

# Matching Expressions by using Structural Belief Propagation: First Results

Alexandre Noma  
Centro de Matemática, Computação e Cognição  
Universidade Federal do ABC  
São Paulo, Brazil  
Email: alexandre.noma@ufabc.edu.br

Frank D. Julca-Aguilar, Nina S. T. Hirata  
Instituto de Matemática e Estatística  
Universidade de São Paulo  
São Paulo, Brazil  
Email: {faguilar, nina}@ime.usp.br

**Abstract**—Ground-truthed datasets are fundamental for performance evaluation of handwritten mathematical expression recognition methods. In order to automate the construction of such datasets, some approaches consider transcription of model expressions followed by automatic assignment of symbols in the transcribed expression to the corresponding symbols in the model expression. In order to cope with observed difficult cases, we propose a structural approach based on belief propagation to match the corresponding symbols in the expressions. Preliminary results suggest that the proposed approach can outperform previous methods.

**Keywords**—Handwritten mathematical expression; ground-truth; graph matching; structural belief propagation

## I. INTRODUCTION

Given a task, ground-truthed datasets serve as common base to evaluate and compare performance of different methods for that task. In the problem of recognition of handwritten mathematical expressions (MEs), a main concern with the creation of such datasets is the effort needed to manually attach ground-truth at the different levels of the expressions – that is, grouping of strokes that represent a same symbol, labeling of individual symbols, and identification of relations between symbols – and the error prone nature of this task.

In [1], [2], the authors proposed an expression matching-based approach to automate much of the ground-truthing work. Given a model expression with ground-truth data and a transcription of it (called input expression) with no ground-truth data, symbols of the input expression are associated to the corresponding ones in the model. The resulting correspondences are used to automatically transfer all ground-truth information from the model to the input. Thus, expression datasets with ground-truth data can be created from a set of model expressions, making different people transcribe them, and using the proposed expression matching method. Two examples of expression matching, both to a same model expression, are illustrated in Fig. 1.

The average correct matching rate on the *ExpressMatch* dataset (comprising 56 types of expressions and 901 input expressions) are, according to [2], around 97%. However, it has been observed that due to handwritten variability and structural differences, the matching performance was very poor for some particular pairs of model-input expressions. For

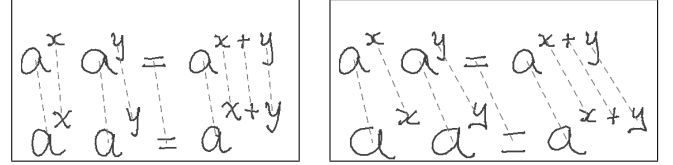


Fig. 1. Matching of input expressions (bottom ones) to the model (top one). Symbols correspondences are shown with dashed lines.

example, Fig. 2 shows several matching errors that are due to the horizontal numerator displacement: while the numerator of the expression at the bottom (input) is horizontally placed approximately at the center of the fraction line, in the model expression (top) the numerator is displaced to the left relative to the center of the fraction line.

$$A_{2K} = \frac{2RAK}{2R + \sqrt{4R^2 + A_K^2}}$$

$$A_{2K} = \frac{2RAK}{2R + \sqrt{4R^2 + A_K^2}}$$

Fig. 2. Matching errors due to numerator horizontal displacement. Lines indicate incorrect symbol correspondences.

We propose a structural matching approach based on belief propagation [3] to improve the correspondence between symbols of a pair of expressions. Preliminary results suggest that the proposed approach can outperform previous methods, specially under challenging scenarios as the ones described above. The key step is to find a robust structure capturing a global arrangement of symbols that is invariant among corresponding expressions. Currently, our research efforts are being dedicated to find such robust structures within expressions.

## II. BACKGROUND

The matching problem [1] mentioned above is modeled as a linear assignment problem. Expressions are modeled as

graphs and matching costs between symbols consider both local (symbol) as well as global (structural) features. The optimal one-to-one assignment is computed using the well-known Hungarian algorithm. This formulation can be viewed as a simple **graph matching** problem: each expression is represented by a graph, where each vertex corresponds to a symbol, and edges encode the spatial configuration between symbols.

Through experimentation on *ExpressMatch* dataset [2], some characteristics that affect the performance of the matching technique were observed and reported: (i) horizontal displacement of numerators or denominators in relation to a fraction line (such as the case shown in Fig. 2), (ii) symbols with irregularly proportional sizes (for example, the square root and fraction line of the expression at the top in Fig. 3), and (iii) non-uniform spacing, either decreasing or increasing, between symbols as the symbol positions advance toward the right extreme of the expression, or yet non-uniform spacing between sub-expressions (as in the bottom expression in Fig. 3). In this work, the proposed method is evaluated on these challenging cases.

$$V_h(r) \approx \frac{n^{-(n+D)/2}}{\sqrt{\pi}} (2\pi e)^{n/2} r^n$$

$$\frac{b^2 c^2 - 4 b^3 d - 4 a c^3 + 18 a b c d - 27 a^2 d^2}{a^4}$$

Fig. 3. Problematic expressions: symbols with irregularly proportional sizes (top), and irregular spacing between symbols (bottom).

#### A. Graph Matching

Popular algorithms for graph matching include approaches based on search [4], [5] and continuous optimization [6], [7], among others.

In some cases, the graph matching problem can be simplified. Recently, deformed graphs have been proposed for matching and interactive segmentation [4], [5]. Both cases involve a method based on a global cost evaluating the vertex attributes and structural information. The structural information, or spatial configuration between vertices, is used by deformed graphs to significantly reduce the search space. The matching method proposed in [1] and briefly described above is based on these ideas. Unfortunately, most of search-based methods are based on heuristic approaches, and generally do not consider the context of the neighborhood, in the sense that the matching of a pair of vertices can influence the matching of their neighbors.

Another branch of techniques involves continuous optimization to approximate the original graph matching problem by relaxing the integer constraint. An example is the graduated assignment algorithm, which was popularized by Chui and

Rangarajan [7]. They improved the original method by including a transformation estimation step (e.g. thin-plate splines) for the correspondence estimation. The idea is to alternate two steps, correspondence and transform estimation. The same technique can be combined with other local features, such as shape contexts [6]. Here, we combine shape contexts with spatial aspects in order to improve the correspondences.

While the expression matching method described in [1] calculates matching using deformed graphs, we propose an approach that is an extension of the efficient belief propagation algorithm [3], which takes into account the contextual information given by the neighbors. Here, the smoothness constraints are replaced by spatial constraints in order to evaluate the global arrangements of vertices. The goal is to minimize the ambiguities when comparing pairs of vertices.

### III. METHODOLOGY

To calculate the expression graphs, each symbol is represented by a vertex. Each vertex is associated to a  $(x, y)$  coordinate in the plane. Thus, edges are defined to encode the spatial configuration of symbols.

#### A. Technical background

The proposed approach is based on Markov random fields (MRFs) and follows previous works [8], [9].

Let  $G = (V, E, \mu, \nu)$  be an attributed graph,  $V$  be the set of vertices,  $E \subseteq V \times V$  be the set of edges,  $\mu$  be an attribute vector assigned to each vertex, and  $\nu$  be an attribute vector assigned to each edge. The goal is to match a pair of graphs, an input graph  $G_i$  representing an input expression, and a model graph  $G_m$  representing a model expression.

Given  $G_i = (V_i, E_i, \mu_i, \nu_i)$  and  $G_m = (V_m, E_m, \mu_m, \nu_m)$ , we define a MRF on  $G_i$ . For each input vertex  $p \in V_i$ , we must find a corresponding model vertex, by computing a suitable mapping (labeling)  $f : V_i \rightarrow V_m$  which minimizes Eq. 1. Let  $f_p \in V_m$  be the label of  $p \in V_i$ .

$$E(f) = \sum_{p \in V_i} D_p(f_p) + \lambda_1 \sum_{(p,q) \in E_i} M(f_p, f_q). \quad (1)$$

Each vertex has an attribute vector  $\mu_i(p)$  in  $G_i$  (similarly in  $G_m$ ). Each directed edge has an attribute vector  $\nu_i(p, q)$  in  $G_i$  (similarly in  $G_m$ ). The linear term  $D_p(f_p)$  assigns a cost proportional to the vertex attributes  $\mu_i(p)$  and  $\mu_m(f_p)$ . The quadratic term  $M(f_p, f_q)$  evaluates the structural information and assigns a cost proportional to the edge attributes  $\nu_i(p, q)$  and  $\nu_m(f_p, f_q)$ . Parameter  $\lambda_1$  weights the influence of the quadratic term.

1) *Belief propagation*: The max-product belief propagation [3] is applied to minimize Eq. 1, based on a message propagation strategy. For each iteration  $t$ , each vertex  $p$  sends a message vector to each neighbor  $q$ , with dimension defined by the number of labels.

$$m_{pq}^t(f_q) = \min_{f_p} \left( M(f_p, f_q) + D_p(f_p) + \sum_{s \in \mathcal{N}_p \setminus \{q\}} m_{sp}^{t-1}(f_p) \right) \quad (2)$$

where  $\mathcal{N}_p \setminus \{q\}$  denotes the neighbors of  $p$  except  $q$ . After  $T$  iterations, each vertex can compute its belief vector: each entry is based on vertex dissimilarity and contextual information given by the neighbors. The goal is to choose a label corresponding to the minimum entry in the belief vector.

$$b_q(f_q) = D_q(f_q) + \sum_{p \in \mathcal{N}_q} m_{pq}^t(f_q) \quad (3)$$

2) *Efficiency via min convolution*: Each message vector can be efficiently computed by rewriting Eq. 2 in the form of a min convolution [3], as shown in Eq. 4, where  $h(f_p) = D_p(f_p) + \sum m_{sp}^{t-1}(f_p)$ . For instance, in Eq. 5, we replicate the min convolution form for the Potts model, proposed by Felzenszwalb and Huttenlocher [3].

$$m_{pq}^t(f_q) = \min_{f_p} \left( M(f_p, f_q) + h(f_p) \right) \quad (4)$$

$$m_{pq}^t(f_q) = \min \left( h(f_q), \min_{f_p} h(f_p) + d \right) \quad (5)$$

In our case, we define  $H(f_q)$  to include the structural evaluation in the optimization process and rewrite Eq. 4 by replacing  $h(f_q)$  with  $H(f_q)$ , as shown in Eqs. 6 and 7, respectively.

$$H(f_q) = \min_{f_p \in \mathcal{N}_{f_q} \cup \{f_q\}} \left( h(f_p) + M(f_p, f_q) \right), \quad (6)$$

where  $\mathcal{N}_{f_q}$  denotes the neighbors of  $f_q$  in the model graph.

$$m_{pq}^t(f_q) = \min \left( H(f_q), \min_{f_p} h(f_p) + d \right) \quad (7)$$

### B. Expression matching

To match pairs of expressions, we exploited shape contexts (SC) [6] as vertex attributes. In our preliminary experiment, each symbol is represented by its centroid, defined by the respective stroke coordinates – thus each expression can be viewed as a point set. Then, we calculate (global) SC of each vertex, considering the coordinates of the remaining vertices.

To minimize the ambiguities between vertex attributes, we exploited geometric properties based on the spatial distribution of symbols. Ideally, the structure should be invariant among the corresponding expressions.

In this preliminary study, we tested the proposed approach by matching isomorphic graphs. In general, non-isomorphic graphs are expected, but the goal here was to evaluate the capabilities of the method. In order to build isomorphic graphs for each pair of expressions, firstly a Delaunay triangulation was computed from the first expression. Then, the same structure was replicated to the second expression, based on the true matchings. Although this may sound very artificial, the distortions can be quite significant, as shown in Fig. 4.

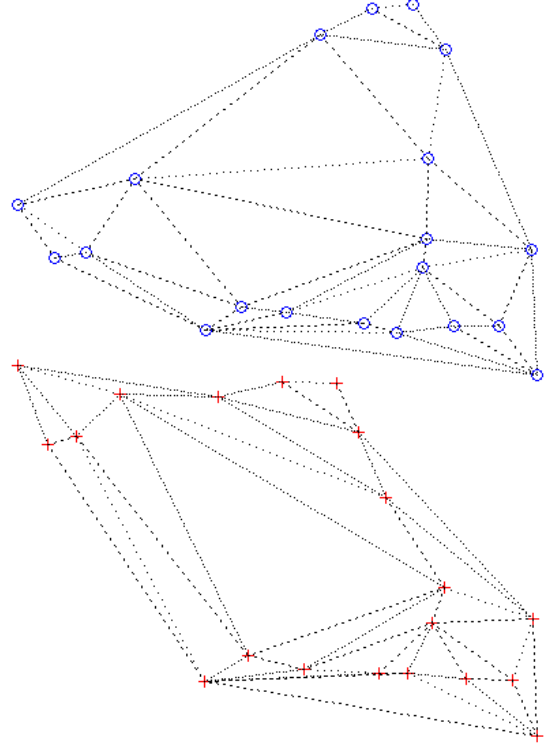


Fig. 4. Two isomorphic graphs based on a Delaunay triangulation of the above point set.

1) *Linear term*: The linear term of Eq. 1 can be defined as the SC metric shown in Eq. 8 – if both  $h_r(k)$  and  $h_s(k)$  are zero, the term is not considered in the summation. The key idea of SC is to characterize each point in terms of the distribution of the remaining points in its neighborhood. Firstly, the neighborhood is divided into sectors by using polar coordinates. Then, a normalized histogram can be computed by counting the number of points in each sector. For each point  $r$ , and for each bin  $k$ , we denote the corresponding normalized histogram by  $h_r(k)$ .

$$d_{\chi^2}(SC(r), SC(s)) = \frac{1}{2} \sum_k \frac{[h_r(k) - h_s(k)]^2}{h_r(k) + h_s(k)} \quad (8)$$

In the present work, the vertex attributes are defined as the global SC computed from the point sets. In this case, there is one point per symbol and the information is based on the spatial configuration of neighboring symbols. Note that, as suggested in [1], we can also exploit local SC, by sampling points from the symbol strokes, taking into account the local information of each symbol shape.

2) *Quadratic term*: The structural term  $M(f_p, f_q)$  is divided into two cases, defined by Eq. 9, following previous works [8], [9]. The first case evaluates the edge attributes (geometric properties of the spatial distribution of symbols), where two vectors are compared in terms of length,  $|\cdot|$  denotes the absolute value and  $|\vec{v}|$  denotes the length of vector  $\vec{v}$ . During the experiments, all vector lengths were normalized

TABLE I  
SELECTED 7 CHALLENGING CASES

Id	#Symbols	Model	Input
1	20	caue_356	miguel_759
2	20	caue_356	danilo_284
3	20	caue_356	silas_366
4	20	miguel_759	daniel_214
5	66	caue_307	Fabricio_176
6	66	caue_307	rosario_447
7	45	caue_343	david_546

between 0 and 1. Similar to the Potts model, the second case encourages adjacent vertices to have the same label, by penalizing when there is no such corresponding edge in the model graph.

$$M(f_p, f_q) = \begin{cases} |\nu_i(p, q) - \nu_m(f_p, f_q)|, & \text{if } (f_p, f_q) \in E_m \\ d, & \text{if } (f_p, f_q) \notin E_m \text{ and } f_p \neq f_q \end{cases} \quad (9)$$

#### IV. PRELIMINARY EXPERIMENT

To perform our experiments, we used expressions taken from *ExpressMatch* dataset [2]. This dataset is composed of 56 models and 901 transcribed (input) expressions. From the dataset, we have selected 7 challenging cases such as the ones described in Section II and those are shown in Table I. Model and input expressions are identified by their corresponding ID defined in the *ExpressMatch* dataset.

Table II compares the results obtained with the original [2] and the proposed techniques. The proposed approach outperformed the original method in all cases. It is interesting to note that even for the largest expressions – pairs 5 and 6, that include matrix and fraction structures – the proposed algorithm achieved perfect correspondence. Fig. 5 shows expressions of pair 5.

Fig. 5. Pair 5 of the evaluating expressions: caue\_307 (top) and Fabricio\_176 (bottom)

Although the results in Table II seem promising, it is important to stress that they are based on an ideal structure, which results from isomorphic graphs.

#### V. CONCLUSIONS AND FURTHER WORK

In this paper, we have introduced a structural matching technique based on belief propagation to match corresponding

TABLE II  
PRELIMINARY COMPARISON RESULTS

Id	Original		Proposed	
	#Errors	%Errors	#Errors	%Errors
1	10	50%	0	0%
2	11	55%	2	10%
3	5	25%	3	15%
4	9	45%	1	5%
5	8	12.1%	0	0%
6	23	34.8%	0	0%
7	17	37.8%	0	0%

symbols from a pair of handwritten mathematical expressions. In our preliminary experiments, we used spatial configuration between symbols to produce isomorphic graphs to make the structural information invariant among the corresponding expressions. The results suggest that the proposed approach can significantly improve the correspondences under such assumptions. Thus, if it were possible to build nearly isomorphic graphs for the expressions, the proposed method has potential to improve current matching rates.

Current research is being dedicated to test the proposed technique to non-isomorphic graphs by exploiting local shape contexts (for each symbol shape) and to find realistic robust structures to be used in practice. Future work includes a research of more realistic structures, where edges are calculated independently for each graph (input and model), and testing the proposed approach on all expressions of the *ExpressMatch* dataset.

#### ACKNOWLEDGMENT

F. D. Julca-Aguilar is supported by FAPESP (Grant number 2012/08389-1). N. Hirata is partly supported by CNPq.

#### REFERENCES

- [1] N. S. T. Hirata and W. Y. Honda, "Automatic labeling of handwritten mathematical symbols via expression matching," in *Proceedings of the 8th International Conference on Graph-based Representations in Pattern Recognition*, 2011, pp. 295–304.
- [2] F. D. J. Aguilar and N. S. T. Hirata, "ExpressMatch: a system for creating ground-truthed datasets of online mathematical expressions," in *Proceedings of 10th IAPR International Workshop on Document Analysis Systems*, 2012, pp. 155–159.
- [3] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *Intl. J. Comput. Vision*, vol. 70, no. 1, pp. 41–54, 2006.
- [4] A. Noma, A. Pardo, and R. M. Cesar-Jr, "Structural matching of 2D electrophoresis gels using deformed graphs," *Pattern Recognition Letters*, vol. 32, pp. 3–11, 2011.
- [5] A. Noma, A. B. V. Graciano, R. M. Cesar-Jr, L. A. Consularo, and I. Bloch, "Interactive image segmentation by matching attributed relational graphs," *Pattern Recognition*, vol. 45, pp. 1159–1179, 2012.
- [6] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, April 2002.
- [7] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Underst.*, vol. 89, no. 2-3, pp. 114–141, 2003.
- [8] A. Noma, L. Velho, and R. M. Cesar-Jr, "A computer-assisted colorization approach based on efficient belief propagation and graph matching," in *CIARP '09: Proceedings of the 14th Iberoamerican Conference on Pattern Recognition*. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 345–352.
- [9] A. Noma and R. M. Cesar-Jr, "Sparse representations for efficient shape matching," in *Proceedings of the 23rd SIBGRAPI (Conference on Graphics, Patterns and Images)*, 2010, pp. 186–192.