Automated Spot Matching in Autoradiogram Images of Two-Dimensional Electrophoresis of Genomic DNA

Junichi Isamikawa [1]
isami@infor.kanazawa-it.ac.jp
Katsutoshi Takahashi [1]
sltaka@infor.kanazawa-it.ac.jp
Masayuki Nakazawa [2]
nakazawa@nakazawa.ishikawa-pc.ac.jp
Yasuo Watanabe [1]
watanabe@infor.kanazawa-it.ac.jp

[1] Kanazawa Institute of Technology, 7-1 Ohgigaoka Nonoichi-machi, Ishikawa-gun, Ishikawa 921, Japan

[2] Ishikawa Polytechnic College, I 45-1 Yuigaoka Anamizu-machi, Hugeshi-gun, Ishikawa 927, Japan

In this paper, we present the pattern matching of autoradiogram images obtained with two-dimensional gel electrophoresis based on the RLGS (restriction landmark genomic scanning) method[1]. After detecting several thousand of spots, the structured features of the spot pattern is extracted and represented as a structured graph.

The problem is matching two or more images to detect differences among RLGS patterns. Because of nonlinear distortion and density variation of a RLGS image, a simple pattern matching algorithm base on correlation does not succeed. We have developed a robust and fast matching algorithm by means of Delauney net and relative neighborhood graph (referred to as DN and RNG, respectively.) The graph matching starts with the initial points which are given manually in both graphs. And then, it progresses with breadth-first search, using RNG as a guide for matching with DN and evaluating both geometrical factor and image density of corresponding spots. That is, the algorithm traverses RNG, minimizing the following function for the depth k.

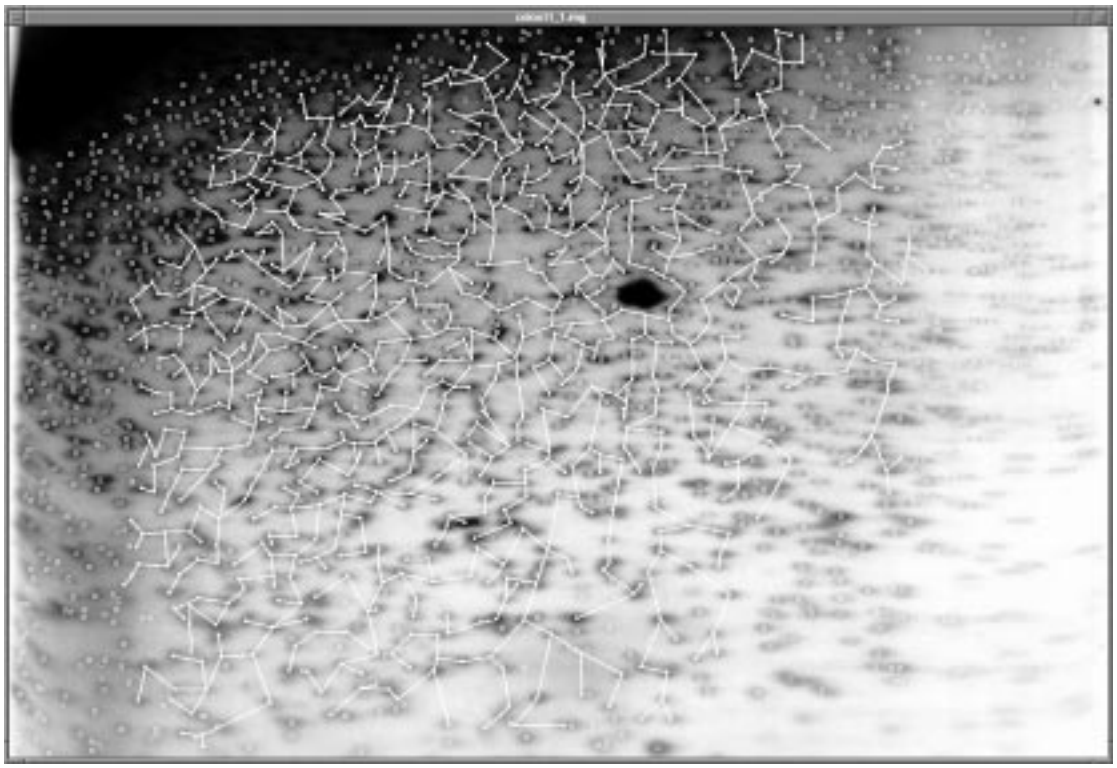
$$E_k = \sum_i \Big[\{1 - S(a_i^{(k)}, F(a_i^{(k)}))\}^2 + \{1 - S(P_r(a_i^{(k)}), P_o(F(a_i^{(k)})))\}^2 \Big],$$

where for the vectors u and v, $S(u,v) = \frac{(u,v)}{|u||v|}$, $a_i^{(k)}$ is either the i-th vector corresponding the directed arc to be traversed or the virtual vector directed to the node which is directly connected from the destination node of the "real" directed arc. F is the mapping from arcs of the RNG in the reference image to arcs of the DN in the object. In addition, $P_r(x)$ is the vector representing a reference subimage centering the destination point of x and $P_o(x)$ means the similar vector with respect to the object subimage. In addition, our algorithm eliminates redundant paths, by using thresholds of difference of the corresponding arcs, i.e, (1) $\arccos(S(u,v))$, (2) $||u| - |v||$, and (3) i-th term of E_k. Here, we should select the respective values as thresholds, taking statistics of a variety of patterns. Applying our matching algorithm to the whole RLGS patterns, one is the reference RNG (a) with 2104 nodes and 2891 arcs and the other is the object DN (b) with 1336 nodes and 3797 arcs as shown in figure 1, about a thousand pairs were matched, of which 97.5% are correct. And it took only 9.8 sec to perform the matching on DOS/V computer(Pentium/75MHz with 16MB EDO-DRAM) running with Linux.

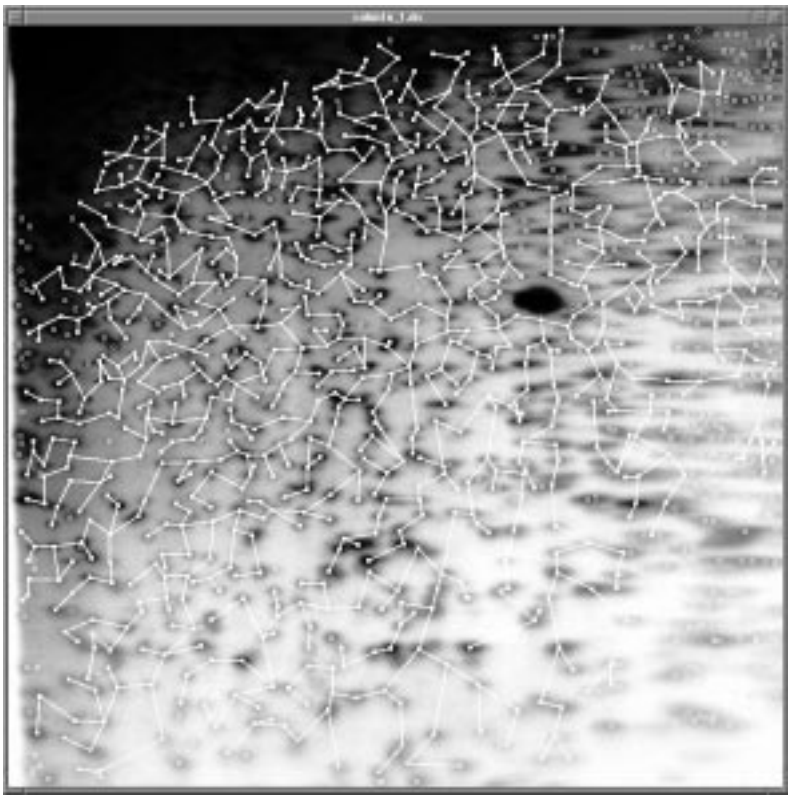
References

[1] Y. Hayashizaki, et al. "Restriction landmark genomic scanning method and its various application," *Electrophoresis*, 14, pp. 251–258, 1993.

[2] K.Takahashi, M.Nakazawa, Y.Watanabe. "DNAinsight: An Image Processing System for 2-D Gel Electrophoresis of Genomic DNA," *Proc. of the 8th Workshop on Genome Informatics*, 1997.



(a) Matched RNG of the reference RLGS Pattern.



(b) Matched DN of the object RLGS Pattern.

Figure 1: Matching RNG with DN.