# Layered Shape Matching and Registration: Stochastic Sampling with Hierarchical Graph Representation

Xiaobai Liu[1,3], Liang Lin[2,3], Hongwei Li[3], Hai Jin[1], Wenbing Tao[1]
[1] SCTS&CGCL, School of Computer Science and Technology, HUST, Wuhan, 430074, China
[2] School of Info. Sci. and Tech., Beijing Institute of Technology, China
[3] Lotus Hill Research Institute, China
hjin@hust.edu.cn

## Abstract

*To automatically register foreground target in cluttered images, we present a novel hierarchical graph representation and a stochastic computing strategy in Bayesian framework. The graph representation, which contains point-(image primitives), seedgraph-, and subgraph- three levels, are built up following the primal sketch theory to capture geometric, topological, and spatial information both in local and global scale. We use two types of bottom-up algorithms for searching matching candidates to generate the point-level and seedgraph-level representations respectively. Then, the Swendsen-Wang Cuts and Gibbs sampling methods are performed for global optimal solution to generate the final subgraph-level representation, where a mixture bending function and a set of topological operators are defined for matching measurement. Experiments with comparison are demonstrated on standard dataset with outperforming results. Results show that our method can work well even with clutter noise and complex background.*

## 1. Introduction

Although shape matching is extensively studied in computer vision community in last a few years, using template matching for object instance detection (registration) in cluttered images is still a challenging task, due to the main difficulties: 1) developing representations to capture important shape variations; 2) searching correct correspondences based on geometrical and/or topological information. Addressing these two points, this paper proposes a integrative framework, which includes a novel shape representation and a shape matching algorithm.

The previous literatures of shape representation can be roughly divided into three categories: 1) Point set based, such as shape context [1] and Data-Driven EM [2]. These approaches represent shape via sampled points from shapes and they only work well with the objects with less topological structure in clean images. 2) Curve based, including Hausdorff distance [3] and its variants. This category defines the matching functions of shape based on curve measurement. 3) Graph based, including skeleton graphs [4] and layered graph [5]. They represent shapes with explicit graph structures and graph editing. The main limitations for them are the expensive computational cost and may fail under the situation of partial occlusion.

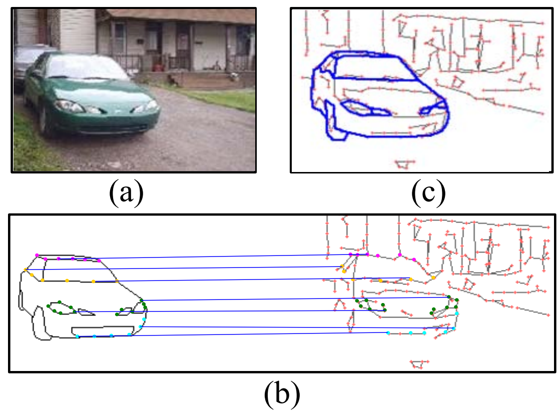



Figure 1: Typical example. (a)Test image; (b)Matching between the template shape and the sketch image; (c) Registered image.

The work in this paper belongs to the third categories, and naturally combines the advantages of other two. We represent the shape using hierarchical graphs, which contains three graphs with vertex being key-points(image primitives), seedgraph(several connected points) and subgraph(several composite seedgraph) respectively. This model is approved to capture local and global information in each scale. In computation, we formulate the shape matching problem in Bayesian framework and propose a layered stochastic algorithm. First, we build two point-layer (flat) attribute graphs

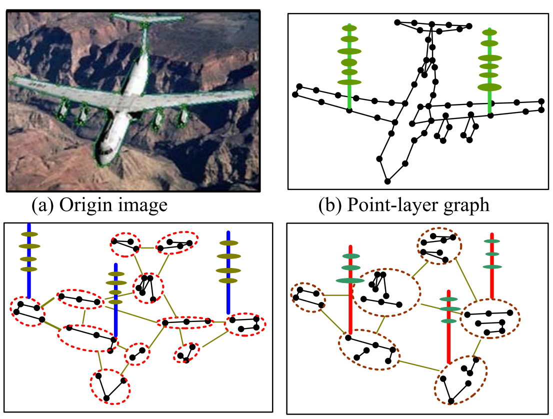
978-1-4244-2175-6/08/$25.00 ©2008 IEEE

and keep the matching candidates through local features. Second, a branch-and-bound strategy is applied to generate the seedgraph-layer graph and their candidates. Last, two MCMC samplers, SW-Cut and Gibbs, are performed to maximize the posterior probability following the Metropolis-hastings principle. Fig. 1 illustrates a representative result.

Compared with the state-of-the-art shape matching algorithm, the proposed approach has the following contributions: 1) A novel hierarchical graph is introduced to represent the shape at different scale; 2) The matching problem is formulated as maximizing a posterior probability (MAP) and two MCMC samplers are performed for nearly global optimal solution in a coarse-to-fine fashion. 3) a set of graph editing operators is proposed to measure the structural distance.

The remainder of this paper is organized as follows. We first introduce the hierarchical shape model with local feature descriptors in Sec. 2. Then, we formulate the problem and define matching measurement in Sec. 3. The matching algorithm is introduced in Sec. 4 and the experiments ares presented in Sec. 5. The paper is concluded in Sec. 6 with the future works.

## 2. H-Graph Representation



(a) Origin image (b) Point-layer graph (c) Seedgraph-layer graph (d) Subgraph-layer graph

Figure 2: H-Graph shape representation

Following the primal sketch theory [4], we represent the shape using a hierarchical attribute graph which is denoted as $G_i = (V_i, E_i, A_i), i = 1, 2, 3$, with $V$ being a set of vertices for the key point(image primitives), seedgraph or subgraph, $E$ being a set of edges for the connectivity of the nodes and $A$ a set of attributes defined for local features. In our model, one seedgraph node from $V_2$ consists of several connected key points and itself is a point-level graph defined as $G_1$. One subgraph node from $V_3$ consists of several distinct seedgraph and itself is a seedgraph-level attribute graph defined as $G_2$. This compositional representation is called "H-Graph", as illustrated in Fig. 2.

We define the attributes set and the matching measures for different layer in H-Graph as follows:

I. For the point-level graph $G_1$, the attribute set $A_1$ is defined as, $A_1 = \{F_1(x_i), x_i \in V_1, i = 1, \ldots, N_1\}$, with $N_1$ being the number of the vertices, and $F_1$ the local feature descriptor of $x_i$. To obtain descriptor $F_1$ for point $x_i$, we draw an circle with radius r and collect all the points that fall with in the circle. The angle of these points relative to $x_i$ are computed and the angle histogram is used as the local feature: $F_1(x_i) = \{h(\theta_j), j = 1, \ldots, 6\}$.

II. For the seedgraph-level graph $G_2$, the attributes set $A_2$ is denoted as: $A_2 = \{F_2(g_i), i = 1, \ldots, N_2\}$, $F_2(g_i) = \{F_1(x_j), x_j \in g_i, j = 1, \ldots, |g_i|\}$. Thus, the similarity of two seedgraphs is denoted as $H_{D1}$, which can be computed using Hausdorff Distance as:

$$H_{D1}(g_a, g_b) = max(D(g_a, g_b), D(g_b, g_a)) \quad (1)$$
$$D(A, B) = \max_{p_1 \in A} \min_{p_2 \in B} \|h_1(p_1) - h_1(p_2)\|$$

III. For the subgraph-level graph $G_3$, attributes set $A_3$ is defined as: $A_3 = \{F_3(\bar{g}_i), i = 1, \ldots, N_3\}, F_3(\bar{g}_i) = \{F_2(g_j), g_j \in \bar{g}_i\}$. The matching measures between two subgraphs can be calculated through:

$$H_{D2}(\bar{g}_i, \bar{g}_j) = \max(\bar{D}(\bar{g}_i, \bar{g}_j), \bar{D}(\bar{g}_j, \bar{g}_i)) \quad (2)$$
$$\bar{D}(A, B) = \max_{A \in \bar{g}_i} \min_{B \in \bar{g}_j} H_{D1}(g_a, g_b)$$

Generally, the shape model proposed has following benefits: 1)*H-graph* is a compositional description of shape, which can capture both the local and global shape information. 2)Combining with the geometric and spatial constraints, we can build the *H-Graph* and prune the matching candidates simultaneously.

## 3. Bayesian Formulation

The task of shape matching is to find the corresponding relation between the source shape and the target shape. Given the first layer graph $G_1^S$ and $G_1^T$, the solution configuration is defined as $W = \{K, \Pi, \Psi, \Phi\}$, where $K$ denotes the number of subgraphs, $\Pi$ the subgraph partition, $\Psi$ the matching relation between source vertex and target vertex, $\Phi$ the matching energy from $G_1^S$ and $G_1^T$. The subgraph partition is defined as: $\Pi = \{\bar{g}_0, \bar{g}_1, \ldots, \bar{g}_k\}, \cup_{i=0}^{K} \bar{g}_i = G_1$, where each subgraph $\bar{g}_i = \cup_j g_j = \cup_j \{V_1^j, E_1^j, A_1^j\}$, is a separate layer and itself an attribute graph.

Letting $V_i^S$ and $V_i^T$ denote the vertices from $G_1^S$ and $G_1^T$ respectively, we have the matching relation as: $\Psi = \cup_{i=0}^{N1} \psi_i, \psi_i : V_i^S \rightarrow V_i^T \cup \{\emptyset\}$.

Given matched subgraphs pair $(\bar{g}_i, \bar{g}_j), i = 1, \ldots, K)$, we define the matching energy in three aspects: $\Phi =$

$(H_{D2}, \Phi^{geo}, \Phi^{top})$, where $H_{D2}$ has been defined in Sec.2. The geometric transform $\Phi^{geo}$ is defined using a global affine transformation $A_i$ and a TPS (Thin-Plate-Spline) warping for deformation $F_i(x, y)$ on a 2D domain $\wedge_i$ covered by $\bar{g}_i$. Thus, we have: $\Phi_{geo}(\bar{g}_i, \bar{g}_j) = \omega_1 \cdot E_A(\bar{g}_i, \bar{g}_i) + \omega_2 \cdot E_F(\bar{g}_i, \bar{g}_j)$ where $E_A(\bar{g}_i, \bar{g}_j)$ indicates the affine transformation and $E_F(\bar{g}_i, \bar{g}_j)$ the TPS bending. The constant coefficients $\omega_1$ and $\omega_2$ can be assigned empirically ($\omega_1 = 0.08, \omega_2 = 0.92$).

The topological distance $\Phi^{top}$ is used to preserve the graph connectivity and correct errors which are caused by inference uncertainties, cluttered noise, and complex background structure. To simplify the computation, only two basic operators are defined in our approach, that is, $\sigma^A and \sigma^B$. The former remans missing/adding node and the latter missing/adding arm. Thus, given two subgraphs $\bar{g}_i$ and $\bar{g}_j$, we get the topological cost as: $\phi^{top}(\bar{g}_i, \bar{g}_j) = M \cdot cost(\sigma^A) - N \cdot cost(\sigma^B)$, which suppose the operator $\sigma^A$ and $\sigma^B$ are used times $M$ and $N$ to compensate the topology difference of the two subgraphs. As Fig. 3 illustrates, the top row gives two basic operators and the other rows give the derived ones.

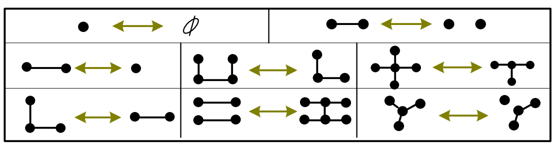


Figure 3: Graph editing operators

Thus, the shape matching problem is formulated as maximizing a posterior (MAP) of a layered partition, given the source and target attribute graphs:

$$W^* = \arg\max_W P(W|G_1^S, G_1^T) = \arg\max_W P(W|G_1^S) \cdot P(G_1^T|W, G_1^S) \quad (3)$$

We assume that subgraphs are independent with each other, thus, the prior and likelihood terms are defined as follows:

$$P(W|G_1^S) = p(K)p(\Pi|K, G_1^S)p(\bar{g}_0)p(\bar{g}'_0) \cdot \prod_{i=1}^{K} p(\bar{g}'_i|\psi_i, \phi_i, \bar{g}_i) \quad (4)$$

$$P(G_1^T|W, G_1^S) = \prod_{i=1}^{K} p(\phi_i|\psi_i, \bar{g}_i)p(\psi_i|\bar{g}_i) \quad (5)$$

In this model, the term $p(K)$ penalizes the number of layers and can be predicted in the exponential form,

$$p(K) = \exp\{-\lambda_1 K\} \quad (6)$$

The partition prior $p(\Pi|K, G_1^S)$ is modeled as the Potts model, as,

$$p(\Pi|K, G_1^S) = \exp\{-\lambda_2 \sum_{<s,t>\in E_1^S} 1(l(s) = l(t))\} \quad (7)$$

where $1(x) \in \{-1, +1\}$ is an indicator function for a Bool variable $x$. To narrow the number of the free vertices, that doesn't belong to any seedgraphs, we model $p(\bar{g}_0)$ and $p(\bar{g}'_0)$ as ,

$$p(\bar{g}_0) = \exp\{-\alpha|\bar{g}_0|\}, p(\bar{g}'_0) = \exp\{-\beta|\bar{g}'_0|\} \quad (8)$$

As the matching relation between $\bar{g}_i$ and $\bar{g}'_i$ is deterministically, we can easily predict the term $p(\phi_i|\psi_i, \bar{g}_i)$. However, for $N$ seedgraphs with average $M$ candidates for each of them, the whole solution space is of order $O(N^M)$. To sample this space efficiently, we introduce an integrative algorithm which combine the stochastic sampling in a coarse-to-fine fashion.

## 4. Stochastic Layered Shape Matching

With *H-graph* representation and the formulation, our shape matching method proceeds in 4 steps.

**Step 0**: Initialization and point matching. According to the mathematic model proposed in [4], we first build two graphs $G_1^S$ and $G_1^T$, given source and target images. Then, we match each single node in $G_1^S$ to all nodes in $G_1^T$ and calculate the matching energy using local feature descriptor $h(\cdot)$ as: $Cost(v_i, v_j) = \exp(KL(h(v_i)||h(v_j)) + KL(h(v_j)||h(v_i)))$, where $v_i \in G_1^S, v_j \in G_1^T$, and $KL(\cdot)$ is the Kullback-Leibler divergence between any two histograms.

In our method, most $M_1$ matches which have less matching energy, are selected as the candidates. As Fig. 2(b) illustrates, the blobs on the green line indicates the candidates in the target attribute graph. The size of these blobs represents their weights of matching similarity.

**Step 1**: Seedgraph detection and matching. The branch and bound algorithm is applied to filter the point-level matching and generate the seedgraph-level attribute graph. We start with a random vertex of $G_1^S$, and grow it into small seedgraph by adding neighboring points and re-calculate the matching probability using similarity measures discussed above. Meanwhile, the seedgraph matches are reserved as illustrated in Fig. 2(c). For most real images, we don't prune all ambiguous matches directly and need to keep most $M_2$ candidates for the next two sampling step.

**Step 2**: Subgraph detection and matching pruning with the SW-Cut sampling method. This step is the core of the proposed framework. To rapidly search the complex solution space, the efficient SW-Cut computation method is applied for the second layer graph $G_2^S$ to sampling the posterior probability defined in Eq.3, and the bottom-up information is use to drive the MCMC search for the joint MAP solution.

The SW-Cut proceeds in two steps: (i) generating a key subgraph through composing the initial seedgraphs; and (ii)assigning the matching label(candidates) of the selected subgraph collectively with an acceptance probability calculated following the Metropolis-Hastings principle. The sampling results are the "layered" subgraph with its matches, as Fig.2(d) illustrates.

**Step 3**: The reset isolating graph nodes sampling using Gibbs method. Due to the imperfect sketches from real images, there are still many independent points out of any subgraph. Thus, a Gibbs sampler is adopted for the independent points. To improve efficiency, our method visits the "strong" nodes, which have less matching cost, and propagates the matches to neighbor points.

## 5. Experiments

We test our approach on the LHI Dataset [6] for image tasks as object registration and detection.

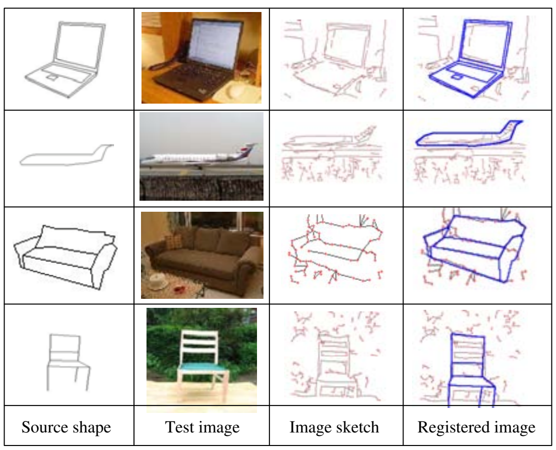



Figure 4: Results of our method

In experiments, we use the *H-Graph* for the template representation and the stochastic sampling as matching algorithm. Let the maximum candidates $M_1$ and $M_2$ be 100, 50 respectively. The maximum iteration times for the SW-Cut sampling is 200, while that for the Gibbs sampling is 100. Fig.1 and Fig.4 give some examples of object registration. Results show that our method can tolerate cluttered noise and complex background. The matching time for each image is about 30-50 seconds on a PC without code optimization.

To quantitatively compare the performance of our approach with shape context [1] and the data-driven EM [2], we run the three algorithms on the same image set. Given a specified template image for each category, our task is to recognize the object type in test image. The data set we used is LHI Dataset [6], which contains 75 types of man-made objects, each of which has about 100 different contours, giving a total of 7500 test image. For each type of objects, we add 100 positive images. Some results with recall precision and false alarm rate are shown in Fig.5.

| | Shape Context | | Data-driven EM | | Our Method | |
|---|---|---|---|---|---|---|
| | Recall Precision | False Alarm | Recall Precision | False Alarm | Recall Precision | False Alarm |
| Sedan(Frontal View) | 67% | 36% | 78% | 15% | 93% | 8% |
| Desk lamp | 73% | 31% | 67% | 14% | 91% | 7% |
| Handbag | 79% | 24% | 69% | 24% | 83% | 8% |
| Teapot | 69% | 38% | 80% | 26% | 89% | 12% |
| Monitor | 67% | 25% | 74% | 21% | 93% | 5% |
| Bicycle( Side View) | 56% | 28% | 67% | 19% | 89% | 9% |

Figure 5: Performance comparison

## 6. Summary

In this paper, we present a novel shape representation, the *H-Graph*, and an efficient matching algorithm for shape matching and registration. We are considering the future works as follows: 1) Integrating graph extraction step with the matching process, and they are approved to be co-related and complementary for each other; 2) Applying this work into a general object recognition framework as a top-down verification module.

## 7. Acknowledgements

This work is done at the Lotus Hill Institute and is supported by NSFC(Grant No.60603024). The work is also partially supported by China 863 program(Grant No.2007AA01Z340 and No.2006AA01Z121).

## References

[1] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts", *IEEETrans.onPAMI*, Vol.24, No.4, 2002, pp.509-522.

[2] Z. W. Tu, S. F. Zheng, and A. Yuille, "Shape Matching and Registration by Data-driven EM", *CVIU*, Vol.103, No.3, 2007, pp.290-304.

[3] O. K. Kwon, D. G. Sim, and R. H. Park, "Object Matching Algorithms Using Robust Hausdorff Distance Measures". *IEEETrans.onImageProcessing*, Vol.8, No.3, 1999, pp.425-429.

[4] S. C. Zhu and A. Yuille, "Forms: A Flexible Object Recognition and Modeling System", *IJCV*, Vol.20, No.3, 1996, pp.187-212.

[5] L. Liang, S. C. Zhu, and Y. T. Wang, "Layered Graph Match with Graph Editing", *InProc.CVPR'*07, Vol.1, June 18-23, 2007, Minneapolis, Minnesota, USA, pp.885-892.

[6] B. Yao, X. Yang, and S. C. Zhu, "Introduction to A Large Scale General Purpose Groundtruth Dataset: Methodology, Annotation Tool, and Benchmarks", *InProc.EMMCVPR'*07, LNCS, Vol.4697, Springer, 2007, pp.169-183.