

Variational Maximum A Posteriori Model Similarity and Dissimilarity Matching

John Chiverton, Majid Mirmehdi,
Department of Computer Science,
University of Bristol, UK
{johnc,majid}@cs.bris.ac.uk

Xianghua Xie
Department of Computer Science,
University of Wales Swansea, UK
x.xie@swansea.ac.uk

Abstract

A new variational Maximum A Posteriori (MAP) contextual modeling approach is presented that minimizes the product of two ratios: (a) the ratio of the model distribution to the distribution of currently estimated foreground pixels; (b) the ratio of the background distribution to the model distribution for all estimated background pixels. This approach provides robust discrimination to identify the division between foreground and background pixels, which is useful for applications such as object tracking.

1 Introduction

Image based photometric information is commonly used in image segmentation. However, many region based segmentation approaches often do not provide an accurate definition of the bounds of the region of interest, e.g. [1]. Those that do may only rely on simplifying image assumptions, e.g. [6, 2], or an effective competition between the foreground and background regions, e.g. [7, 9]. In this paper, we propose a MAP solution to the matching process (in contrast to [8]), where the prior information includes a regularity condition imposed via the gradient magnitude of the labelling function and a model probabilistic definition of the region of interest. The formulation described here combines both model to foreground *matching* and model to background *mismatching*. This can be compared with existing probabilistic region based modeling approaches, many of which only rely on a type of competition between the foreground and background. These depend on the region of interest possessing sufficiently different photometric information in comparison to the neighboring structures, which otherwise can lead to the foreground region leaking into background regions, or vice

versa. Further information is therefore often introduced to constrain the identified region such as image derived edge based energy terms, e.g. [7]. Hence in contrast to such approaches, we advocate the use of foreground matching and background mismatching using a stable probabilistic description to reduce the possibility of leakage or retention from one region to another, without resorting to dedicated generalization of object shape, such as in the active appearance model [4].

2 Methodology

An image is considered here to be composed of pixels \mathbf{x} labeled as either foreground $\mathcal{F} = \{\forall \mathbf{x} : f(\mathbf{x}) = 1\}$ or background $\mathcal{B} = \{\forall \mathbf{x} : b(\mathbf{x}) = 1 - f(\mathbf{x})\}$, where $f(\mathbf{x}) \in \{0, 1\}$. The set of pixels in the entire image space is given by $\Omega = \mathcal{F} \cup \mathcal{B}$. Pixels in the foreground and background regions are associated with photometric image information, $I_{\mathbf{x}}$. The foreground *a posteriori* probability is given by $P(f|I_f) = \frac{p(I_f|f)P(f)}{p(I_f)}$, where $p(I_f|f)$ is the foreground data likelihood, $P(f)$ is the foreground prior probability, $p(I_f)$ is the marginal data PDF and $I_f = \{I_{\mathbf{x}}|f(\mathbf{x})\}$. The *a posteriori* probability $P(b|I_b)$ for the background region can be similarly defined.

Statistical techniques, such as maximum likelihood or MAP are often used to maximize the *a posteriori* probabilities via an optimization process to obtain an optimal classification and hence division of the image space into foreground and background regions, e.g. [7]. Prior knowledge can be introduced into a probabilistic model somewhat indirectly by specifying the form of prior distributions. Alternatively, a more direct way of introducing prior knowledge is via a reference foreground data likelihood, $p(I|m)$. This likelihood term can be obtained via an *a posteriori* model probability, $P(m|I) = \frac{p(I|m)P(m)}{p(I)}$, where m is a hypothetical model labeling. The foreground region in the image will hypo-



Figure 1. Example use of proposed model similarity and dissimilarity matching method in object tracking for varying amounts of noise (tracking framework from the authors' work in [3]).

thetically exist, i.e. $\mathfrak{M} = \{\forall \mathbf{x} : m(\mathbf{x}) = 1\}$, that is obtained if $P(f|I_f)$ has been optimized for the data at hand. This is of course also dependent on the availability and selection of appropriate features, the probabilistic modeling and optimization. However, we do not seek to utilize \mathfrak{M} directly. Instead, we assume the availability of the model likelihood, $p(I|m)$ which is possible from the result of prior processing (e.g. object tracking) or supervised training. Fig. 1 gives an example where the bootstrapping stage of the object tracking application provides a good estimation of the foreground object, whose photometric model can be fed to the method proposed in this paper to estimate the location of the object in the following frames to, in turn, further assist object tracking.

Our optimization process will determine a foreground that possesses a very similar *a posteriori* probability in comparison to our hypothetical model. This similarity, $R_{f,m}$, can be quantified via a logarithmic probability ratio test of the *a posteriori* probabilities for the foreground and the hypothetical model:

$$R_{f,m} = -\ln \frac{P(m|I_f)}{P(f|I_f)} = -\ln \frac{p(I_f|m)P(m)}{p(I_f|f)P(f)} \quad (1)$$

We can also define a similar but reciprocal test for the background in comparison to the model, $\bar{R}_{b,m}$.

$$\bar{R}_{b,m} = -\ln \frac{P(b|I_b)}{P(m|I_b)} = -\ln \frac{p(I_b|b)P(b)}{p(I_b|m)P(m)}. \quad (2)$$

This will quantify the dissimilarity of the model with the background. We show later this dissimilarity measure is very useful in accurately estimating the foreground. Minimization of (1) and (2), i.e. $\arg \min_{\mathfrak{f}, \mathfrak{B}} (R_{f,m} + \bar{R}_{b,m})$, will produce two mutually exclusive regions. The foreground and background priors, $P(f)$ and $P(b)$ can be improper or they can be used to impose some regularity condition in the image space. However, the model prior $P(m)$ is considered constant throughout the image space and therefore not dependent on the location. Hence, the minimization problem can

be formulated as:

$$\arg \min_{\mathfrak{f}, \mathfrak{B}} \left(-\int_{\mathfrak{f}} \ln \frac{p(I_{\mathbf{x}}|m)}{p(I_{\mathbf{x}}|f)} d\mathbf{x} - \int_{\mathfrak{B}} \ln \frac{p(I_{\mathbf{x}}|b)}{p(I_{\mathbf{x}}|m)} d\mathbf{x} + \int_{\mathfrak{f}} \ln P(f) d\mathbf{x} - \int_{\mathfrak{B}} \ln P(b) d\mathbf{x} \right). \quad (3)$$

This expression optimizes the model similarity with the estimated foreground and its dissimilarity with the estimated background.

Minimization of (3) can be accomplished via gradient descent in a variational framework. The method used here treats the entire image globally. Hence, both the foreground and background likelihood ratio integral terms are now extended to the entire image space, Ω . Noting that $\int_{\mathfrak{f}} \ln(\cdot) \mathbf{x} = \int_{\Omega} f \ln(\cdot) \mathbf{x}$, $\int_{\mathfrak{B}} \ln(\cdot) \mathbf{x} = \int_{\Omega} b \ln(\cdot) \mathbf{x}$, and $b = 1 - f$, then (3) is equivalent to

$$\arg \min_{\mathfrak{f}, \mathfrak{B}} \left(-\int_{\Omega} \left(f \ln \frac{p(I_{\mathbf{x}}|m)}{p(I_{\mathbf{x}}|f)} + (1-f) \ln \frac{p(I_{\mathbf{x}}|b)}{p(I_{\mathbf{x}}|m)} - f \ln P(f) + (1-f) \ln P(b) \right) d\mathbf{x} \right). \quad (4)$$

A condition that the foreground and background are considered optimally defined is if they are spatially consistent, which is equivalent to obtaining a classification with minimal spatial gradient magnitude of the labeling function across the entire image space. Therefore, minimizing the last two terms in (4) is considered here to be equivalent to (by analogy) $\arg \min_{\mathfrak{f}, \mathfrak{B}} (\int_{\Omega} |\nabla f| d\mathbf{x})$. Thus we may seek to minimize

$$-\int_{\Omega} \left(f \ln \frac{p(I_{\mathbf{x}}|m)}{p(I_{\mathbf{x}}|f)} + (1-f) \ln \frac{p(I_{\mathbf{x}}|b)}{p(I_{\mathbf{x}}|m)} - |\nabla f| \right) d\mathbf{x}. \quad (5)$$

The action of (5) combines three global and simultaneous minimizing actions. The first two are the minimization of the logarithmic difference between (a) the model PDF $p(I_{\mathbf{x}}|m)$ and the foreground PDF $p(I_{\mathbf{x}}|f)$



Figure 2. Convergent results obtained for poor initial conditions using proposed model.

for the currently defined foreground pixels f , and (b) the background PDF $p(I_{\mathbf{x}}|b)$ and the model PDF for the currently defined background pixels b , and the third imposes smoothness throughout the image space by minimizing the foreground labeling gradient magnitude.

Photometric representation The photometric information in the foreground and background regions are described by the histograms $\mathcal{H}_f(I(\mathbf{x} \in \mathfrak{F}))$ and $\mathcal{H}_b(I(\mathbf{x} \in \mathfrak{B}))$ respectively. This histogram information is then used to approximate the PDFs, $p(I_{\mathbf{x}}|f)$ and $p(I_{\mathbf{x}}|b)$. The probability space is modeled here with a finite number of Gaussian components with parameters that are estimated via Expectation Maximization. It is worth noting that other statistics, such as filtering responses, can also be included in the model.

Minimization To solve (5), we first make a simple substitution of the labeling function, $f(\mathbf{x}) \in \{0, 1\}$, with a slightly regularized Heaviside function $H_\epsilon(\cdot)$ in combination with a signed distance function $\phi(\mathbf{x})$, similar to [2]. The signed distance function is Lipschitz continuous and implicitly encodes the boundary between the foreground and background. Thus, the minimization problem is now

$$\arg \min_{\mathfrak{F}, \mathfrak{B}} \left(- \int_{\Omega} \left(H_\epsilon(\phi) \ln \frac{p(I_{\mathbf{x}}|m)}{p(I_{\mathbf{x}}|f)} \right) \right. \\ \left. + (1 - H_\epsilon(\phi)) \ln \frac{p(I_{\mathbf{x}}|b)}{p(I_{\mathbf{x}}|m)} - |\nabla H_\epsilon(\phi)| \right) d\mathbf{x}. \quad (6)$$

The solution to (6) can be found via gradient descent using the derivative of the functional, equivalent to

$$\frac{\partial \phi}{\partial t} = -\delta_\epsilon(\phi) \left(\ln \frac{p(I_{\mathbf{x}}|m)}{p(I_{\mathbf{x}}|f)} + \ln \frac{p(I_{\mathbf{x}}|m)}{p(I_{\mathbf{x}}|b)} - \kappa \right), \quad (7)$$

where κ is the curvature and $\delta_\epsilon(\cdot)$ denotes a slightly regularized Dirac delta function, ([2]).

Model interpretation An interpretation of (7) can be considered via a number of possible solution states. Table 1 summarizes three possible solution states: (a) a good solution state where the currently defined foreground is in good agreement with the hypothetical

Table 1. Intuitive simplified summary of possible solution states. H - high; M - medium; L - low.

	Solution state					
	Good		Medium		Bad	
Model Var.	f	b	f	b	f	b
P_m	H	L	H	L	H	L
P_f	H	L	M	M	L	H
P_b	L	H	M	M	H	L
$P_m - P_f$	0	0	+1	-1	+2	-2
$P_m - P_b$	+2	-2	+1	-1	0	0
<i>Combined</i>	+2	-2	+2	-2	+2	-2

model, (b) a medium solution state where the currently defined foreground is in moderate agreement with the hypothetical model, and (c) a bad solution state where the currently defined background has more in common with the hypothetical model than the current foreground. The bottom three rows of Table 1 illustrate that the combination of foreground matching and background mismatching in (7) provides a consistent cost function. A good solution state will heavily penalize ($\approx \times 2$) background pixels in the foreground region and heavily penalize ($\approx \times 2$) foreground pixels in the background region. As the state changes, so do the costs.

This can be compared with approaches that only rely on foreground matching (cf. row $P_m - P_f$ in Table 1), see e.g. [5], and similarly for methods that only perform model foreground and background comparison (cf. row $P_m - P_b$ in Table 1). Combining these two brings extra enforcement, which is demonstrated in the next section.

3 Results

The importance of the properties summarized in Table 1 are now illustrated experimentally. The model likelihood, $p(I_{\mathbf{x}}|m)$ was defined here, for the purposes of our work via a manually defined model region $\mathbf{x} \in \mathfrak{M}$ on the test images. This region can be defined automatically e.g. via the outline of the object in a preceding frame for tracking applications, e.g. see [3]. The results in Figs. 2 and 3 illustrate the performance of the

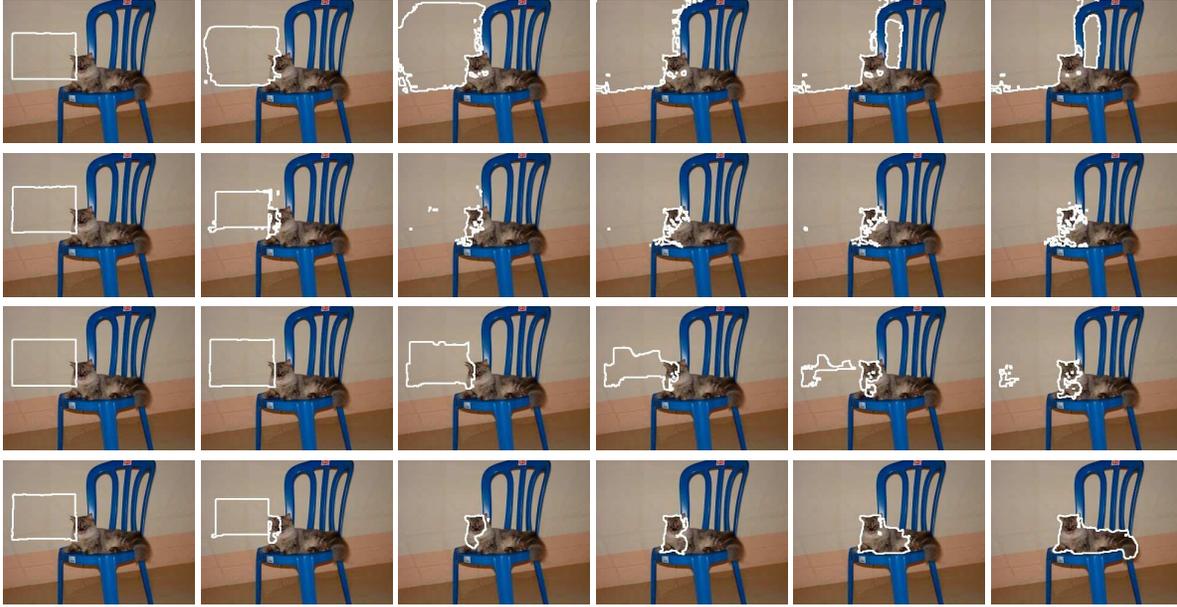


Figure 3. Comparative performance of different components of (7). 1st Row: foreground background competition with context (see e.g. [7]); 2nd row: foreground similarity; 3rd row: background dissimilarity; 4th row: foreground similarity and background dissimilarity with context.

proposed approach for difficult initial conditions which may arise in e.g. a tracking application, where the region of interest has moved significantly between frames (cf. Fig. 1). It is important to note that the initial contour has to *touch* the region of interest. This is highly likely for tracking applications and furthermore is of considerable interest when an image may possess more than one region with similar photometric properties.

The relative performance of the model proposed here in relation to just foreground matching, background mismatching or a competition between the active foreground and background are also illustrated in Fig. 3. The first row provides the result of foreground background competition without prior model m , e.g. [7]. It performed badly due to difficult initialization and lack of prior knowledge. The next two rows show the results using foreground matching alone and background mismatching alone, respectively. Foreground similarity measure removes false positives, while background dissimilarity comparison increases true positives. However, working individually, neither of them produced reasonable results. The proposed method which combines these two actions provides the best results across our range of experiments, e.g. last row of Fig. 3.

4 Conclusions

An approach that applies both foreground matching and background mismatching in combination with a MAP solution was proposed. The performance of the method in relation to other potential model configura-

tions has also been illustrated, demonstrating its unique potential for use in a wide variety of applications.

Acknowledgements

This research was supported by Leverhulme Trust UK.

References

- [1] G. Bradski. Computer vision face tracking as a component of a perceptual user interface. In *Workshop on Applications of Computer Vision*, pages 214–219, 1998.
- [2] T.F. Chan and L.A. Vese. Active contours without edges. *IEEE IP*, 10(2):266–277, 2001.
- [3] J. Chiveron, X. Xie, and M. Mirmehdi. Tracking with active contours using dynamically updated shape information. In *BMVC*, 2008. Accepted for publication.
- [4] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE PAMI*, 23(6):681–685, 2001.
- [5] D. Freedman and T. Zhang. Active contours for tracking distributions. *IEEE IP*, 13(4):518–526, 2004.
- [6] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *CPAM*, 27(5):577–685, 1989.
- [7] N. Paragios and R. Deriche. Geodesic active regions: a new framework to deal with frame partition problems in computer vision. *JVCIR*, 13(1-2):249–268, 2002.
- [8] T. Zhang and D. Freedman. Improving performance of distribution tracking through background mismatch. *IEEE PAMI*, 27(2):282–287, 2005.
- [9] S. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE PAMI*, 18(9):884–900, 1996.