Entropy-of-likelihood Feature Selection for Image Correspondence

M. Toews and T. Arbel
Centre for Intelligent Machines
McGill University
3480 University St.
Montreal, Quebec, CANADA H3A 2A7
{mtoews,arbel}@cim.mcgill.ca

Abstract

Feature points for image correspondence are often selected according to subjective criteria (e.g. edge density, nostrils). In this paper, we present a general, non-subjective criterion for selecting informative feature points, based on the correspondence model itself. We describe the approach within the framework of the Bayesian Markov random field (MRF) model, where the degree of feature point information is encoded by the entropy of the likelihood term. We propose that feature selection according to minimum entropy-of-likelihood (EOL) is less likely to lead to correspondence ambiguity, thus improving the optimization process in terms of speed and quality of solution. Experimental results demonstrate the criterion's ability to select optimal features points in a wide variety of image contexts (e.g. objects, faces). Comparison with the automatic Kanade-Lucas-Tomasi feature selection criterion shows correspondence to be significantly faster with feature points selected according to minimum EOL in difficult correspondence problems.

1. Introduction

Image correspondence refers to the task of finding the optimal mapping of feature points in a model image to points in a new image. In this paper, we address the question of how to select the optimal feature points for image matching. Most existing strategies base their selection of "informative" feature points on *subjective* criteria. For example, finding correspondences between face images tends to involve a search for typical facial features such as eyes or noses [5, 20]. In contour-based approaches [9], evenly-spaced points lying on the boundary of an object are assumed to be equally informative. Automated strategies identify informative image feature points based on criteria such as edge density [17]. Although such techniques

work reasonably well in the context for which they were designed, because of the underlying subjectivity of the feature selection, they cannot be guaranteed to work in new contexts. In general, selection of non-subjective feature points for image correspondence has largely been ignored in the literature.

Typically, once feature points are selected, a correspondence model combining an image similarity metric and a regularization strategy is developed, based on the chosen feature points. Next, an optimization strategy is invoked with the objective of attaining an optimally fast, unambiguous correspondence solution. In this paper, we choose to *reverse* this procedure and introduce a novel, non-subjective approach to automatic feature point selection. We show that by basing the choice of features precisely on the given correspondence model, optimization over the selected set will necessarily result in a fast, unambiguous correspondence solution. This approach is non-subjective in the sense that the model itself determines which image feature points are best, eliminating the need for additional context-specific selection techniques.

We embed our approach within a general Bayesian image correspondence model, where image similarity is represented by the likelihood probability distribution. We show how the *entropy of the likelihood (EOL)* distribution can then be used to define a map describing the degree of information content in all points in the input image. Feature points selected according to a minimum EOL criterion are least likely to suffer from correspondence ambiguity, given the similarity metric. This, in turn, leads to a faster and more reliable optimization solution. By describing our criterion with the language of information theory, the strategy is generalizable to any context where the local image similarity metric can be defined through a probability distribution over the space of possible mapping solutions.

The remainder of our paper is organized as follows: Section 2 describes our approach within the context of the Bayesian MRF image correspondence model. We define

a global optimization strategy, and describe how feature points with low EOL result in a fast, unambiguous, global solution. Experimental results in Section 3 show how the EOL criterion identifies optimal feature points for correspondence for a wide variety of images using a simple Bayesian MRF implementation. Correspondence trials are performed comparing EOL feature point selection with the Kanade-Lucas-Tomasi (KLT) [17] operator, a popular automatic feature point selection approach based on edge density. Finally, Section 4 offers a brief conclusion.

2. The Entropy-of-Likelihood (EOL) Criterion

The problem of image correspondence is defined as follows: Given a model image of a scene, I_1, determine the locations of the image pixels in a new image of the scene, I_2, where some arbitrary deformation has taken place in the scene between image acquisitions. Specifically, correspondence seeks a displacement field $\mathbf{T}$ of random vectors, $\mathbf{t}_i$, each of which maps a fixed (x, y) pixel location p_{1i} in I_1 to a random (x, y) pixel location p_{2i} in I_2. Image correspondence requires a model to evaluate the fitness of $\mathbf{T}$, and an optimization strategy to find an optimal instance of $\mathbf{T}$ based on the model.

The goal of correspondence is to find a fast, unambiguous mapping solution. Ideally, a feature point selection criterion should evaluate points according to their potential to fulfill this goal. Such a criterion can only be defined by explicitly considering the correspondence model used, and possibly the optimization strategy. We therefore begin by describing a Bayesian image correspondence model and optimization strategy, which provide the context in which a general and useful criterion can be defined. We then define the EOL criterion for feature point selection.

2.1. Bayesian Image Correspondence Model

The Bayesian Markov random field (MRF) correspondence model provides a general framework for combining local image features with prior knowledge of spatial relationships between features. In the model, similarity and regularization terms are defined as probability densities, whose form can easily be changed without modifying the overall formulation. The model has been widely applied; approaches such as active blobs [15], deformable models [9], and active shape models [2] can all be described using the Bayesian MRF model [10].

The Bayesian approach to image correspondence defines a posterior probability density function over displacement vector field $\mathbf{T}$, conditional on the images I_1 and I_2:

$$p(\mathbf{T}|I_1, I_2) \propto p(I_2|I_1, \mathbf{T})p(\mathbf{T}), \qquad (1)$$

where $p(I_2|I_1, \mathbf{T})$ and $p(\mathbf{T})$ are referred to as the *likelihood* and $p(\mathbf{T})$ *prior* distributions, respectively.

The Markov random field (MRF) is a generalization of the Markov process to higher dimensions, stating that a random vector $\mathbf{t}_i \in \mathbf{T}$ is independent of the rest of the field $\mathbf{T}$ given a subset $\mathbf{N}_i \subseteq \mathbf{T}$ of neighboring random vectors:

$$p(\mathbf{t}_i|I_1, I_2, \mathbf{T}) = p(\mathbf{t}_i|I_1, I_2, \mathbf{N}_i). \qquad (2)$$

A MRF is equivalent to a Gibbs distribution over a random field $\mathbf{T}$ [7]:

$$p(\mathbf{T}) = \frac{exp(-E(\mathbf{T}))}{Z}, \qquad (3)$$

where $E(\mathbf{T})$ is referred to as an energy functional, and Z is a normalization constant. The energy functional is equal to the sum of individual energy terms over all Markov neighborhood *cliques*, i.e. all fully connected singles, pairs, triples, etc... of variables $\mathbf{t}_i$ in neighborhoods $\mathbf{N}_i$ such that:

$$E(\mathbf{T}) = \sum_i E(\mathbf{t}_i) + \sum_{i,j} E(\mathbf{t}_i, \mathbf{t}_j) + ... \qquad (4)$$

The posterior density function in (1) can be described as a MRF by defining both the likelihood and the prior as Gibbs distributions [6, 16]:

$$
\begin{aligned}
p(\mathbf{T}|I_1, I_2) &= \frac{exp(-E(\mathbf{T}|I_1, I_2))}{Z} \\
&= \frac{exp(-E(I_2|\mathbf{T}, I_1)) \, exp(-E(\mathbf{T}))}{Z}. \quad (5)
\end{aligned}
$$

The two energy functionals $E(I_2|\mathbf{T}, I_1)$ and $E(\mathbf{T})$ define the likelihood and the prior, and are specified according to the particular task at hand. We consider the likelihood energy functional as representing a measure of local image similarity between a point in I_1 and it's displacement in I_2. The prior energy functional describes the prior assumptions as to how the neighboring vectors $\mathbf{t}_i$ and $\mathbf{t}_j$ can vary spatially. Our posterior energy functional is therefore of the form:

$$E(\mathbf{T}|I_1, I_2) = \sum_i E_L(I_2|I_1, \mathbf{t}_i) + \sum_{i,j} E_P(\mathbf{t}_i|\mathbf{t}_j), \quad (6)$$

where $E_L(I_2|I_1, \mathbf{t}_i)$ and $E_P(\mathbf{t}_i|\mathbf{t}_j)$ represent the likelihood and the prior energy terms, respectively.

2.2. Optimization Strategy

Once the model has been specified, correspondence is calculated by optimizing over a fitness function. In the Bayesian MRF model, the fitness function is the posterior in (5), and the goal is to find the maximum a posterior (MAP) displacement field instance T_{MAP} such that:

$$T_{MAP} = \underset{T}{\mathrm{argmax}} \, \{ \, p(\mathbf{T}|I_1, I_2) \, \}. \qquad (7)$$

Existing optimization strategies can be broadly classified as either global or local in nature [10]. Local strategies include iterative or gradient based schemes, which are generally sub-optimal and based on assumptions that can produce spurious, incorrect local solutions [8]. Global strategies require searching the entire posterior solution space for a global maximum. Such a search is combinatorial in nature and intractable in general, but feasible for correspondence models based on a small number of feature points [3]. We motivate the EOL criterion using an exhaustive search global strategy, although any optimization strategy can benefit from feature points selected according to the criterion.

An exhaustive search for the global optimum of the posterior in (5) requires a search over all possible combinations of instances of $\mathbf{T}$. If $\mathbf{T}$ consists of P random variables $\mathbf{t}_i$, and the domain of random variables $\mathbf{t}_i$ consists of N discrete instances $t_i^k, k = 1..N$, an exhaustive search is of complexity $O(N^P)$. The search for a globally optimal solution can be viewed as the traversal of a search tree of fixed depth P, where branching at depth i corresponds to choosing instances of $\mathbf{t}_i$. Such a search can be solved using techniques such as dynamic programming [3] or the Viterbi algorithm [13].

In order to perform the search in reasonable time, search heuristics [14] must be used to intelligently prune, or disregard, provably sub-optimal solutions. We consider a depth-first tree search where instances of $\mathbf{t}_i$ are evaluated in a best-first manner, and alpha-beta pruning is used [14]. Such a strategy can efficiently eliminate large regions of the search space when T_{MAP} represents a significant, unique posterior maximum, and is found early on in the search.

2.3. EOL Feature Point Evaluation

With a general model and the optimization strategy defined, we proceed to identify feature points which are likely to result in a fast and unambiguous correspondence solution. We begin by noting that in the tree search strategy, each level i of the tree corresponds to a branching factor of N, where N is the number of discrete instances t_i in the domain of random variable $\mathbf{t}_i$. Using the best-first search heuristic [14], instances t_i are selected for evaluation from most to least probable, in order to locate T_{MAP} early on and effectively prune the search tree.

During the search, the probability of instances t_i can be evaluated using the correspondence model posterior in (5), conditional on instances $t_1..t_{i-1}$ already selected:

$$p(\mathbf{t}_i|I_1, I_2, t_1...t_{i-1}) \propto$$
$$exp(-E_L(I_2|I_1, \mathbf{t}_i) - \sum_{j=1}^{i-1} E_P(t_j|\mathbf{t}_i)). \quad (8)$$

Due to the exponential formulation, the posterior expression in (8) represents an upper bound on the maximum posterior value obtainable at a stage in the search tree. It can

thus be used to prune the tree if another branch has previously yielded a higher solution, i.e. if the branch corresponding to T_{MAP} has already been encountered. In the worst case, the posterior in (8) is uniformly distributed, with many equally probable instances t_i to be evaluated. This results in a large branching factor, and the probability is high that many false branches will be taken before finding T_{MAP}. In the best case, it is highly informative, containing a small number of highly probable locations to be evaluated, after which pruning can effectively eliminate sub-optimal solutions.

The strategy we propose in this paper is to evaluate points $\mathbf{t}_i$ according to how well they lead to an unambiguous, informative posterior in (8). This can be done by calculating the Shannon entropy [4], a well-known measure of probability distribution ambiguity:

$$H(\mathbf{t}_i|I_1, I_2, t_1...t_{i-1}) =$$
$$\sum_{k=1}^{N} p(t_i^k|I_1, I_2, t_1...t_{i-1}) \, log \, \frac{1}{p(t_i^k|I_1, I_2, t_1...t_{i-1})}. \quad (9)$$

In order to simplify the entropy calculation in (9), we propose considering only the likelihood term of the posterior in (8), which represents the likelihood of a match occurring between a feature point in I_1 and a random point in I_2. In effect, the fewer potential matches, the smaller the branching factor of the global optimization strategy, and the smaller the number of potentially ambiguous correspondence solutions. We thus introduce the notion of computing the *entropy of the likelihood* distribution (EOL) as a measure of how likely $\mathbf{t}_i$ is to result in a fast, unambiguous correspondence solution:

$$H(\mathbf{t}_i|I_1, I_2, t_1...t_{i-1}) \approx H(\mathbf{t}_i|I_1, I_2) =$$
$$\sum_{k=1}^{N} exp(-E_L(I_2|I_1, t_i^k))E_L(I_2|I_1, t_i^k). \quad (10)$$

Calculation of (10) is computationally intensive, requiring a sum over the entire domain of $\mathbf{t}_i$ for each point in I_1. This is necessary because intuitively, the EOL criterion provides a measure of the ambiguity of the likelihood over all possible matches in I_2 of a fixed point in I_1. In practice, under the assumption that I_1 and I_2 contain approximately the same local image features, the EOL can be calculated off-line from I_1 alone from the likelihood of $p(\mathbf{t}_i|I_1, I_1)$.

To select feature points for correspondence, an EOL map can be calculated over all image points from which a subset is to be chosen for correspondence. Those chosen according to minimum EOL are more likely to result in fast, unambiguous image correspondence. An example of an EOL map generated for the image of a cat can be found in Figure 1. Note how informative feature points agree with intuitive notions of image importance (e.g. eyes, ears).
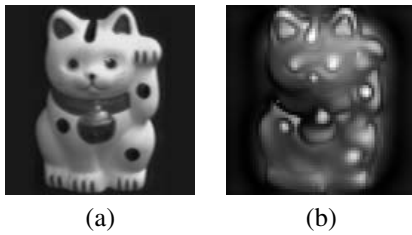
Figure 1. (a) Cat image [12] (b) Entropy-of-likelihood image, mapped such that light pixels correspond to low EOL, highly informative points. In this case, the likelihood of the correspondence model represents a difference-of-Gaussian similarity metric described in Section 3.1.1.

3. Experimentation

In this section, we present a series of experiments performed to demonstrate that the EOL criterion can be used to select optimal feature points for correspondence. We begin by describing the specific Bayesian MRF implementation used throughout the experimentation. Next, we describe our implementation of feature point selection based on the EOL map, and the optimization. Finally, we perform correspondence experiments comparing EOL feature selection with the popular Kanade-Lucas-Tomasi (KLT) automatic feature selection criterion [17], over a wide variety of image pairs. Throughout all experimentation, the same model and optimization is used, and only the feature selection criterion is varied. The trials demonstrate that EOL feature points result in faster, more reliable image correspondence for an arbitrarily specified model.

3.1. Bayesian MRF Implementation

We first describe the particular implementation of the Bayesian MRF model used throughout experimentation. We emphasize that the particular model chosen was only one possible implementation, and the EOL criterion can be used in any correspondence model for which local image similarity can be defined as a probability distribution over the domain of $\mathbf{t}_i$. For convenience, we use p_{1i} and p_{2i} to denote the points in I_1 and I_2 associated with displacement vector $\mathbf{t}_i$.

3.1.1 Likelihood Energy

As mentioned in 2.1, we define the likelihood of $\mathbf{t}_i$ to be a *local* image similarity metric, that is independent of the similarity at neighboring $\mathbf{t}_j$. The likelihood energy thus is defined over single cliques, expressing local similarity as:

$$E_L(I_2|I_1, \mathbf{t}_i) = \sum_k \frac{(f_k(I_1, p_{1i}) - f_k(I_2, p_{2i}))^2}{\sigma_L^2}, \quad (11)$$

where $f_k()$ are local image features, and σ_L^2 is a variance term. Specifying $f_k()$ is task specific; any relevant image features can be used. In this paper, we use multi-scale difference-of-Gaussian (DOG) features [11]. The Gaussian blurring kernel provides partial robustness to small perturbations in local image geometry and spurious noise. DOG features are also rotationally invariant, marginalizing orientation and simplifying the domain of $\mathbf{t}_i$ to (x, y) locations in I_2. For experimentation, $f_k()$ consisted of 4 levels of DOG features, calculated using Gaussian blurring kernels of sigma 2, 4, 8, 16, and 32 pixels, with variance parameter $\sigma_L^2 = 100$.

3.1.2 Prior Energy

The prior energy function is designed to reflect any prior assumptions made regarding the possible spatial variations between features. We define our prior to reflect the general assumption that the relative distance between feature points remains approximately constant between I_1 and I_2:

$$d(p_{1i}, p_{1j}) \approx d(p_{2i}, p_{2j}). \quad (12)$$

where $d()$ is the Euclidean distance. The prior energy is then defined over all neighboring $\mathbf{t}_i$ and $\mathbf{t}_j$ as:

$$E_P(\mathbf{t}_i|\mathbf{t}_j) = \frac{(d^2(p_{1i}, p_{1j}) - d^2(p_{2i}, p_{2j}))^2}{\sigma_{P_{i,j}}^4}, \quad (13)$$

where $\sigma_{P_{i,j}}$ reflects the amount of variation permitted in the relative distance assumption. As for neighborhood relationships, we evaluate the relative distance assumption between all pairs $\mathbf{t}_i$ and $\mathbf{t}_j$. We define $\sigma_{P_{i,j}}$ as proportional to $d(p_{1i}, p_{1j})$ to reflect the assumption that the greater the separation between neighboring points in I_1, the higher the expected variation in their relative distances in I_2. For all experiments, we use $\sigma_{P_{i,j}} = 2d(p_{1i}, p_{1j})$.

3.2. EOL Feature Point Selection and Optimization

Image correspondence trials were performed on input images pairs. For each image pair, the EOL map was calculated off-line for each pixel in image I_1. A set of n^1 feature points is then chosen from n lowest valleys of the map, as these reflect the lowest entropy points. We use $n = 10$ for experimentation. To ensure that feature points were not clustered in a small image region, a minimum distance of 5 pixels was enforced between all feature points.

Optimization was performed on-line as described in Section 2.2, to find the optimal T_{MAP} displacement of feature points from I_1 to I_2. In the search, variables were ordered in decreasing significance according to the particular feature selection criterion, and instances of variables were

---

[1]n is an arbitrary feature set size chosen empirically.

(a) I_1: Face image   (b) EOL map

(c) EOL feature points I_1 (d) Corresponding points in I_2
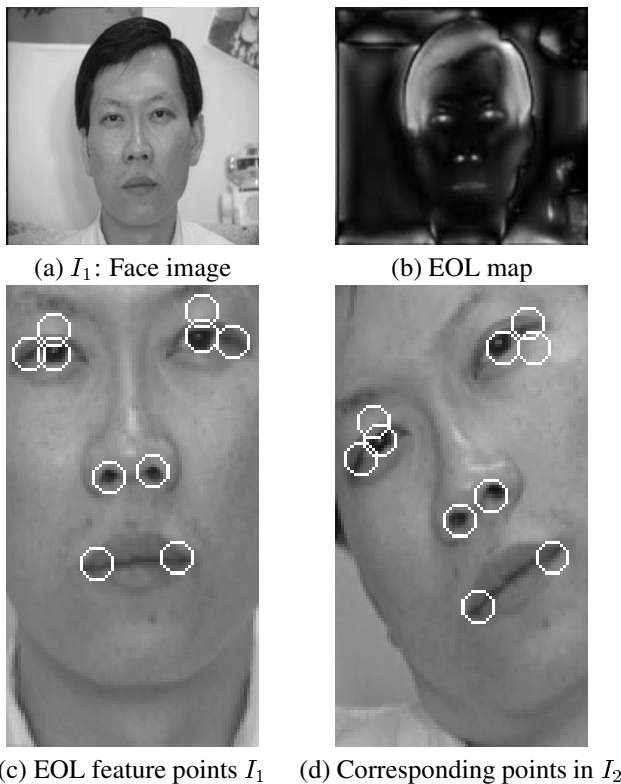
Figure 2. Feature selection and correspondence for a face image (a) using the EOL map (b). In (c) and (d), the EOL feature points are found at the center of the white circles overlaying I_1 and I_2.

evaluated in order of decreasing likelihood. Computational complexity of optimization was reduced by constraining the domain of $\mathbf{t}_i$ to instances t_i corresponding to peaks of the likelihood of $\mathbf{t}_i$.

In order to emphasize the generality of the approach, the strategy proposed in this paper was tested on a wide variety of image types. We tested face image correspondence using the MIT face database [19]. Figure 2 illustrates an EOL map (b) computed off-line on the face image depicted in (a). Note how the most informative facial feature points (shown in (c)) according the EOL agree with intuitive notions of interesting facial features, namely eyes, nostrils, etc. The generality of the approach lies in the fact that these points were chosen without the requirement of explicitly specifying a face model. The result of correspondence can be found in (d). A slight correspondence mismatch occurs on the left eye in (d) because the distance prior in (13) permits two symmetric solutions, above and below the eye.

3.3. Comparison of EOL and KLT Feature Selection

To demonstrate that the EOL can be used to select optimal points for image correspondence, we performed correspondence trials comparing feature points selected according to minimum EOL with selection according to the KLT criterion [17], which favors points with significant, non-uniformly oriented edge density. The goal of testing was to compare EOL and KLT feature point selection according correspondence optimization time. We expected EOL features would generally result in faster optimization. Feature points were selected as in Section 3.2: maps were computed off-line based on the KLT and EOL criteria, and the best 10 peaks (or valleys) were chosen for correspondence.

In terms of implementation, the KLT criterion required specification of 5 subjective parameters (i.e. local window size, minimum edge eigenvalue, etc.). We used the defaults as specified in [1]. The EOL criterion is defined entirely by the likelihood of the correspondence model in Section 3.1.1, and required no additional parameters.

The test set for included 35 image pairs from the VASC database [18] depicting a wide variety of scenes (e.g. objects, natural and synthetic scenes, aerial photographs, etc.). Images pairs consisted of different viewpoints of the same scene. In the interest of a fair comparison, we intentionally chose image pairs such that all feature points selected are visibly present in both images, i.e. we did not consider occluded features. Trials were deemed successful if a correspondence solution was found in less than 20 minutes. All images were scaled to 128×128 resolution for processing.

The time required to find T_{MAP} ranged from milliseconds to minutes, depending on the complexity of the scenes in the images. All correspondence solutions found appeared qualitatively plausible, although no ground truth comparison was performed. EOL feature points resulted in 33 successful solutions, as compared with 18 for KLT feature points. Figure 3 illustrates a plot comparing optimization time for each image pair. Optimization of EOL feature points was faster in 27 of the 33 solutions.

Of the 35 image pairs tested, 20 were relatively simple, uncluttered scenes such as the as one shown in Figure 4. For these image pairs, optimization time for EOL feature points was only marginally superior to KLT feature points. Intuitively, such images have relatively few feature points to choose from, and therefore fewer potential false matches. All 20 correspondence trials were successful for EOL features, verses 16 successful trials for KLT features. For 6 of the 20 trials (house1, toys, book, ball, fruit, Pepsi [18]), optimization was faster for KLT features.

The remaining 15 image pairs contained a larger degree of feature complexity and were therefore prime candidates for testing the two approaches in more difficult image matching contexts. EOL features outperformed KLT fea-
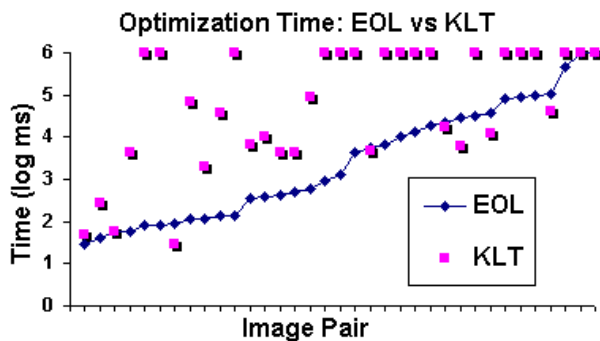
Figure 3. Comparison of EOL and KLT feature selection. Optimization time (log scale) is shown for a test set of 35 image pairs, sorted according to EOL optimization time. Points above the EOL curve represent image pairs for which optimization was slower for KLT feature points. When no correspondence was found after 20 minutes, time 6 is displayed.

tures in terms of optimization speed in all trials. Out of 15 trials, 13 were successful using EOL features, but only 2 using KLT features (See Figure 5 for example). Interestingly, in the 2 trials resulting in successful correspondence of KLT features, several feature points chosen according to the KLT criterion were the same as those chosen by the EOL criterion. Figure 6 depicts such a case.

It is not surprising that the EOL criterion outperforms the KLT criterion, since it is designed to identify feature points which are optimal for the particular correspondence model used. This is precisely the strength of the approach; if the model is modified to use a similarity metric specific to different local image characteristics, the EOL criterion will define a new set of optimal feature points to achieve fast, unambiguous correspondence. A subjective approach, such as the KLT feature selection, will always identify the same sub-optimal set of feature points.

4. Conclusions

In this paper, we introduced the EOL as a novel criterion for selecting optimal feature points for image correspondence. The criterion is optimal in the sense that it selects precisely the points which drive optimization towards a fast and unambiguous solution based on the correspondence model itself, not on subjective notions of feature point importance. We described the criterion in the context of a general Bayesian MRF model, which can be adapted to a wide variety of local image similarity measures and images. Experimental results show that feature points selected according to the EOL criterion outperform the well-known KLT criterion in terms of optimization time. Future work
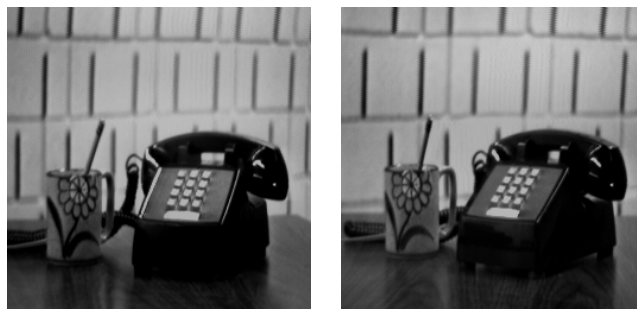
(a) Telephone: I_1 and I_2
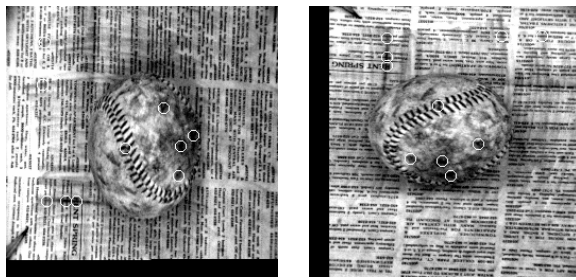

(b) EOL feature correspondence

(c) KLT feature correspondence

Figure 4. Telephone scene. In (b) and (c), EOL and KLT feature points are found at the center of the white circles seen overlaying I_1 and I_2. The default KLT parameters used were: window size = (7,7), smoothing = true, gradient sigma = 1.0, min eigenvalue = 1.

will involve further validation of the EOL criterion, including testing over a range of correspondence models and optimization strategies, and ground truth verification of correspondence accuracy.

References

[1] S. Birchfield. KLT: An Implementation of the Kanade-Lucas-Tomasi Feature Tracker, MIT media lab. http://vision.stanford.edu/ birch/klt.

[2] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models: Their training and application. *CVIU*, 61(1):38–59, January 1995.

[3] J. Coughlan, A. Yuille, C. English, and D. Snow. Efficient optimization of a deformable template using dynamic programming. In *CVPR98*, pages 747–752, 1998.

[4] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley & Sons, New York, 1991.

[5] G. Edwards, T. Cootes, and C. Taylor. Face recognition using active appearance models. In *ECCV98*, pages 581–595, 1998.
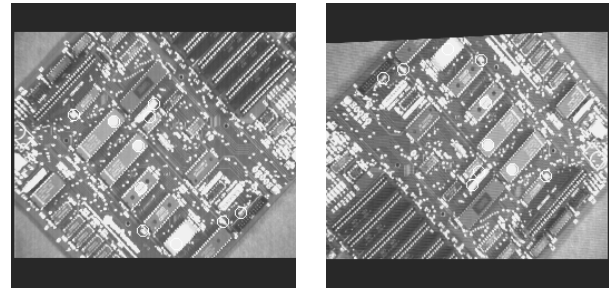
(a) Baseball: I_1 and I_2
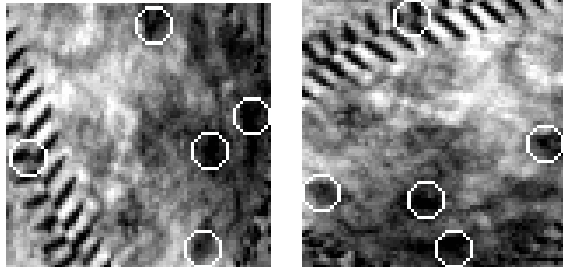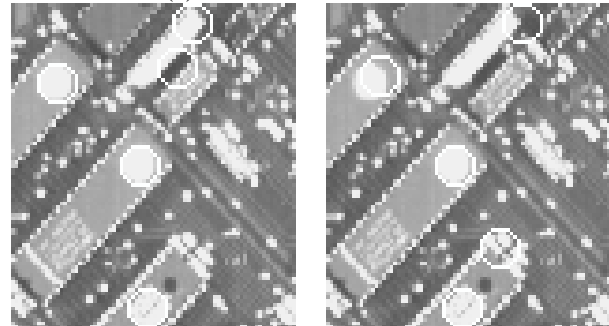

(b) EOL features: I_1 and correspondence in I_2.
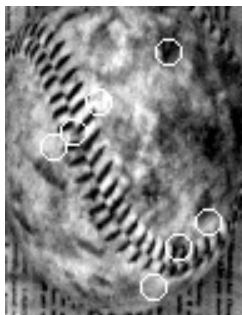
(c) KLT features: I_1, no correspondence found in I_2.

Figure 5. Highly textured baseball scene. EOL features were successfully matched in 80 ms. No solution was found in less than 20 minutes using KLT features.

(a) Circuit board: I_1 and I_2

(b) EOL feature points   (c) KLT feature points

Figure 6. Circuit board, one of two complicated correspondence trials where KLT feature points resulted in a correspondence solution. The best KLT features are mostly the same as those chosen by the EOL criterion.

[6] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(99):712–741, 1984.

[7] J. Hammersley and P. Clifford. Markov fields on finite graphs and lattices. Unpublished manuscript.

[8] A. Jain, Y. Zhong, and M. Dubuisson-Jolly. Deformable template models: A review. *SP*, 71(2), December 1998.

[9] A. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *Pattern Analysis and Machine Intelligence*, 18(3):267–278, March 1996.

[10] S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer-Verlag, 2001.

[11] D. Lowe. Object recognition from local scale-invariant features. In *ICCV99*, pages 1150–1157, 1999.

[12] S. A. Nene, S. K. Nayar, and H. Murase. Columbia object image library (coil-100). http://www.cs.columbia.edu/CAVE/.

[13] R. Nohre. Deformed template matching by the viterbi algorithm. *PRL*, 17(14):1423–1428, December 1996.

[14] J. Pearl. *Heuristics: intelligent search strategies for computer problem solving*. Addison-Wesley, 1984.

[15] S. Sclaroff and J. Isidoro. Active blobs. In *ICCV98*, pages 1146–1153, 1998.

[16] R. Szeliski. Bayesian modeling of uncertainty in low–level vision. *International Journal of Computer Vision*, 5(3):271–301, 1990.

[17] C. Tomasi and J. Shi. Good features to track. In *CVPR:94*, pages 593–600, 1994.

[18] Vision and Autonomous Systems Center. Stereo data, CMU. http://vasc.ri.cmu.edu/idb/html/stereo/.

[19] Vision and Modeling Group. Face database, MIT media lab. ftp://whitechapel.media.mit.edu/pub/images/.

[20] A. Yuille, D. Cohen, and P. Hallinan. Feature extraction from faces using deformable templates. In *CVPR89*, pages 104–109, 1989.