

# A Discrete Labelling Approach to Attributed Graph Matching using SIFT Features

Gerard Sanromà<sup>a</sup>, René Alquézar<sup>b</sup>, Francesc Serratosa<sup>a</sup>

<sup>a</sup>DEIM, Universitat Rovira i Virgili, Tarragona, Spain

<sup>b</sup>Institut de Robtica i Informtica Industrial, CSIC-UPC, Barcelona, Spain

gerard.sanroma@urv.cat, ralquezar@iri.upc.edu, francesc.serratosa@urv.cat

## Abstract

*Local invariant feature extraction methods are widely used for image-features matching. There exist a number of approaches aimed at the refinement of the matches between image-features. It is a common strategy among these approaches to use geometrical criteria to reject a subset of outliers. One limitation of the outlier rejection design is that it is unable to add new useful matches. We present a new model that integrates the local information of the SIFT descriptors along with global geometrical information to estimate a new robust set of feature-matches. Our approach encodes the geometrical information by means of graph structures while posing the estimation of the feature-matches as a graph matching problem. Some comparative experimental results are presented.*

## 1. Introduction

Image-features matching based on local stable representations has become a topic of increasing interest over the last decade. Mikolajczyk and Schmid [6] evaluated a number of approaches and identified Lowe's SIFT descriptors [4] as the most stable representations.

SIFT features (keypoints) are located at the salient points of the scale-space. Each SIFT feature provides local texture information invariant, at a considerable extent, to image distortions.

A number of approaches have been presented aimed at enhancing the SIFT matching between two images with the use of higher-level information. To cite some examples, RANSAC [3] has been successfully applied to outlier rejection by fitting a geometrical model. More in the topic of the present paper, Aguilar et al. [2] have recently presented an approach that use a graph transformation to select a subset of structurally robust

matches.

One drawback of the two mentioned approaches, and all the outlier rejectors in general, is that they cannot add additional useful matches but only remove incorrect ones. To overcome this limitation we face the enhancement of the SIFT matches as an attributed graph matching problem. We propose a new model for the assignment probabilities based on both the local descriptors and the surrounding matches. We use a discrete labelling scheme in order to update the configuration of matches according to our new model.

We have evaluated the matching precision and recall of our method, which is described in section 2, under different sources of noise. We present in section 3 comparative results with Aguilar et al.'s outlier rejector [2], RANSAC used to fit a fundamental matrix [1], and Luo and Hancock's structural graph matching approach [5]. In section 4 some conclusions are drawn.

## 2 Attributed Graph Matching with SIFTs

We define a graph  $G_0$  representing a set of SIFT keypoints from an image  $I_0$  as a three tuple  $G_0 = (V_0, B, Y)$  where  $v \in V_0$  is a node associated to a SIFT keypoint,  $B$  is the adjacency matrix (thus,  $B_{v,v'} = 1$  indicates that nodes  $v$  and  $v'$  are adjacent and  $B_{v,v'} = 0$  otherwise) and,  $y_v \in Y$  is the SIFT descriptor associated to node  $v$  (a column-vector of length 128).

Consider another image  $I_1$  showing the same scene as  $I_0$  but with some random changes in the viewing conditions such as viewpoint, position, illumination, etc ... Consider the graph  $G_1 = (V_1, A, X)$  to represent a set of keypoints from  $I_1$ .

Our approach to attributed graph matching aims to estimate an assignment function  $f : V_1 \rightarrow V_0$  with the use of SIFT descriptors as attributes. Therefore,  $f(u) = v$  means that node  $u \in V_1$  is matched to node  $v \in V_0$ , and  $f(u) = \emptyset$  means that it is not matched to any node.

## 2.1 A Discrete Labelling Scheme

We want to maximise the joint probability of a graph given the assignment function  $f$ . To do so, our iterative algorithm visits all the nodes of the graph at each iteration, and updates  $f$  in order to increase this joint probability

$$P(G_1|f) = \prod_{u \in V_1} P(u|f(u)) \quad (1)$$

The update equation of the assignment function is

$$f^{(n+1)}(u) = \arg \max_{\{f^{(n)}(u) \in V_0\} \cup \{\emptyset\}} P(u|f^{(n)}(u)) \quad (2)$$

We have designed our matching criterion as a product of two quantities:

$$P(u|f(u)) = P_u Q_u \quad (3)$$

where  $P_u$  and  $Q_u$  stand for the matching probability according to the current node attributes and its adjacency relations, respectively.

## 2.2 Matching probability according to the current node attributes

This measure represents the likelihood of a node association regardless of the adjacency relations. We want our algorithm to behave the same as SIFT matching when no graph structure is present.

Consider two sets of SIFT keypoints  $X, Y$  from two images  $I_1$  and  $I_0$ . According to SIFT matching, a keypoint  $i$  from  $I_1$  with descriptor (column) vector  $x_i$  is matched to a keypoint  $i'$  from  $I_0$  iff:  $i' = \arg \min_{\hat{i}'} \|x_i - y_{\hat{i}'}\|$ , and  $\frac{\|x_i - y_{i'}\|}{\|x_i - y_{i,2min}\|} < \rho$ , where  $\|x\| = \sqrt{x^\top x}$  is the Euclidean length ( $L^2$  norm),  $y_{i,2min}$  is the descriptor with the second smallest distance from  $x_i$ , and  $0 < \rho \leq 1$  is a parameter defining the ratio of acceptance of candidates. If these conditions are not met, then keypoint  $i$  is leaved unmatched.

Our model for the matching probability according to the node attributes between nodes  $u \in V_1$  and  $f(u) \in V_0$  is

$$P_u = \mathcal{N}(\|x_u - y_{f(u)}\|; 0, \sigma) \quad (4)$$

where  $x_u$  and  $y_{f(u)}$  are the SIFT descriptors of nodes  $u \in V_1$  and  $f(u) \in V_0$ , and  $\mathcal{N}(\bullet; 0, \sigma)$  is the normal probability function with zero mean and sample standard deviation

$$\sigma = \frac{1}{|V_1||V_0|} \sum_{u \in V_1} \sum_{v \in V_0} \|x_u - y_v\| \quad (5)$$

The SIFT matching criterion states that a keypoint  $i$  can only be matched to another keypoint when the distance between their descriptors is below  $\rho \|x_i - y_{i,2min}\|$ . Accordingly, we define the threshold probability for sending a node  $u \rightarrow \emptyset$  as

$$P_{u\emptyset} = \mathcal{N}(\rho \|x_u - y_{u,2min}\|; 0, \sigma) \quad (6)$$

## 2.3 Matching probability according to the adjacency relations

Our aim is to define a model that takes into account the matching quality of the adjacent nodes.

We define the assignment variable  $S$  such as  $s_{u,v} \in S$  and  $s_{u,v} = 1$  if  $f(u) = v$  and  $s_{u,v} = 0$  otherwise.

Luo and Hancock [5] showed how to factorise, using the Bayesian theory, the hard-to-model matching probability given the entire state of the assignments  $S$  into easy-to-model unary assignment probability terms:

$$P(u|f(u), S) = g_u \prod_{u' \in V_1} \prod_{v' \in V_0} p(u|f(u), s_{u',v'}) \quad (7)$$

where  $g_u = [1/p(u)]^{|V_1||V_0|-1}$  is a constant depending on the identity of node  $u$ .

It is a well-known strategy to state that a match from a node  $u \in V_1$  to a node  $f(u) \in V_0$  is more likely to occur as more nodes adjacent to  $u$  are assigned to nodes adjacent to  $f(u)$  [5].

We define a *hit* as a node  $u' \in V_1$  adjacent to  $u$  that is matched to a node  $f(u') \in V_0$  adjacent to  $f(u)$ . The model for the unary assignment probabilities presented in [5] used the Bernoulli distribution in order to accommodate *hits* and *no hits* with fixed probabilities  $(1 - P_e)$  and  $P_e$  (being  $P_e$  the probability of error). We present a new model aimed at giving a more fine-grained measure by assessing the *hits* according with their quality, while giving room for possible structural errors in the case of *no hit*. The proposed expression is

$$p(u|f(u), s_{u',v'}) = \frac{P_{u'}^{A_{uu'} B_{f(u)v'} s_{u',v'}}}{[\xi P_{u'\emptyset}]^{(1 - A_{uu'} B_{f(u)v'} s_{u',v'})}} \quad (8)$$

where  $A$  and  $B$  are the adjacency matrices of  $G_1$  and  $G_0$ , respectively;  $P_{u'}$  is the quality term of matching node  $u' \in V_1$  according to eq. (4) and,  $[\xi P_{u'\emptyset}]$  is the ground-level contribution in the case of *no hit* expressed in reference to  $P_{u'\emptyset}$  (eq. (6)). The parameter  $0 < \xi \leq 1$  regulates the ground-level contribution. When  $\xi \rightarrow 0$ , there is small room for structural errors and then, the update equation (3) relies mostly on the structural model. On the other hand, when  $\xi \rightarrow 1$ , the ground-level approaches the quality term and the structural model becomes ambiguous.

In a similar manner that is done in [5], we state equations (7) and (8) in the exponential form, obtaining the following expression

$$Q_u = h_u \exp \left[ \sum_{u'} \sum_{v'} \log \left( \frac{P_{u'v'}}{\xi P_{u'\emptyset}} \right) A_{uu'} B_{f(u)v'} s_{u'v'} \right] \quad (9)$$

where  $h_u = \exp \left[ \sum_{u'} \sum_{v'} \log(\xi P_{u'\emptyset}) \right] g_u$  is a constant that does not depend on either the graph structure or the state of the correspondences.

Finally, we define the threshold probability for sending a node  $u \in V_1$  to *null* as

$$Q_{u\emptyset} = h_u \exp \left[ K_\emptyset \log \left( \frac{1}{\xi} \right) \right] = h_u \exp \left[ -K_\emptyset \log(\xi) \right] \quad (10)$$

where  $K_\emptyset \geq 0$ ,  $K_\emptyset \in \mathfrak{R}$  is a parameter defining the minimum number of *hits* with quality term  $P_{u'} \geq P_{u'\emptyset}$  required for the match  $u \rightarrow f(u)$  to be more structurally likely than  $u \rightarrow \emptyset$ .

### 3 Experiments and Results

We have compared our attributed graph matching method (AGM) to the following approaches: Graph Transformation Matching (GTM) [2], RANSAC used to fit a fundamental matrix [1] and, Structural Graph Matching with the EM Algorithm (GM-EM) [5]. We have evaluated the matching *Precision* and *Recall* scores of each method under the following types of perturbations: *image distortions*, *geometrical noise* and *clutter* (point contamination). We have used the *F-measure* to plot the results. F-measure is defined as the weighted harmonic mean of Precision and Recall and its expression is:  $F = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$ .

Graph structures have been generated using a K-nearest-neighbours approach with  $K = 4$  in all the methods needed (i.e., edges are placed joining a keypoint with its  $K$  nearest neighbours in space). Additionally, in the case of our method (AGM) we have included the SIFT descriptors as attributes of the nodes. All the methods have been initialized with the configuration of matches returned by a classical SIFT matching using a ratio  $\rho = 1$  (the best value possible for the outlier rejectors). The keypoint-sets size used in the experiments has been  $N = 20$ . Our method (AGM) has done 20 iterations, and we have used  $\xi = 0.5$  and  $\rho = 1$ . We have empirically set  $K_\emptyset = 1.6$  in the clutter experiments, and  $K_\emptyset = 0$  in the others. The tolerance threshold for RANSAC has been set to 0.01, and the number of iterations to 1000 (as suggested in [1]). The probability of error  $P_e$  for the GM-EM method has been set to 0.0003, and the number of iterations to 100.

For each experiment we have arbitrarily chosen a grayscale image  $I_0$  from the Camera Movements and Deformable Objects' databases used in [2].

In the image distortion experiments, we generate  $I_1$  by simultaneously applying the following types of perturbations to  $I_0$ : image resizing, to simulate changes in the distance from the objects in the image; image rotation, to simulate changes in viewpoint; image intensity adjustment, to simulate illumination changes and; gaussian white noise addition to pixel intensity values, to simulate deterioration in the viewing conditions.

We extract the SIFT keypoints from images  $I_1$  and  $I_0$ , obtaining coordinate vector-sets  $\mathbf{P}$  and  $\mathbf{Q}$ , and SIFT descriptor-sets  $X$  and  $Y$ , respectively. We define  $\tilde{\mathbf{P}}$  as the result of the mapping from points in  $\mathbf{P}$  back to the reference of  $I_0$ . We compute  $\tilde{\mathbf{P}}$  by applying to  $\mathbf{P}$  the inverse resizing and rotation from the perturbation. We set the ground truth assignments on the basis of the proximity between the points in  $\mathbf{Q}$  and  $\tilde{\mathbf{P}}$ . Then, for a given  $\mathbf{q}_i \in \mathbf{Q}$ , we select as its ground truth assignment the most salient  $\tilde{\mathbf{p}}_j \in \tilde{\mathbf{P}}$  among the ones falling inside a certain radius  $r$  from  $\mathbf{q}_i$ . Saliency is decided according to the gradient magnitude of the SIFT features [4]. The proximity radius has been set to  $r = 0.03 \times l$ , where  $l$  is the diagonal-length of the image. The keypoints that are not involved in any ground truth assignment are discarded. So, at the end of this step we end up with keypoint-sets  $\mathbf{Q}' = (\mathbf{q}'_1, \dots, \mathbf{q}'_N)$  and  $\mathbf{P}' = (\mathbf{p}'_1, \dots, \mathbf{p}'_N)$ , and a bijective mapping  $f_{gtr} : \mathbf{P}' \rightarrow \mathbf{Q}'$  of ground truth assignments.

Once the  $N$  ground truth assignments have been established, we implement the clutter by adding a certain amount of the remaining points in both  $\mathbf{P}$  and  $\mathbf{Q}$  to  $\mathbf{P}'$  and  $\mathbf{Q}'$ . Clutter points are carefully selected not to fall inside the radius of proximity  $r$  of any pre-existent point. Thus, we can safely assume that they have no correspondence in the other point-set.

Finally, geometrical noise consists on adding random gaussian noise with zero mean and a certain standard deviation  $\sigma_g$  to the point positions  $\mathbf{p}_i = (p_x, p_y)$ . This type of noise simulates nonrigid deformations in the position of the features.

Each plot is the average of the experiments on 10 images. Due to the random nature of the noise, we have run 10 experiments for each image.

Figure 1 shows the F-measure plots for an increasing amount of image distortions. Both geometrical noise and clutter have been set to zero.

Figure 2 shows the results for geometrical noise with  $\sigma_g$  ranging from 0% to 5% of  $\mu_d$  (where  $\mu_d$  is the mean of the pairwise distances between the points). Neither image distortions nor clutter have been introduced.

Figure 3 shows the results for an increasing num-

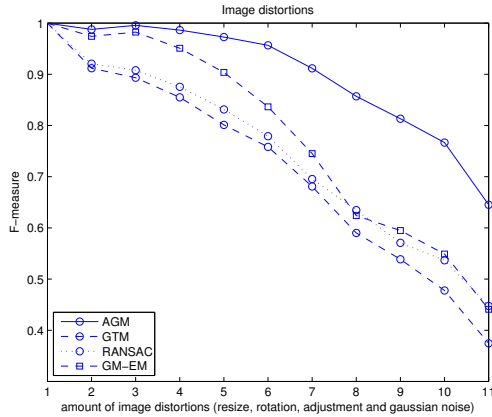


Figure 1. Image distortions.

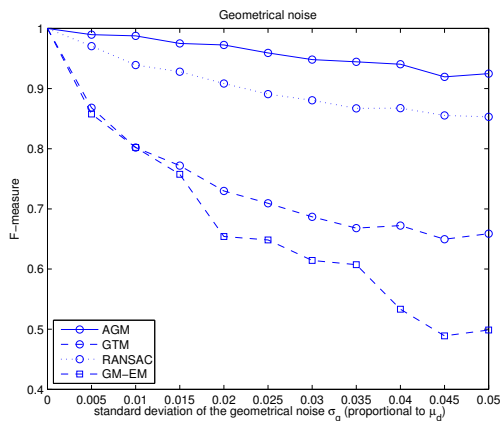


Figure 2. Geometrical noise.

ber of clutter points. The amount of point contamination has ranged from 0% to 80% of the total  $N$  points. Neither background geometrical noise nor image distortions have been introduced.

## 4 Conclusions

We have presented a graph matching model that combines both SIFT features and structural relations. Unlike outlier rejectors, our method faces the enhancement of the SIFT matches as a graph matching problem.

In the image distortion experiments, the methods that are not based on outlier rejection (AGM, GM-EM) recover better than the others from matching misplacements. Specifically our combined approach (AGM) performs better than a purely structural one (GM-EM). In the experiments with geometrical noise, the methods that only use structural information (GTM, GM-EM) experience a considerable decreasing in performance.

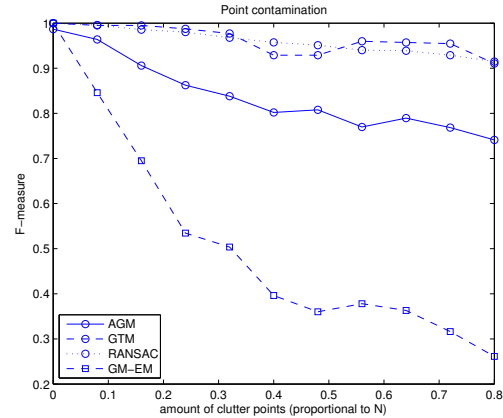


Figure 3. Point contamination.

Our approach (AGM) remains the most stable even in severe noise conditions. In the point contamination experiments, outlier rejectors (GTM, RANSAC) show the best performance. Results suggest us to work towards the achievement of a better stability in front of point contamination.

## Acknowledgements

We want to acknowledge Wendy Aguilar and Francisco Escolano for facilitating us the image databases used in the experiments. This research was partially supported by Consolider Ingenio 2010, project CSD2007-00018, by the CICYT project DPI 2007-61452.

## References

- [1] <http://www.csse.uwa.edu.au/pk/research/matlabfns/>.
- [2] W. Aguilar, Y. Frauel, F. Escolano, and M. E. Martinez-Perez. A robust graph transformation matching for non-rigid registration. *Image and Vision Computing*, 27:897–910, 2009.
- [3] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [4] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), January 2004.
- [5] B. Luo and E. R. Hancock. Structural graph matching using the em algorithm and singular value decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10), October 2001.
- [6] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, October 2005.