pen trajectory data is known as digital ink and it is captured for the recognition process.

For the past decade, different techniques such as Support Vector Machines (SVM) [3], Hidden Markov Models (HMM) [8], hybrid SVM-HMM [7] and Support Vector Machine-Elastic Matching (SVM-EM) [6] have been proposed for chemical symbol recognition. The current image-based approaches [3,5,4] mainly considered the geometrical and statistical information from the captured images of users' handwriting strokes. As such, users' writing styles during the writing process are not captured. On the other hand, the current trajectory-based recognition approaches [8,7] only considered extracting temporal symbol features based on stroke points. These techniques only focused on capturing features about the writing process which are dependent on users' writing styles.

To recognize chemical symbols accurately, it is important to identify an effective set of important chemical features from users' handwritten input data. In this paper, we propose a novel CF44 chemical feature set which consists of writer dependent features, writer independent features and context environment features. We limit the feature set to a total of 44 important features in order not to over influence the processing speed of the feature extraction process. The rest of the paper is organized as follows. Section 2 reviews the related work. Section 3 presents the proposed chemical symbol features. Experimental results are then presented in Section 4. Finally, Section 5 concludes the paper.

2 Related Work

Different kinds of features have been proposed based on the image-based symbol recognition approaches. In [5], geometrical features based on lines and polygons were extracted and recognized with text recognition. In [3], Ouyang and Davis used both the geometrical and statistical features including number of strokes, bounding-box dimensions, stroke ink density, inter-stroke distance and inter-stroke orientation for SVM symbol recognition. In [4], Ouyang and Davis further proposed to extract additional geometrical features from the stroke segments including segment length, segment count, stroke diagonal, symbol diagonal and symbol ink density for symbol recognition using the Conditional Random Field (CRF). However, the current image-based approaches can only capture the visual features about the handwritten chemical symbols.

For trajectory-based symbol recognition approaches, different types of chemical features have also been extracted. In [8,7], Zhang et al. proposed 11 types of chemical features including normalized stroke point coordinates, normalized first derivatives and second derivatives of the stroke points, curvature, writing direction, aspect, curliness and linearity for symbol recognition in HMM and SVM-HMM. In [6], Tang et al. used some simple chemical features including number of strokes, stroke point coordinates, horizontal angle and turning angle for the SVM-EM symbol recognition. As the chemical features extracted by the current trajectory-based approaches are quite limited, in this paper we propose an effective set of chemical features based on the trajectory-based approach for chemical symbol recognition.

3 Chemical Symbol Features

Chemical symbols consist of digits, alphabetic characters, bonds, and operators. For the purpose of handwritten chemical symbol and formula recognition, we propose a set of 44 chemical symbol features to capture the various aspects of handwritten chemical symbols. They are classified into 11 writer dependent features, 31 writer independent features and 2 context environment features.

First, we define the following fundamental concepts on *stroke*, *chemical symbol pattern* and *bounding box*:

Definition 1 (Stroke). A stroke s is a sequence of m two-dimensional points (p_1, p_2, \dots, p_m) :

$$s = (p_1, p_2, \dots, p_m),$$

where $p_i = (x_i, y_i)$, $1 \leq i \leq m$, p_1 is the pen-down point, p_m is the pen-up point and p_2, \dots, p_{m-1} are the pen-move points.

Definition 2 (Chemical Symbol Pattern). A chemical symbol pattern S is a valid sequence of k strokes which are recognized together as a chemical symbol. Therefore, it is also a sequence of n two-dimensional points (p_1, p_2, \dots, p_n) :

$$S = (s_1, s_2, \dots, s_k) = (p_1, p_2, \dots, p_n),$$

p_1 and p_n are the starting point and ending point of the chemical symbol pattern.

Definition 3 (Bounding Box). The bounding box $B = (w, h)$ of a symbol S is the smallest rectangle which encloses S . The edges of the rectangle are in parallel with the coordinate axes. w and h are the width and height of the bounding box. And the center of the bounding box is defined as the center of S .

3.1 Writer Dependent Features

The writer dependent features are dynamic features including the order of points, writing direction, and the number and order of strokes which model the writing process of handwritten chemical symbols.

Pattern Starting and Ending Point Features. The positions of the starting point and ending point of the chemical symbol pattern are important features for the recognition of chemical symbols since users often write the same symbol with similar positions on its starting point and ending point. However, if a handwritten chemical symbol has very small width or height such as chemical bond symbols, we will extract the pattern starting point and ending point features by standardizing them with respect to a bounding square. The bounding square is defined as follows:

Definition 4 (Bounding Square). The bounding square of a symbol pattern S is the smallest virtual square which encloses S . The center of the bounding square $c = (x_c, y_c)$ is located at the center of the chemical symbol pattern. The edges of the bounding square are in parallel with the coordinate axes. The length l of the bounding square is the larger of the width w and height h of the bounding box.

For a chemical symbol pattern $S = (p_1, \dots, p_n)$ with starting point $p_1 = (x_1, y_1)$ and ending point $p_n = (x_n, y_n)$, the pattern starting point features (f_1 and f_2) and ending point features (f_3 and f_4) are formulated as follows:

$$\text{Starting Point: } f_1 = \frac{x_1 - x_c}{l} + \frac{1}{2}, \quad f_2 = \frac{y_1 - y_c}{l} + \frac{1}{2}, \quad (1)$$

$$\text{Ending Point: } f_3 = \frac{x_n - x_c}{l} + \frac{1}{2}, \quad f_4 = \frac{y_n - y_c}{l} + \frac{1}{2}. \quad (2)$$

where l is the length of the bounding square.

Pattern Writing Direction Features. The pattern writing direction features model users' general writing direction when users write a chemical symbol.

Definition 5 (Pattern Writing Direction). For a chemical symbol pattern $S = (p_1, p_2, \dots, p_n)$, the pattern writing direction of S is defined as a vector from the starting point p_1 to the ending point p_n .

The writing direction features are important since users often write the same symbol with the same direction. The writing length feature (f_5) and direction features (f_6 and f_7) are formulated as follows:

$$\text{Writing Length: } f_5 = \|\mathbf{v}\| = \|p_1 p_n\|, \quad (3)$$

$$\text{Writing Direction: } f_6 = \frac{\mathbf{v}_x \cdot \mathbf{u}_x}{\|\mathbf{v}\|}, \quad f_7 = \frac{\mathbf{v}_y \cdot \mathbf{u}_y}{\|\mathbf{v}\|}, \quad (4)$$

where \mathbf{v} is the starting point to ending point vector $\overrightarrow{p_1 p_n}$ which models the writing direction. \mathbf{u}_x and \mathbf{u}_y denote the unit vector of the horizontal and vertical axes respectively.

The two direction features f_6 and f_7 become unstable when the writing length feature f_5 is too small. To tackle this problem, the values of the two direction features will be set to zero when the writing length feature is less than a minimal distance threshold which is set empirically to $\max(w, h)/4$.

Pattern Closure Feature. It is used to distinguish the difference between chemical symbols with closed or circular writing trajectory pattern (such as 'O') and those with straight line writing trajectory pattern (such as 'I' and chemical bond symbols). This feature is defined based on the writing length feature f_5 .

Definition 6 (Pattern Closure). The pattern closure of a symbol pattern is defined as the ratio between the writing length and the trajectory length of the handwritten chemical symbol pattern.

For a chemical symbol pattern $S = (p_1, p_2, \dots, p_n)$, the pattern closure feature (f_8) is formulated as follows:

$$\text{Closure: } f_8 = \frac{\|\mathbf{v}\|}{L}, \quad (5)$$

where $L = \text{pathdist}(p_1, p_n) = \sum_{i=1}^k \sum_j \|\overrightarrow{p_j p_{j+1}}\|$, L is the path distance between the pattern starting point p_1 and ending point p_n . And it is calculated by summing up the length of each stroke segment $\overrightarrow{p_j p_{j+1}}$ in S . It indicates the length of the writing trajectory by the user.

Pattern Inflection Features. The pattern inflection features are used to distinguish the convex and concave curvature patterns shown in the handwritten chemical symbol.

Definition 7 (Pattern Inflection). The pattern inflection of a chemical symbol pattern S is a measure of the relative positioning of the middle-path point p_{mid} of S with respect to the middle point p_m of the straight line $\overrightarrow{p_1 p_n}$.

The pattern inflection features (f_9 and f_{10}) are formulated as follows:

$$X \text{ and } Y\text{-Inflection: } f_9 = \frac{1}{w}(x_{mid} - x_m), \quad f_{10} = \frac{1}{h}(y_{mid} - y_m), \quad (6)$$

where $p_m = (x_m, y_m) = (\frac{x_1+x_n}{2}, \frac{y_1+y_n}{2}) = \frac{p_1+p_n}{2}$.

The pattern inflection features f_9 and f_{10} are normalized by the width w and height h of the bounding box of the chemical symbol pattern respectively.

Total Stroke Number Feature. It is an important feature to indicate the writing complexity of the handwritten chemical symbol. The total stroke number feature (f_{11}) accounts for the number of strokes k in the handwritten chemical symbol pattern as follows:

$$\text{Total Stroke Number: } f_{11} = k. \quad (7)$$

3.2 Writer Independent Features

The writer independent features are visual features which model the appearance of the handwritten chemical symbols. These features will not be affected by changes in users' stroke order or writing direction.

Bounding Box Features. The bounding box features model the basic information about the shape of the handwritten chemical symbol. The diagonal angle feature (f_{13}) and diagonal ratio feature (f_{14}) are formulated as follows:

$$\text{Diagonal Angle and Ratio: } f_{13} = \arctan \frac{h}{w}, \quad f_{14} = \frac{h+w}{L}. \quad (8)$$

The diagonal angle feature measures the angle between the bounding box diagonals with respect to the horizontal axis. And the diagonal ratio is a relative measure of the bounding box size which characterizes the complexity when writing the chemical symbol.

Deviation Feature. The deviation feature is used to measure the dispersion of stroke points in the handwritten chemical symbol. For a chemical symbol pattern $S = (p_1, p_2, \dots, p_n)$, the deviation feature (f_{15}) is formulated as follows:

$$\text{Pattern Deviation: } f_{15} = \frac{1}{n} \sum_{i=1}^n \|p_i \mu\|, \quad (9)$$

where $\mu = \sum_{i=1}^n p_i/n$, and p_i is the i -th stroke point in the handwritten chemical symbol. μ is the center of gravity which is calculated as the mean of all the stroke points in the chemical symbol pattern. And $\|p_i \mu\|$ is the Euclidean distance between the i -th stroke point p_i and the center of gravity μ .

Average Direction Feature. The average direction feature is computed by averaging the writing directions of all the stroke segments defined in the trajectory of the handwritten chemical symbol. For a chemical symbol pattern $S = (p_1, p_2, \dots, p_n)$, the average direction feature (f_{16}) is formulated as follows:

$$\text{Average Direction:} \quad f_{16} = \frac{1}{n} \sum_{i=1}^{n-1} \arctan\left(\frac{y_{i+1} - y_i}{x_{i+1} - x_i}\right). \quad (10)$$

Curvature Feature. The curvature feature measures the curvature of the handwritten chemical symbol. Here, we introduce *stroke turning angle* which is used to measure the local curvature for two consecutive stroke segments.

Definition 8 (Stroke Turning Angle). *The stroke turning angle is the turning angle between two consecutive stroke segments within the same stroke. For a chemical symbol pattern $S = (p_1, p_2, \dots, p_n)$, the stroke turning angle between the stroke segments $\overrightarrow{p_{i-1}p_i}$ and $\overrightarrow{p_i p_{i+1}}$ is defined as follows:*

$$\theta_i = \arccos\left(\frac{\overrightarrow{p_{i-1}p_i} \cdot \overrightarrow{p_i p_{i+1}}}{\|\overrightarrow{p_{i-1}p_i}\| \|\overrightarrow{p_i p_{i+1}}\|}\right).$$

An example on stroke turning angle is shown in Figure 1a. The curvature feature (f_{17}) is then formulated by summing up all the local curvatures in S :

$$\text{Pattern Curvature:} \quad f_{17} = \sum_{i=2}^{n-1} \theta_i. \quad (11)$$

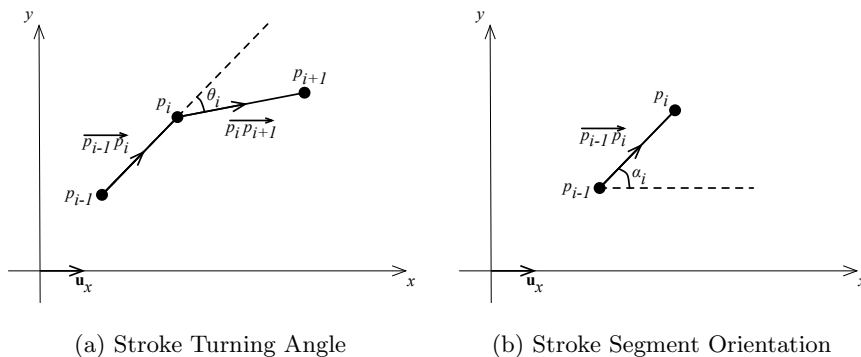


Fig. 1. Stroke Turning Angle and Stroke Segment Orientation

Therefore, symbols which consist of straight lines will have low curvature, whereas symbols which consist of curved strokes will have high curvature. For example, the symbol ‘H’ will have lower curvature measure than the symbol ‘O’.

Perpendicularity Features. The perpendicularity features are important features to identify the abrupt changes in the stroke trajectory, especially for symbols with sharp turning strokes such as ‘A’ and ‘N’. For a chemical symbol pattern $S = (p_1, p_2, \dots, p_n)$, the perpendicularity features (f_{18} and f_{19}) are formulated as follows:

$$\text{Perpendicularity:} \quad f_{18} = \sum_{i=2}^{n-1} \sin^2 \theta_i, \quad (12)$$

$$2\text{-th Perpendicularity:} \quad f_{19} = \sum_{i=2}^{n-1} \sin^2(\theta_i^2), \quad (13)$$

where 2nd-turning angle $\theta_i^2 = \arccos \left\{ \frac{\overrightarrow{p_{i-2}p_i} \cdot \overrightarrow{p_i p_{i+2}}}{\|\overrightarrow{p_{i-2}p_i}\| \|\overrightarrow{p_i p_{i+2}}\|} \right\}$.

Directional Histogram Features. The directional histogram features help to capture information about writing direction of each stroke segment of the chemical symbol pattern. First, we define *stroke segment orientation* as follows:

Definition 9 (Stroke Segment Orientation). *The stroke segment orientation is a measure for the writing direction of that stroke segment. For a chemical symbol pattern $S = (p_1, p_2, \dots, p_n)$, the stroke segment orientation for a stroke segment $\overrightarrow{p_i p_{i+1}}$ is calculated as:*

$$\alpha_i = \arccos \left\{ \frac{\overrightarrow{p_i p_{i+1}} \cdot \mathbf{u}_x}{\|\overrightarrow{p_i p_{i+1}}\|} \right\}. \quad (14)$$

Figure 1b illustrates the definition of stroke segment orientation.

Since a chemical bond symbol can have 12 possible directions in chemical formulas, the stroke segment range is divided into 12 directional histogram bins ($sh_1 - sh_{12}$) and each directional histogram bin indicates one of the 12 possible directions for stroke segment orientations. The range between two consecutive directional histogram bins is 30° . A fuzzy quantization approach is used to quantize the stroke segment orientation into histogram bins. A stroke segment orientation α_i contributes to two histogram bins with two different weights:

$$w_{(m,i)} = 1 - \frac{6}{\pi} \cdot \arccos\{\mathbf{u}_{\alpha_i} \cdot \mathbf{v}_m\}, \quad w_{(m+1,i)} = \frac{6}{\pi} \cdot \arccos\{\mathbf{u}_{\alpha_i} \cdot \mathbf{v}_{m+1}\}, \quad (15)$$

where \mathbf{u}_{α_i} is the unit vector which has the same orientation as stroke segment orientation α_i and \mathbf{v}_m is the unit vector for the orientation of directional histogram bin sh_m . Each histogram bin is then calculated as the sum of weighted contributions from each stroke segment:

$$sh_m = \sum_{i=1}^{n-k} w_{(m,i)}. \quad (16)$$

We define 6 directional histogram features based on the stroke segment orientation by inspecting the number of stroke segments oriented in the 12 directional histogram bins. The 6 directional histogram features (f_{20} to f_{25}) for the stroke segment orientation are formulated as follows:

$$\text{Stroke Directional Histogram: } f_{20} = \frac{sh_1 + sh_7}{n - k}, \dots, f_{25} = \frac{sh_6 + sh_{12}}{n - k}. \quad (17)$$

Similarly, we define 6 directional histogram features (f_{26} to f_{31}) for the stroke turning angle, which provides additional curvature information about the writing trajectory of the symbol pattern, as follows:

$$\text{Turning Directional Histogram: } f_{26} = \frac{th_1 + th_7}{n - 2 \times k}, \dots, f_{31} = \frac{th_6 + th_{12}}{n - 2 \times k}. \quad (18)$$

where th_1 to th_{12} are the 12 directional histogram bins for stroke turning angle.

2-Dimensional (2D) Histogram Features. The 2D histogram features measure which area of the symbol pattern is denser or has more stroke points. To extract the features, the bounding box of a symbol is divided into 3×3 2D histogram bins. Figure 2 shows the 3×3 2D histogram bins for a handwritten chemical symbol ‘C’. Then, each stroke point will be quantized into 4 nearest histogram bins with the respective weightings. The weighted contributions to the 4 histogram bins are determined by the distance from the stroke point to the center of the 4 respective histogram bins.

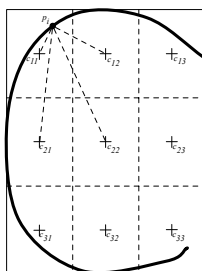


Fig. 2. 2D Histogram Bins

Nine 2D histogram features are extracted and each feature is calculated by inspecting all the contributions to each 2D histogram bin from the stroke points in the handwritten chemical symbol pattern. The 2D histogram features (f_{32} to f_{40}) are formulated as follows:

$$\text{2D Histogram: } f_{32} = \frac{h_{11}}{n} = \frac{1}{n} \sum_{i=1}^n w_{11}(p_i), \dots, f_{40} = \frac{h_{33}}{n} = \frac{1}{n} \sum_{i=1}^n w_{33}(p_i) \quad (19)$$

Convex Hull Features. The convex hull of the chemical symbol pattern is a set of points which can be used to construct the smallest convex shape to enclose the chemical symbol pattern. Therefore, the convex hull features are a perfect measure to model the geometric information of the handwritten chemical symbol pattern. In this research, the convex hull of the chemical symbol pattern is computed using the Graham algorithm [2]. The two convex hull features (f_{40} and f_{41}) are formulated as follows:

$$\text{Convex Hull Box Ratio: } f_{41} = \frac{A_H}{w \times h}, \quad (20)$$

$$\text{Convex Hull Trajectory Ratio: } f_{42} = \frac{L^2}{A_H}, \quad (21)$$

where the convex hull H is a sequence of points $(v_1, \dots, v_i, \dots, v_m)$ with $v_i = (v_{i,x}, v_{i,y})$ and the convex hull area $A_H = \frac{1}{2} \left| \sum_{i=1}^{p-1} (v_{i,x} \times v_{i+1,y} - v_{i+1,x} \times v_{i,y}) \right|$.

3.3 Context Environment Features

The context environment features capture the related context information on the handwritten chemical symbols when users write the chemical formulas. The context environment features are only possible after the first symbol of a handwritten chemical formula is recognized. The context environment features are extremely useful to distinguish symbols with the same shape and writing direction, for example, capital letters and its non-capital. There are two context environment features: previous symbol type (f_{43}) and relative symbol height (f_{44}). They are formulated as follows:

$$\text{Previous Symbol Type: } f_{43} = \begin{cases} 0 & \text{if } S_P \text{ is a digit} \\ 1 & \text{if } S_P \text{ is a character} \\ 2 & \text{if } S_P \text{ is an operator} \\ 3 & \text{if } S_P \text{ is a bond} \end{cases}, \quad (22)$$

$$\text{Relative Symbol Height: } f_{44} = \frac{h}{h_P}, \quad (23)$$

where S_P is the previous recognized symbol of the current unknown symbol in the chemical formula. h_P is the height of the previous symbol. f_{43} models the symbol type of the previous symbol and f_{44} captures the relative height information between the previous symbol and the current unknown symbol.

4 Performance Evaluation

In this research, we have developed a handwritten chemical formula recognition system on the iOS platform, called iDrawChem. In the system, LibSVM [1] is used for symbol recognition with the proposed chemical features. The performance evaluation was conducted as follows. First, we introduced the iDrawChem system to 10 users. Then, the users spent about 10 minutes to familiarize themselves with the system and experimented a few simple expressions which were different from those used for testing. We prepared 468 valid simple chemical expressions and the previous symbol of the chemical symbol was also contained in each chemical expression. In the experiment, each user was required to write all the chemical expressions. Then, the writer dependent features, writer independent features and context environment features of the chemical symbols were recorded. As a result, a total of 4680 valid symbol expressions were collected

Table 1. Performance Results based on Symbol Categories

Category	Precision@1	Precision@3	Precision@5
Digits	95.67%	98.33%	99.33%
Alphabetical Characters	92.56%	96.22%	98.85%
Bonds & Operators	91.9%	97.50%	98.96 %
Average	93.80%	96.75%	98.93 %

Table 2. Performance Results based on Feature Types

Features	Precision@1	Precision@3	Precision@5
WD	69.49%	75.56%	83.50%
WI	86.97%	90.47%	94.15%
WD+WI	92.82%	95.13%	97.61%
WD+WI+CE	93.80%	96.75%	98.93%

from the 10 users. Among them, 2340 of the symbol expressions were used for training the SVM classifier and the rest were used for testing.

Table 1 shows the performance results based on different symbol categories. As our symbol recognition is able to return a ranked list of candidate symbols for user selection, we use precision@1, precision@3 and precision@5 for performance evaluation. Precision@ n reports the recognition accuracy that the correct symbol is in the top- n candidate symbols. As shown in the table, the performances at precision@1, precision@3 and precision@5 are 93.80%, 96.75% and 98.93% respectively which are quite promising. The performance of digits is better than that of alphabetical characters, bonds and operators, as digits generally have more distinguishing features.

Table 2 gives the performance results based on different types of chemical features: Writer Dependent (WD) features, Writer Independent (WI) features, Context Environment (CE) features. As shown in the table, the performance with WI features (86.97% at precision@1) is much higher than that with WD features (69.49% at precision@1), as writer independent features which capture the visual appearance of the symbols may help recognize the symbols more effectively than writer dependent features. In addition, using WD features and WI features together, the performance has improved to 92.82% at precision@1. Furthermore, the CE features can further improve the performance to 93.80% at precision@1 since they can help distinguish symbols with the same visual shape and writing direction.

5 Conclusion

In this paper, we have proposed a set of 44 chemical symbol features which can effectively model the user writing process, visual appearance and context environment of each handwritten chemical symbol. The proposed CF44 chemical symbol features are designed by considering both writer dependent and writer independent features as well as the context environment features. The proposed features have been evaluated using our chemical formula recognition system called iDrawChem. The performance evaluation has shown promising results with high accuracy up to 98.93% at precision@5. Therefore, the proposed chemical features are effective to be used for the recognition of handwritten chemical symbols for chemical formula recognition.

References

1. Chang, C., Lin, C.: Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2(3), 27 (2011)
2. Graham, R.L.: An efficient algorithm for determining the convex hull of a finite planar set. *Information Processing Letters* 1(4), 132–133 (1972)
3. Ouyang, T., Davis, R.: Recognition of hand drawn chemical diagrams. In: *Proceedings of the National Conference on Artificial Intelligence*, vol. 22, p. 846. AAAI Press; MIT Press, Menlo Park; Cambridge (2007)
4. Ouyang, T., Davis, R.: Chemink: A natural real-time recognition system for chemical drawings. In: *Proceedings of the 16th International Conference on Intelligent User Interfaces*, pp. 267–276. ACM (2011)
5. Ramel, J., Boissier, G., Emptoz, H.: Automatic reading of handwritten chemical formulas from a structural representation of the image. In: *Proceedings of the Fifth International Conference on Document Analysis and Recognition*, pp. 83–86. IEEE (1999)
6. Tang, P., Hui, S., Fu, C.: Online chemical symbol recognition for handwritten chemical expression recognition. In: *Proceedings of the 12th International Conference on Computer and Information Science* (2013)
7. Zhang, Y., Shi, G., Wang, K.: A svm-hmm based online classifier for handwritten chemical symbols. In: *Proceeding of International Conference on Pattern Recognition 2010*, pp. 1888–1891. IEEE (2010)
8. Zhang, Y., Shi, G., Yang, J.: Hmm-based online recognition of handwritten chemical symbols. In: *Proceedings of 10th International Conference on Document Analysis and Recognition*. pp. 1255–1259. IEEE (2009)