

# Feature-level Fusion for Object Segmentation using Mutual Information

Vinay Sharma and James W. Davis

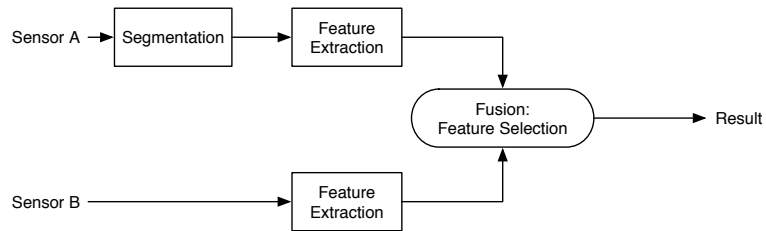
Dept. of Computer Science and Engineering  
Ohio State University  
Columbus OH 43210 USA  
{sharmav,jwdavis}@cse.ohio-state.edu

**Abstract.** *In this chapter, a new feature-level image fusion technique for object segmentation is presented. The proposed technique approaches fusion as a feature selection problem, utilizing a selection criterion based on mutual information. Starting with object regions roughly detected from one sensor, the proposed technique aims to extract relevant information from another sensor in order to best complete the object segmentation. First, a contour-based feature representation is presented that implicitly captures object shape. The notion of relevance across sensor modalities is then defined using mutual information computed based on the affinity between contour features. Finally a heuristic selection scheme is proposed to identify the set of contour features having the highest mutual information with the input object regions. The approach works directly from the input image pair without relying on a training phase. The proposed algorithm is evaluated using a typical surveillance setting. Quantitative results, and comparative analysis with other potential fusion methods are presented.*

## 1 Introduction

Vision applications, such as video surveillance and automatic target recognition, are increasingly making use of imaging sensors of different modality. The expectation is that a set of such sensors would benefit the system in two ways; first, the complementary nature of the sensors will result in increased capability, and second, the redundancy among the sensors will improve robustness. The challenge in image fusion is thus combining information from the images produced by the constituent sensors to maximize the performance benefits over using either sensor individually.

In order to better quantify the performance benefits, and to enable the use of fusion algorithms in automatic vision systems, we adopt in our work a more “goal-oriented” view of image fusion than is traditionally used. Instead of merely improving the context (or information) present in a scene, we focus on the specific task of using image fusion to improve the estimation of the object shape (as defined by a silhouette, or a boundary).



**Fig. 1.** Flowchart of proposed fusion method.

We propose a novel solution to this task, one that approaches fusion as essentially a feature-selection problem. Our approach is applicable to any combination of imaging sensors, provided that the sensors are co-located and registered. The processing pipeline and the main computation stages of our algorithm are shown in the flowchart of Fig. 1. As can be seen from the flowchart, the feature extraction stage of our algorithm is preceded by an object segmentation routine employed in only one of the input sensors (denoted by sensor *A*). The object segmentation routine is used only to bootstrap the feature selection process, and hence any method that provides even a rough, incomplete object segmentation can be employed at this stage. Contour-based features are then extracted from the rough object segmentation results of sensor *A* and from the corresponding image region of sensor *B*. These features are then used within a mutual information framework so as to extract a subset of features from sensor *B* that are most *relevant* (complementary and redundant) to the features obtained from the initial segmentation of sensor *A*.

In order to estimate the probability distribution required for computing the mutual information across sets of features, we present a method that relies on the regularities in shape and form found in most objects of interest. We extend the notion of affinity, originally defined to measure the smoothness of the curve joining two edge elements [37], to our contour features. Using this affinity measure, we formulate conditional probability distributions of contour features from sensor *A* with respect to sensor *B*. We then compute the mutual information between contour features from the two sensors based on these conditional distributions. Then we identify the set of contour features from *B* that maximize the mutual information with the features from *A*. The contours from sensor *A* overlaid with the selected contours from sensor *B* form the fused result, which can then be completed and filled to create silhouettes.

Image fusion algorithms can be broadly classified into low-, mid-, and high-level techniques based on their position in the information processing pipeline. As can be seen from Fig. 1, based on such a classification, the proposed algorithm can be categorized as a goal-oriented, mid-level fusion technique relying on contour-based features. Apart from the proposed algorithm, there have been several other approaches adopted for the purpose of fusing information across two imaging sensors. Next, we briefly outline two of the other popular fusion

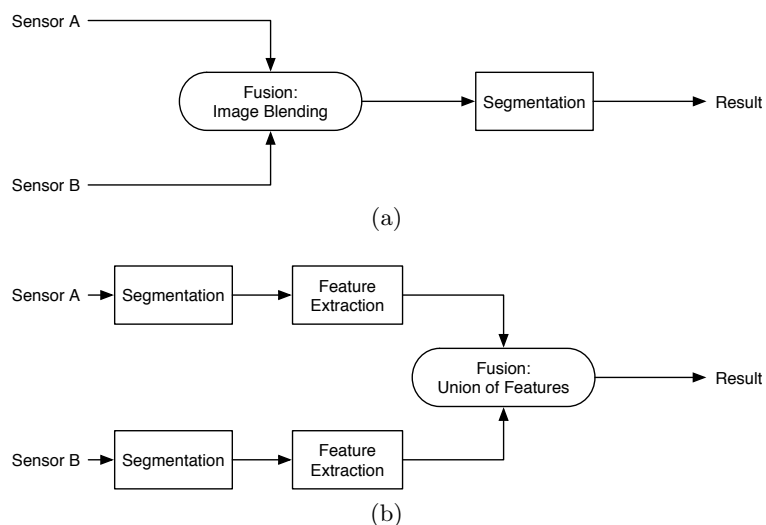
strategies that could potentially be employed with the specific aim to provide robust and accurate object segmentation.

### 1.1 Alternate Fusion Methodologies

**Image blending:** Perhaps the most commonly used approach for fusion is to first to create a single, fused image stream by blending images from each of the sensors [7, 34, 19]. Object location and shape is then obtained by applying relevant detection and segmentation algorithms to the fused image stream. In Fig. 2(a) we show the main processing stages typically employed by algorithms adopting this strategy. Such approaches have two potential drawbacks. First, since the fusion procedure simply blends image pairs at a low-level (pixel-level), the fused image is likely to contain unwanted image artifacts found in each of the two imaging domains. For example, when combining images from the thermal and visible domains, the resulting fused image can contain both shadows (from the visible domain) and thermal halos [9] (found in thermal imagery). Second, since the image characteristics of the fused stream depend on the fusion technique applied, such approaches typically require specialized object segmentation algorithms in order to obtain satisfactory results.

**Union of features:** Another popular approach is to defer the fusion of information to a later stage in the pipeline. Such approaches typically employ complete image segmentation routines in *each* image stream and then combine the segmentation results across the sensing domains. These techniques can either be high-level or mid-level, depending on whether the fusion occurs at the decision level or at the feature level. In decision level fusion, binary segmentation results are obtained independently in each domain and the final result is obtained by combining the individual silhouettes [36]. In feature level fusion, features extracted from the individual segmentation results are utilized to generate the final result. In this work, we employ for comparison a feature level fusion approach [10]. The flowchart of this method is shown in Figure 2(b). As can be seen from the figure, such an approach has the undesirable property of requiring object segmentation routines to be employed independently in the different imaging domains. Since the features are extracted from the segmentation provided in each sensor, the combined result is obtained by simply performing a union of all the features obtained from both sensors. Thus the final result is susceptible to errors in segmentation from both domains.

Based on the described flowcharts (Figs. 1, 2), we note that the proposed approach has the potential to provide clear benefits over these alternate methods. Being a feature-level technique, it does not face the issues that hamper low-level image blending techniques. Additionally, the proposed approach completely decouples the process of segmentation from fusion. Thus any “off-the-shelf” segmentation routine could be used in any one sensor to bootstrap the process. Compared to the feature-union method, the proposed approach provides the obvious benefit of requiring object segmentation in only one sensor. Further, given that the imaging domains employed in fusion systems are generally complementary in nature, different segmentation algorithms are likely to be effective in each



**Fig. 2.** Flowcharts of alternate fusion methods (a) Image blending. (b) Union of features.

domain. While the final result in the feature-union technique will be limited by the worse of the two segmentation results, the proposed fusion technique enables the user to employ only the better of the two segmentation results to bootstrap the fusion process. Depending on the application, factors such as persistence, signal-to-noise ratio, and the availability and complexity of the segmentation scheme can influence which sensor should be chosen for providing the initial segmentation.

## 1.2 Outline

We begin by reviewing in Sect. 2 related work in image fusion. We then present the different aspects of the proposed algorithm, starting with Sect. 3, where we describe the contour-based features used in the algorithm. In Sect. 4 we describe the process of computing the mutual information between different sets of contour features, and in Sect. 5 we describe the contour selection procedure. Then, in Sect. 6, we demonstrate the approach for a video surveillance application using a thermal and color camera as the two input sensors. Based on manually segmented object regions we show the efficacy of the proposed method by comparing segmentation performance using the fusion algorithm over using either input sensor independently. We also compare our algorithm against the two alternate fusion methods introduced earlier, and discuss the advantages of our approach over these other methods. Finally, in Sect. 7 we summarize our approach and provide directions for future work.

## 2 Related Work

Image fusion techniques have had a long history in computer-vision and visualization. We categorize related work into three types, based on the processing level (low, mid, high) at which fusion is performed.

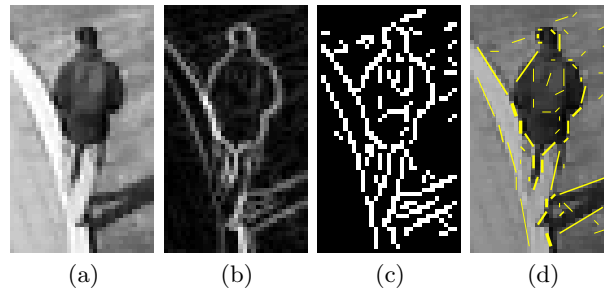
Traditionally, *low-level* techniques have been used to combine information from co-registered multi-sensor imagery. Improving upon simple techniques such as pixel averaging, multi-resolution schemes similar to the pyramid-based approaches of [34, 23, 4] were proposed. More recently, wavelet analysis has emerged as the method of choice in most multi-resolution frameworks [18, 27]. Examples of other low-level techniques include the biologically motivated model based on human opponent color processing proposed in [13]. A PCA-based technique measuring pixel variances in local neighborhoods is used in [7]. Pixel-level combinations of spatial interest images using Boolean and fuzzy-logic operators are proposed in [11], and a neural networks model for pixel-level classification is used in [16].

*Mid-level* fusion techniques have mostly relied on first and second order gradient information. Some of these techniques include directly combining gradients [28], determining gradients in high dimensions [33], and analyzing gradients at multiple resolutions [30, 25]. Other features, such as the texture arrays [3], have also been employed. Model-based alternatives to feature-level fusion have also been proposed such as the adaptive model matching approach of [12], and the model-theory approach of [38]. Other mid-level fusion techniques such as the region-based methods of [17, 39, 26] make use of low-level interactions of the input domains.

*High-level* fusion techniques generally make use of Boolean operators or other heuristic scores (maximum vote, weighted voting, m-of-n votes) [8, 36] to combine results obtained from independently processing the input channels. Other “soft” decision techniques include Bayesian inference [1, 14] and the Dempster-Shafer method [2, 20].

Most of these fusion techniques aim at enhancing the information content of the scene, to ease and improve human interpretation (visual analysis). However, the method we propose is designed specifically to enhance the capabilities of an automatic vision-based detection system. Some techniques such as [11, 12, 3], proposed for Automatic Target Recognition systems, have also been evaluated in terms of object detection performance. These techniques however are not generally applicable to the detection of non-rigid person shapes, and other large, multi-modal objects common in the urban environments considered in this work. Other techniques, such as [13], have been shown to improve recognition performance when used as inputs to separate target recognition modules.

Recently [10] proposed a mid-level fusion algorithm also designed with the aim of improving object segmentation. However, their fusion technique was specific to a thermal and color camera, and required background modeling in *both* domains (see Fig. 2(b)). Our proposed method is an improvement on both counts. First, the current algorithm is independent of the methods used for detecting the initial object regions. Thus as long as the required contour features can be



**Fig. 3.** Feature extraction. (a) Input image. (b) Gradient magnitudes. (c) Thinned binarized edges. (d) Extracted contour features overlaid on input image.

reliably extracted from the images produced by both sensors, the fusion procedure is unaffected by the modality of the sensors (far infrared, long wavelength infrared, short wavelength infrared, etc.). Second, the current algorithm only requires the prior ability (via method of choice) to detect object features in any *one* sensor modality.

### 3 Contour Features

Based only on the preliminary object segmentation obtained from sensor  $A$ , our goal is to be able to extract relevant information from sensor  $B$ , such that the combined result is a better estimation of the object shape. The crucial step in this process is choosing the appropriate features. The importance of first-order gradient information in estimating the shape and appearance of an object is well known [6, 21]. We exploit this information by extracting features that capture the location, orientation, and magnitude of the object gradients.

We first obtain a thinned representation of the gradient magnitude image using a standard non-maximum suppression algorithm. The thinned edges are then broken into short, nearly linear contour fragments based on changes in the gradient direction. A contour fragment is obtained by traversing along a thinned edge using a connected-components algorithm until a change from the initial edge orientation is encountered. To ensure contour fragments of reasonable size, the edge orientations are initially quantized into a smaller number of bins. We represent a contour fragment by a feature vector  $c = [p_1, p_2, E_{mag}]$ , where  $p_1$  and  $p_2$  are the coordinates of the two end-points, and  $E_{mag}$  is the mean edge magnitude along the contour. The set of all contour features  $\{c_1, \dots, c_n\}$  forms the feature representation of the object.

In Fig. 3 we show an example of the feature extraction process. Figure 3(a) shows the input image, and figures 3(b) and (c) show the gradient magnitudes and the thinned binary edges respectively. The short, linear contour features extracted for this examples are shown overlaid on the input image in Fig. 3(d).

The features were obtained by quantizing the edge orientations into 4 equal sized bins.

## 4 Estimating Feature Relevance

Having extracted the contour features, our goal now is to select features from sensor  $B$  that are relevant to the features in sensor  $A$ . Mutual information is considered to be a good indicator of the relevance of two random variables [5]. This ability to capture the dependence, or relevance, between random variables has recently led to several attempts at employing mutual information in feature selection schemes [15, 24, 35].

### 4.1 Preliminaries

Denoting two discrete random variables by  $X$  and  $Y$ , their mutual information can be defined in terms of their probability density functions (pdfs)  $p(x)$ ,  $p(y)$ , and  $p(x, y)$  as

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (1)$$

Based on entropy, the mutual information between  $X$  and  $Y$  can also be expressed using the conditional probability  $p(x|y)$ . The entropy,  $H$ , of  $X$  is a measure of its randomness and is defined as  $H(X) = -\sum_{x \in X} p(x) \log p(x)$ . Given two variables, conditional entropy is a measure of the randomness when one of them is known. The conditional entropy of  $X$  and  $Y$  can be expressed as

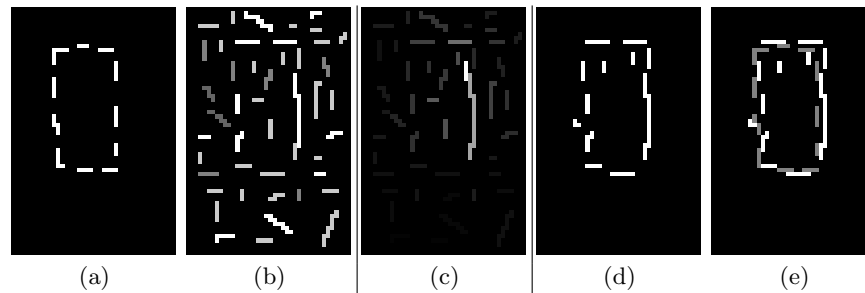
$$H(X|Y) = -\sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log p(x|y) \quad (2)$$

The mutual information between  $X$  and  $Y$  can be computed from the entropy terms defined above by

$$I(X; Y) = H(X) - H(X|Y) \quad (3)$$

Let us associate random variables  $S_1$  and  $S_2$  with the sensors  $A$  and  $B$  respectively. Let  $C_1$  denote the domain of  $S_1$ , and  $C_2$  the domain of  $S_2$ . In order to use either Eqn. 1 or Eqn. 3 to compute the mutual information between  $S_1$  and  $S_2$  we first need to define the domains,  $C_1$  and  $C_2$ , and then estimate the appropriate probability distribution functions. A discretized version of the full contour feature space of  $A$ , and similarly of  $B$ , are natural choices for  $C_1$  and  $C_2$  respectively. In general, obtaining the pdfs, especially the joint and the conditionals, of the contour features  $c_i \in C_1$  and  $c_j \in C_2$  is a difficult task. Indeed, it is this difficulty that primarily impedes the use of mutual information in feature selection schemes [35, 24].

Nevertheless, a typical approach would be to estimate these distributions using a large training data set consisting of manually segmented objects imaged using the sensors in question. The difficulty of generating such a data set



**Fig. 4.** Toy example illustrating the relevant processing stages. (a) Detected object contours from sensor  $A$ . (b) Contours obtained from sensor  $B$ . (c) Relative affinity values of contours in (b) with respect to a contour (shown in white) from (a). (d) Set of contours selected from (b). (e) Overlay of contours from (a), shown in gray, with the selected contours (d).

aside, such an approach has several drawbacks. Importantly, different pdfs will need to be estimated for different object classes in the training set, and there is no guarantee that these would generalize well for novel objects. This is especially cumbersome given the enormous computation and memory requirements of non-parametric estimation techniques. Further, the well-known issue of scale (bandwidth) selection [22] in these methods becomes compounded in high dimensional spaces such as ours.

Instead of relying on a training data-set to learn the distributions of features, we propose a different approach to the problem. Drawing on the observation regarding the “natural structure” of the world, we make the assumption that objects of interest have continuous, regular boundaries. Based on this assumption, we seek to define relationships between samples from  $S_1$  and  $S_2$  that will enable us to identify the set of contours from sensor  $B$  with the highest relevance to sensor  $A$ .

In the context of fusion, we propose that a set of features has high relevance to another, if it provides both redundant *and* complementary information. The choice of contour features (Sect. 3) enables us to further define relevance as the ability of a set of features to coincide with, and complete object boundaries that have been only partially captured by another set. We now address the issue of computing contour feature relevance and folding it into a mutual information framework.

## 4.2 Contour Affinity

Assume that the pair of images shown in Fig. 4(a) and (b) represent the contour features of a rectangular box imaged using two sensors. Let Fig. 4(a) represent the object contours from sensor  $A$ , and Fig. 4(b) the set of contours obtained from sensor  $B$ . Note that the contour features from sensor  $A$  are obtained after an initial segmentation step, and hence lie along the boundary of the rectangular



box. On the other hand, the feature extraction from sensor  $B$  is not preceded by any object segmentation and hence the features are extracted directly from the entire image region.

Visualizing the contour features extracted from sensor  $A$  in image-space, as in Fig. 4(a), we see that the contour fragments form an incomplete trace of the boundary of the viewed object. As described earlier, we desire the subset of contour features from sensor  $B$  that provides the best completion of the broken contour image formed by the features from sensor  $A$ .

Perceptual (and computational) figure completion is a very active field of research, and several figure completion studies, such as [37, 32], have used an “affinity” measure between a pair of edge elements to compute how likely it is that they belong to the same underlying edge structure. The concept of affinity is related to the energy of the simplest curve passing through two edge elements. One popular method of determining the lowest energy (smoothest) completion between two edge elements is minimizing the *elastica* functional, which defines energy as the integral of the square of curvature along the curve. We borrow this notion of affinity and adapt it to deal with contours of finite size instead of the dimension-less edge elements used in the literature.

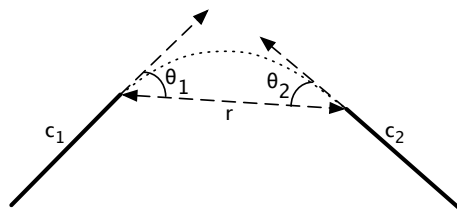
Consider a pair of contours  $c_1$  and  $c_2$ , as shown in Fig. 5, and hypothesize the simplest curve connecting  $c_1$  and  $c_2$ . Any such connection would join one of the end-points of  $c_1$  to an end-point of  $c_2$ . Since  $c_1$  and  $c_2$  have two end-points each, all such curves fall into one of four categories based on which two end-points are connected. Consider one such curve, shown by the dotted line in Fig. 5, between an end-point of  $c_1$  and an end-point of  $c_2$ . Further, consider the vector joining the ends of the curve, pointing from the end-point of  $c_1$  to the end-point of  $c_2$ . As shown in Fig. 5, let  $\theta_1$  denote the angle between this vector and the unit vector at the end-point of  $c_1$ , directed away from the contour along the tangent at  $c_1$ . Let  $\theta_2$  denote the angle from  $c_2$ , analogous to  $\theta_1$ . Finally, let  $r$  denote the Euclidean distance between the two end-points of  $c_1$  and  $c_2$ . These quantities,  $\theta_1$ ,  $\theta_2$ , and  $r$  are computed for each of the four possible sets of curves between end-points of  $c_1$  and  $c_2$ .

We define the contour affinity,  $Aff(c_1, c_2)$ , between two contours  $c_1$  and  $c_2$  as the maximum affinity value over the four possible sets of curves. Following the analytical simplification for the minimization of the *elastica* functional presented in [32], the affinity for a particular curve set is defined as

$$\mathcal{A} = e^{(-r/\sigma_r)} \cdot e^{(-\beta/\sigma_t)} \cdot e^{(-\Delta/\sigma_e)} \quad (4)$$

where  $\beta = \theta_1^2 + \theta_2^2 - \theta_1 \cdot \theta_2$  and  $\Delta = |E_{mag}^{c_1} - E_{mag}^{c_2}|$  (the absolute difference in the intensity of the contours). We write the normalization factors  $\sigma_r$ ,  $\sigma_t$ , and  $\sigma_e$  as  $\sigma_r = R/f_1$ ,  $\sigma_t = T/f_2$ , and  $\sigma_e = E/f_3$ , where  $R$ ,  $T$ , and  $E$  equal the maximum possible value of  $r$ ,  $\beta$ , and  $\Delta$ , and  $(f_1, f_2, f_3)$  are weights that can be used to change the relative influence of each term in the affinity calculation.

Contour pairs that are in close proximity, lie along a smooth curve, and have comparable intensities will have high affinity values. Consider the pair-wise affinity measurements between contour features taken one at a time from  $C_2$ , and



**Fig. 5.** Computation of contour affinity.

the set of contour features  $C_1$ . If a particular contour feature  $c_2 \in C_2$  lies along the object boundary, it would have very high affinity values with neighboring contours features in  $C_1$ . If  $c_2$  represents a non-object contour (e.g., background edge), unless it is in close proximity to some object contour, aligns well with it, *and* has similar intensity values, we expect that it would have a low affinity value with all the contours features in  $C_1$ .

Figure 4(c) shows the relative difference in affinity between the short contour shown in white (selected from Fig. 4(a)) and the other contours (from Fig. 4(b)). The brighter the contour, the higher the affinity. For this computation of affinity, we used the weights  $f_1 = 5$ ,  $f_2 = 5$ , and  $f_3 = 0$  (the intensity of the contours in this example were generated randomly).

### 4.3 Estimation of Conditional Probability using Contour Affinity

As stated earlier, affinity captures the possibility that two contours belong to the same underlying edge structure. If we assume that one of the contours belongs to an object boundary, one can interpret the affinity between two contours to be an indication of the probability that the second contour also belongs to the object boundary. In other words, the affinity between  $c_1$  and  $c_2$  can be treated as an estimate of the probability that  $c_1$  belongs to an object *given* that  $c_2$  does.

Consider once again the random variables  $S_1$  and  $S_2$ . Let  $C_1$ , the domain of  $S_1$ , now contain contour features extracted only from the current input image from sensor  $A$ . Similarly let  $C_2$ , the domain of  $S_2$ , contain contour features extracted from the corresponding image from sensor  $B$ . Based on the pair-wise affinity between contours of  $C_1$  and  $C_2$ , we define

$$P(c_1|c_2) = \frac{Aff(c_1, c_2)}{\sum_{c_i \in C_1} Aff(c_i, c_2)} \quad (5)$$

where  $P(c_1|c_2) \equiv P(S_1 = c_1|S_2 = c_2)$ .

### 4.4 Computing Mutual Information

The definition of the conditional probability in Eqn. 5 enables us to measure the conditional entropy between  $S_1$  and any contour  $c_j \in C_2$ . Using Eqn. 2, this can

be expressed as

$$H(S_1|c_j) = -p(c_j) \sum_{c_i \in C_1} p(c_i|c_j) \log p(c_i|c_j) \quad (6)$$

where the distribution  $p(c_j)$  can be considered as a prior expectation of observing a given contour feature. Similarly, assuming  $p(c_i)$  to be a known distribution (e.g., uniform), the entropy of  $S_1$  can be computed as

$$H(S_1) = - \sum_{c_i \in C_1} p(c_i) \log p(c_i) \quad (7)$$

Using Eqns. 6 and 7 in Eqn. 3 we can measure the mutual information  $I(S_1; c_j)$ . In order to obtain an estimate of the full joint mutual information  $I(S_1; S_2)$ , we consider each contour independently and use the approximation suggested in [24], which is the mean of all mutual information values between contour features  $c_j \in C_2$  and  $S_1$

$$I(S_1; S_2) = \frac{1}{|C_2|} \sum_{c_j \in C_2} I(S_1; c_j) \quad (8)$$

If we assume the prior distribution of contours features  $p(c_i)$  and  $p(c_j)$  to be uniform, the entropy of  $S_1$  (Eqn. 7) is constant. Maximizing the mutual information is then equivalent to finding the set of features from  $S_2$  that *minimizes* the conditional entropy  $H(S_1|S_2)$ . In other words, we seek those contours features from  $S_2$  that minimize the randomness of the object contour features in  $S_1$ . Rewriting Eqn. 8 using Eqns. 6 and 7, and using the assumption of uniform distributions for  $p(c_i)$  and  $p(c_j)$ , the conditional entropy of  $S_1$  and  $S_2$  can be expressed as

$$H(S_1|S_2) \propto \sum_{c_j \in C_2} \left( - \sum_{c_i \in C_1} p(c_i|c_j) \log p(c_i|c_j) \right)$$

where the term in parenthesis can be interpreted as the entropy of the distribution of affinity between  $c_j$  and the contours in  $C_1$ . This is indeed the notion of relevance we wish to capture since, as described in Sect. 4.2, the entropy of affinity values is expected to be low only for  $c_j$  lying on object boundaries.

## 5 Contour Feature Selection using Mutual Information

We now address the issue of selecting the most relevant set of contour features from  $S_2$  based on  $S_1$ . This problem statement is very reminiscent of the feature selection problem [24, 15], and the intuition behind the solution is also similar. We seek the subset of contours from  $S_2$  that maximizes the mutual information between  $S_1$  and  $S_2$ .

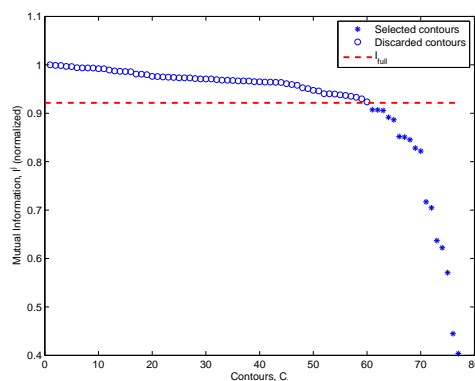
The problem of finding the subset that maximizes the mutual information is intractable since there are an exponentially large number of subsets that would need to be compared. An alternate greedy heuristic involves a simple incremental search scheme that adds to the set of selected features one at a time. Starting from an empty set of selected features, at each iteration, the feature from  $S_2$  that maximizes Eqn. 8 is added to the set of selected features. This solution, as proposed in the feature selection literature [24, 15], has one drawback in that there is no fixed stopping criteria, other than possibly a user-provided limit to the maximum number of features required [24]. Obviously, this is a crucial factor that would impede the use of this greedy selection scheme in most fusion applications.

We present here a modified version of the greedy algorithm that addresses the need for a reliable stopping criterion. Initially, the set  $C_2$  contains contour features that lie along the object boundary as well as a potentially large number of irrelevant contour features due to sensor noise and scene clutter. We start by computing the mutual information,  $I_{full}$ , between  $S_1$  and  $S_2$ . The algorithm is based on the observation that removing a relevant contour feature from  $C_2$  should *reduce* the mutual information ( $< I_{full}$ ), while removing an irrelevant feature should *increase* the mutual information ( $> I_{full}$ ). We iterate over all the individual contours in  $C_2$  and select only those contours that reduce the mutual information when removed from  $C_2$ . The outline of the complete feature selection algorithm is as follows:

1. Compute  $I_{full} = I(S_1; S_2)$ , where  $S_1$  and  $S_2$  are random variables defined over  $C_1$  and  $C_2$  respectively
2. For each  $c_j \in C_2$ 
  - (a)  $C_2^j \leftarrow C_2 \setminus \{c_j\}$
  - (b) Compute  $I^j = I(S_1; S_2^j)$ , where  $S_2^j$  is defined over  $C_2^j$
3. Select all  $c_j$  such that  $I^j \leq I_{full}$

Figure 6 shows the normalized mutual information values ( $I^j$ ) in descending order for the synthetic example images shown in Fig. 4(a) and (b). The dashed horizontal line in the figure corresponds to  $I_{full}$ , and can be considered to be the minimum mutual information required between  $S_1$  and  $S_2$ . The result of the contour selection procedure for this example is shown in Fig. 4(d). As can be seen, apart from a few internal contours, the subset of contours selected is reasonable. Figure 4(e) shows the contours from sensor  $A$  (Fig. 4(a)) overlaid in gray with the selected contours from sensor  $B$ . The slight misalignment in the contours from the two sensors was done intentionally to demonstrate the robustness of the algorithm to small errors in sensor registration.

While using  $I_{full}$  as the threshold in the contour selection procedure is effective, it can sometimes, due to inaccuracies in the estimation of the pdfs, prove to be too strict a threshold in real-world cases. A better threshold can be obtained in practice. Observing the profile of mutual information values  $I_j$  in descending order, we often see that there is a sharp drop (corresponding to the separation of object and non-object contours) in the mutual information at some value  $I_j = I_T$



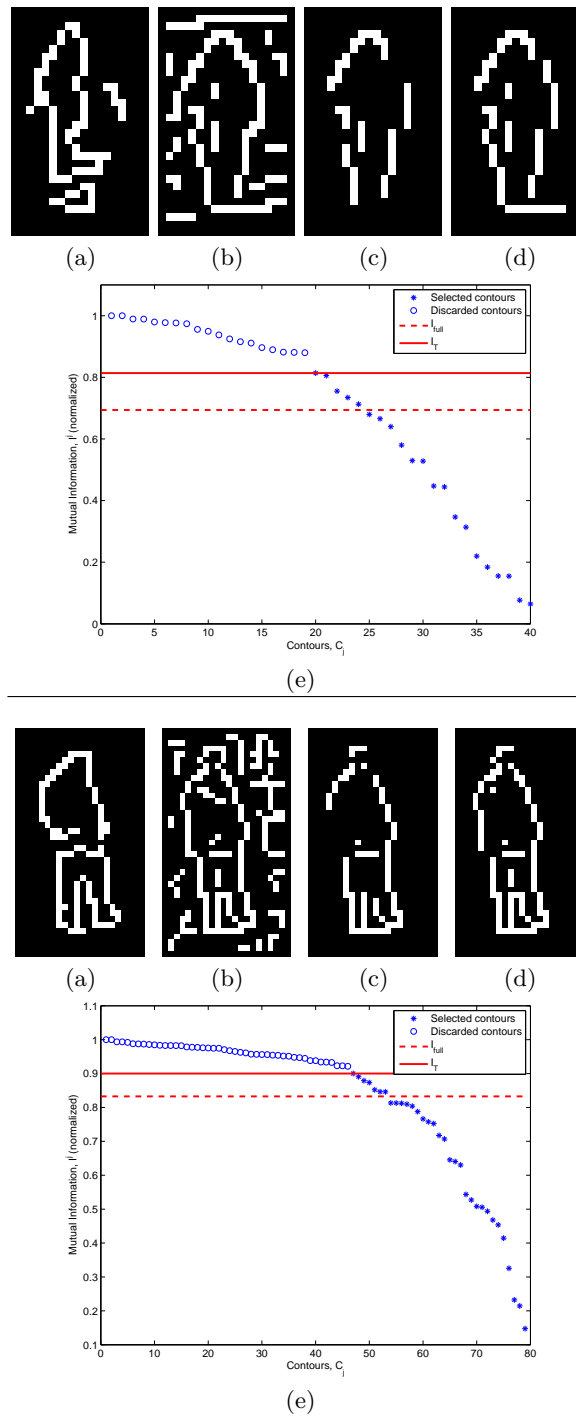
**Fig. 6.** Variation of mutual information values ( $I^j$ ) for different  $C_2^j$  sorted in descending order.

in the vicinity of  $I_{full}$  such that  $I_T \geq I_{full}$ . Using  $I_T$  instead of  $I_{full}$  in Step 3 of the above algorithm typically results in the selection of a better subset of contours.

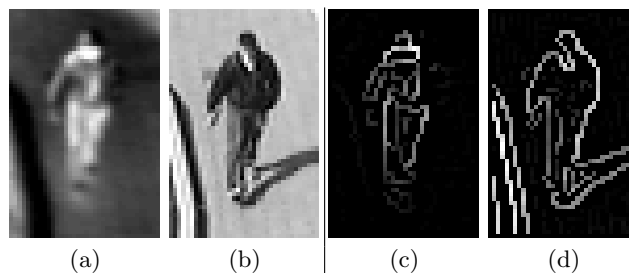
We show two real world examples of the contour feature selection scheme in Fig. 7 and also compare using  $I_T$  and  $I_{full}$  as the thresholds. Figures 7(a) and (b) show the contours extracted from the corresponding sub-images obtained from a thermal (sensor  $A$ ) and visible sensor respectively. In Fig. 7(c) we show the set of contours selected using  $I_{full}$  as the threshold. The contours selected using  $I_T$  as the threshold are shown in Fig. 7(d). The variation of the (normalized) mutual information values  $I^j$  for different  $C_2^j$  is shown in Fig. 7(e). The dashed horizontal line corresponds to  $I_{full}$ . The solid line represents  $I_T$ , the point  $\geq I_{full}$  in the mutual information profile with the *largest* drop. The mutual information profile corresponding to the first example shown in Fig. 7 shows a distinctive drop at  $I_T$ . Under conditions of high clutter, the profile of mutual information values may not contain a point with a distinctly large drop. However, as demonstrated by the second example of Fig. 7, the described heuristic still provides a reasonable separation of object/non-object contours in such cases.

## 6 Experiments

To test our feature-level fusion approach, we consider a video surveillance scenario that employs a pair of co-located and registered cameras. This setting enables us to evaluate the ability of our fusion approach to improve the shape segmentation of objects found in typical urban surveillance scenarios. The two sensors used are a ferroelectric thermal camera (Raytheon 300D core) and a color camera (Sony TRV87 Handycam). We analyzed several different thermal/color video sequence pairs recorded from different locations at different times-of-day. The sequences were recorded on a university campus, and show several people,



**Fig. 7.** Examples of contour feature selection based on variation of mutual information. (a) Contours from sensor *A* (thermal domain). (b) Contours from sensor *B* (visible domain). (c) Contours selected from sensor *B* using  $I_{full}$ . (d) Contours selected from sensor *B* using  $I_T$ . (e) Variation of mutual information values ( $I^j$ ) for different  $C_2^j$  sorted in descending order.



**Fig. 8.** An example input. (a) Thermal sub-image. (b) Visible sub-image. (c) Initial object contours detected from (a). (d) Thinned gradient magnitudes from (b).

some in groups, moving through the scene. We show an example of a typical image pair, cropped to a person region, in Fig. 8(a) and (b).

We begin by describing the choices made for the internal parameters of our algorithm, and providing a visual/qualitative assessment of the fusion results. Then, in Sect. 6.1, we present a detailed quantitative evaluation of the algorithm, including comparative analysis with other competing fusion techniques.

To choose the “reference” sensor ( $A$ ) for our algorithm, we considered the nature of the application and the ease of obtaining an initial segmentation. The need for persistence in a surveillance application, and the ease of background modeling in the relatively stable thermal domain [9], prompted us to choose the thermal camera as sensor  $A$ .

We employ the contour-based background subtraction scheme using Contour Saliency Maps (CSM) [9] along with a minimal threshold to directly obtain a preliminary detection of object contours from the thermal domain. For ease of computation, we break the corresponding thermal and visible input images into sub-images based on the regions obtained from background-subtraction in the thermal domain. Each thermal sub-image consists of contours that belong to a single object, or objects that were close to each other in the input image. The matching visible sub-image consists of all the thinned gradient magnitudes of the image region containing the object(s). In Fig. 8(c) and (d) we show an example of the sub-image pair corresponding to the image regions shown in Fig. 8(a) and (b).

These sub-image pairs form the input to our fusion algorithm. We first extract contour features from each sub-image as described in Sect. 3. We used 4 orientation bins with centers at 0, 45, 90, and 135 degrees, and a standard connected-components algorithm. For every pair of contour features from both domains, we then estimate the probability of a contour feature in the thermal domain conditioned on the occurrence of a feature from the visible domain (as described in Sect. 4.3). For the computation of contour affinity (Eqn. 4), in all the experiments, we used  $f_1 = 5$ ,  $f_2 = 5$ , and  $f_3 = 15$ .

The set of contour features from the visible domain that are most relevant to the object contour features from the thermal domain are chosen using the

steps outlined in Sect. 5. The final fused result is then obtained by overlaying these contour features selected from the visible domain with the contour features originally detected in the thermal domain. In case of misalignments that could arise due to small registration errors, standard morphological techniques are used to ensure that all contours are 1-pixel thick.

Given the contour feature sets  $C_1$  and  $C_2$  extracted from the two sensors, computing the mutual information involves obtaining the pair-wise affinity between contours and takes order of  $O(mn)$  time, where  $|C_1| = m$  and  $|C_2| = n$ . Determining the most relevant subset of features from  $C_2$  is an iterative procedure that has a worst case running time of  $O(m)$ . In our experiments, the three stages of the algorithm, namely extraction of features, computation of mutual information, and selection of the relevant subset of features, took on an average 0.43 seconds per input sub-image pair on a 2.8 GHz Intel P4 machine, using software written partly in Matlab and C.

We show several examples of the fusion results in Fig. 12. All images have been shown in binary to improve clarity. Figure 12(a) shows the detected contours obtained from the thermal domain. The Fig. 12(b) shows the thinned gradients from the visible domain. The set of contours selected by our algorithm from the visible domain are shown in Fig. 12(c). Figure 12(d) shows the final fused result obtained by overlaying Fig. 12(c) with Fig. 12(a). Overall, the results are satisfactory. The algorithm selects contours that both strengthen and complement the set of input object contours. In general, the outer boundaries of the fused result are a reasonable approximation of the true object shape. In spite of the presence of shadows and illumination changes, the proposed fusion framework is effective in obtaining a reasonable contour segmentation in the visible domain, that further improves the original segmentation acquired from the thermal sensor.

After the sub-images of an image pair have been processed, the resulting fused image contains contours extracted from both domains that best represent the objects in the scene. Several different vision applications can benefit from improvements in such a result, especially those that rely on the notion of object shape. Shape could be either extracted directly from the contours, or after using figure completion methods (such as [9]) on these contours. Examples of such applications include activity recognition, object classification, and tracking.

## 6.1 Quantitative Evaluation

As stated in Sect. 1, the challenge for any fusion algorithm is to utilize information from two or more sources so as to maximally improve the performance of the system over using either sensor individually. In this section we analyze how our fusion algorithm stands up to this challenge for the task of shape segmentation. The quantitative evaluation is based on the manual segmentation of the object regions in 73 images-pairs obtained from several thermal/color video sequences. Results of the hand-segmentation (by multiple people) of each pair of images were combined using an element-wise logical-OR operation to obtain the final



manually segmented images.

### Experiment 1: Fusion vs. Independent sensors

Since the final result of our algorithm is a set of contours, let us assume that we have available a module that can generate a closed shape (a silhouette) from such input. For evaluation, we propose then to use this module to generate a segmentation from three different sets of contours,

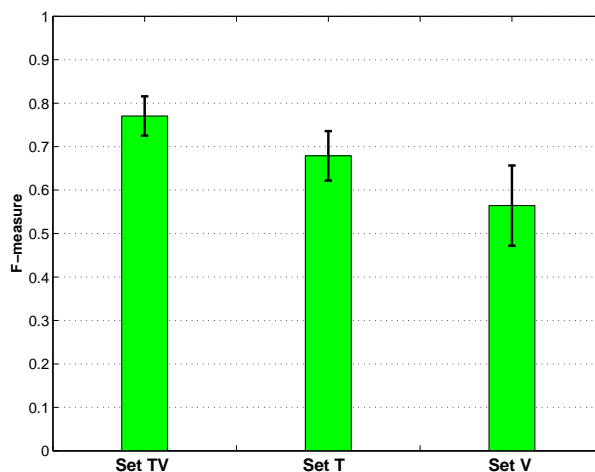
- Set T: contours from the thermal sensor initially detected as lying along the object,
- Set V: subset of contours from the visible sensor selected by the fusion algorithm,
- Set TV: overlay of the thermal and visible contours.

The comparison of the shape segmentation achieved in each of the above scenarios will provide valuable information that can be used to judge the validity of the proposed approach. Several approaches for contour-based figure completion exist. For the purpose of this evaluation, we make use of the method suggested in [9] to complete and fill the shape.

The set of 73 image-pairs generated a total of 208 useable sub-image pairs (a simple size criterion was used to eliminate sub-images that contained person regions that were too small). For each sub-image, the shape segmentation corresponding to the three sets of contours enumerated above were obtained. Examples of the silhouettes obtained from Set TV are shown in Fig. 12(e). To enable a visual assessment of the segmentation result, we show in Fig. 12(f) the manual segmentation of the image regions. Corresponding to each example, we also note the F-measure value obtained by comparing the generated silhouette (Fig. 12(e)) against the manually marked ground truth (Fig. 12(f)).

To quantify the segmentation results we compute Precision and Recall values using the manually segmented object regions as ground-truth. Precision refers to the fraction of pixels segmented as belonging to the object that are in fact true object pixels, while Recall refers to the fraction of object pixels that are correctly segmented by the algorithm. We combine these values into a single measure of performance using the F-measure [29], which is the harmonic mean of Precision and Recall. The higher the F-measure, the better the performance.

In Fig. 9 we present the mean F-measures evaluated for the three different scenarios over all the sub-images. The error-bars correspond to the variation in F-measure values obtained for each case. As can be seen from the plot, the quality of the segmentation obtained using the proposed fusion method ( $\bar{F}_{TV} = 0.77$ ) is clearly superior to that obtained from the initial object segmentation performed in the thermal domain ( $\bar{F}_T = 0.67$ ) (two-tailed t test:  $t(207) = 4.529, p < 5 \times 10^{-6}$ ). The improvement in segmentation shows that the proposed fusion algorithm is indeed able to extract relevant information from the visible domain such that the combination of information from the two sensors generates better results. The plot in Fig. 9 also shows the segmentation achieved using only the contour features chosen from the visible domain ( $\bar{F}_V = 0.56$ ). While clearly

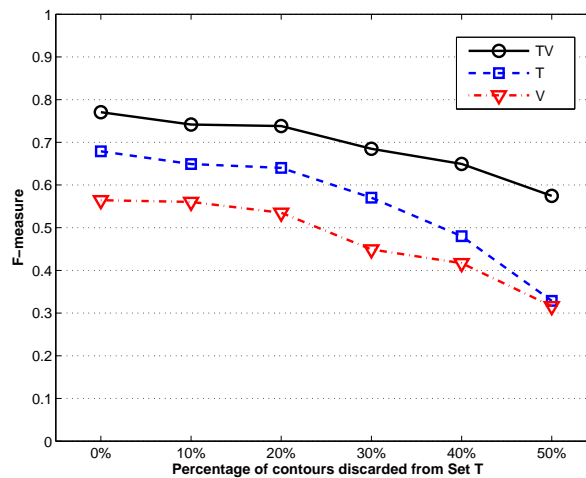


**Fig. 9.** F-measure comparison of fusion results with independent sensors.

lower than the segmentation performance of the thermal sensor (two-tailed t test:  $t(207) = 8.346, p < 1 \times 10^{-14}$ ), it should be noted that the segmentation in the visible domain is not aided by any prior knowledge or background information. It is obtained purely by identifying features that best complement and support the object features detected in the thermal domain.

Overall, these numbers demonstrate the ability of the proposed algorithm to use limited cues from one sensor to extract relevant information from the other sensor. The segmentation performance obtained from the fusion results show that the algorithm is successful in extracting both redundant and complementary information across modalities.

We next subject our algorithm to more adverse conditions and evaluate the ability of the algorithm to identify relevant contours from the visible domain given weaker initial detections in the thermal domain. We perform the same experiment as before, however this time we use only a subset of Set T by randomly discarding  $k\%$  ( $10 \leq k \leq 50$ ) of the contours. This resulting set is then used as input into our fusion algorithm. This experiment tests if the fusion algorithm is capable of estimating the correct set of relevant features from sensor  $B$ , given a more incomplete detection from sensor  $A$ . We vary the value of  $k$  systematically from 10 to 50 at intervals of 10%. At each value of  $k$ , the experiment is repeated 5 times, and the results presented here are averaged over the 5 runs. In Fig. 10 we show the variation in segmentation performance by plotting the F-measure against the different percentages of discarded contours. As expected, the performance for each of the Sets T, V, and TV decreases as the quality of the bootstrap segmentation is impoverished. However, what is interesting is the *rate* at which the performance deteriorates. It is clear from the plot that while the segmentation in the thermal domain (Set T) drops sharply as  $k$  increases,



**Fig. 10.** Variation of silhouette F-measure with smaller subsets of Set T.

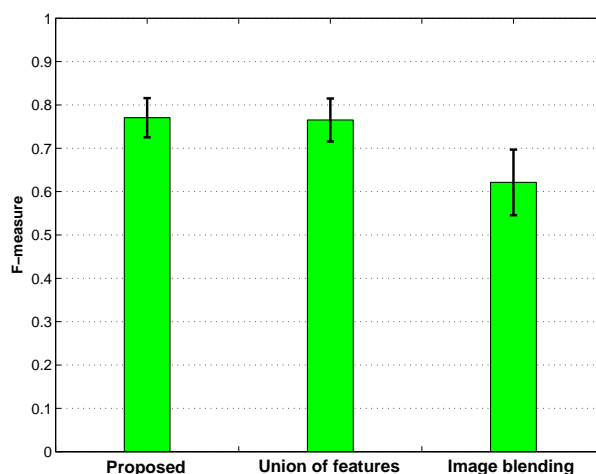
the fusion results (Set TV) show the most gradual decline in performance. It is also worth noting that the rate of change in the performance of the visible domain (Set V) mimics closely that of Set TV.

These results show that instead of being equally or perhaps worse affected by an impoverished input, the outputs of the fusion algorithm (Sets V and TV) show a much more graceful degradation in performance. In fact, the drop in segmentation performance for Set TV (in terms of F-measure score) is systematically lesser than Set T at every value of  $k$ . Thus as the initial bootstrap segmentation becomes weaker, the benefits of using the proposed fusion algorithm to combine information from the visible domain become increasingly apparent. These observations lead us to believe that the algorithm is indeed able to extract information from another sensor to compensate for incomplete information from one sensor.

### Experiment 2: Comparison against other methods

In this experiment, we compare the proposed fusion method against two other fusion approaches that could potentially be employed for object segmentation. The methods we compare against each belong to the two approaches introduced in Sect. 1.1. Here we provide details regarding the specific algorithms employed in each case.

**Image blending:** Fusion is performed by computing a region-wise weighted average of the input sensors. The weights for each circular region are determined using PCA of the pixel-intensities of the input images. For each local region, this results in higher weights being assigned to the sensor that has a higher variance of pixel intensity levels (for details see [8]). We use the method to fuse each color component of the visible domain with the thermal channel resulting in a fused image stream with three components. As is the case with the proposed algorithm,



**Fig. 11.** F-measure comparison of proposed fusion method with alternate approaches.

we employ background-subtraction as the method for obtaining the object segmentation from the fused image stream. Treating the three-channel fused image as a standard color image, we construct single Gaussian background models in the normalized color and intensity spaces. Background-subtraction is performed using each model separately and the results are combined so as to avoid identifying shadows as foreground regions. The final foreground object regions are composed of pixels found to be statistically different from the background in the color space and statistically brighter than the background in the intensity space.

**Union of features:** Binary contour fragments are obtained from the segmented foreground regions in each sensor, and then combined into a single image. Since background-subtraction is employed in each sensor, the combined image is formed by a union of all the extracted features. A gradient-based alignment procedure is also employed to compensate for small errors in registration (for details see [10]). In the thermal domain, background-subtraction is performed using the CSM technique [9]. In the visible domain, background-subtraction is performed using separate models in the intensity and color spaces as described earlier. Background-subtraction in the visible domain is only performed within the foreground regions (or blobs) obtained from the thermal domain. This ensures that the background-subtraction results in the visible domain are not adversely affected by sudden illumination changes commonly found in our data sets. The fused binary contours are completed into silhouette blobs using the same contour completion technique employed for the proposed method.

In order to compare the three methods, we once again rely on the hand-drawn silhouettes as ground truth. We first identify bounding boxes around each of the hand-drawn person silhouette. The output silhouette images generated by each method are evaluated only within these bounding boxes so as not to penalize

any of the techniques for poor background-subtraction results. The silhouette quality is measured, as before, using the F-measure of Precision and Recall.

The mean F-measure of the silhouettes obtained by each of the fusion methods is shown in Fig. 11. The error-bars in the plot correspond to the variation in the F-measure score. As can be seen from the comparison, the quality of the silhouettes obtained by the two feature-level fusion methods is clearly better (two-tailed t test:  $t(207) = 8.346, p < 1 \times 10^{-14}$ ) than that obtained from the image blending algorithm. Among the feature-level fusion methods, the proposed technique is able to generate marginally better (two-tailed t test:  $t(207) = 0.311, p < 0.756$ ) results than the alternate method used for comparison, in spite of requiring only a rough initial segmentation from only one sensor.

## 6.2 Discussion

Both of the experiments described above demonstrate several useful properties of the approach presented in this chapter. The results of the first experiment clearly show that the proposed technique adequately meets the basic requirement of any fusion algorithm, that of providing superior performance than can be achieved from using either sensor individually. More specifically, as shown in Fig. 9, the proposed fusion algorithm generates significantly better object silhouettes than those obtained from employing a standard object segmentation routine in the thermal domain. This clearly shows that, based on the available segmentation in the thermal domain, the fusion algorithm is able to extract relevant information from the visible domain in order to improve the overall segmentation performance.

Since the proposed algorithm requires an initial object segmentation, it is of interest to see how the quality of this initial bootstrap segmentation affects the final segmentation result. As shown in Fig. 10, we see that the proposed approach is able to utilize highly impoverished segmentation information, and yet generate overall silhouettes of much higher quality. Thus, Experiment 1 shows that the proposed algorithm is capable of significantly improving object segmentation results, and further, the algorithm is able to maintain reasonable segmentation quality even as the required bootstrap segmentation is made so weak as to cover only half of the actual object region.

Next, in Experiment 2, we provide a direct comparison of our approach with two other fusion strategies, one a low-level technique and the other a mid-level technique. We see that while low-level fusion techniques are useful for creating blended images for visual inspection, they are unlikely to be very effective in more “goal-oriented” scenarios, such as the object segmentation task under investigation in this work. As discussed in Sect. 1.1, the fusion process in such methods suffer from the lack of higher-level knowledge, such as which image features are likely to be useful and which detrimental to the task at hand. The fused images produced contain combinations of image features peculiar to each imaging sensor. Further, the blended image stream requires the use of segmentation routines specifically tailored to handle the unique image characteristics borne out of the fusion strategy employed.

As can be seen from the plot in Fig. 11, the feature-level fusion strategies fare better for goal-oriented fusion tasks. The “union of features” method used for comparison represents a brute-force approach to fusion, wherein object segmentation is performed in each available sensor, and the final result is obtained by simply combining all the extracted features. Such an approach implicitly assumes that all of the individual object segmentation results are accurate. Any segmentation errors made in either sensor are also manifest in the final segmentation, resulting in poor quality silhouettes. Thus, in order to effectively utilize such an approach, it is essential to have access to high quality object segmentation algorithms in each of the sensors. In terms of overall performance, we see (from Fig. 11) that in spite of requiring object segmentation in only one of the sensors, the proposed fusion algorithm provides silhouette quality that is in fact marginally better than that provided by the brute force “union of features” technique.

Contrary to the “union of features” method, the proposed fusion approach requires only a rough, incomplete segmentation in either one of the sensors. We note that since the algorithm utilizes all of the contour features extracted from the initial segmentation, it is preferable that this segmentation be as reliable as possible, even at the cost of it being considerably incomplete. The intelligent feature selection process utilizes this initial segmentation to extract relevant features from the other sensor, without requiring any *a priori* segmentation information in that domain. Since the proposed approach can be bootstrapped using either sensor, object segmentation needs to be performed in only that sensor in which it is likely to be more reliable. For example, in our experimental set-up, performing background-subtraction in the single channel thermal domain is both more reliable and computationally cheaper than performing background-subtraction in the three channel color space. While our approach requires only thermal background-subtraction, the “union of features” method required additional background-subtraction in the visible domain together with a shadow removal step.

## 7 Summary

We presented a new, goal-oriented, feature-level fusion technique for object segmentation based on mutual information. The proposed algorithm treats fusion as a feature selection problem. The approach utilizes the natural structure of the world within a mutual information framework in order to define a suitable criterion for feature selection. Starting from an initial detection of object features in one sensor, our technique extracts relevant information from the other sensor to improve the quality of the original detection.

We first defined a feature representation based on contour fragments that is rich enough to implicitly capture object shape yet simple enough to provide an easy realization of feature relevance. We then approached fusion as a variation of the mutual information feature selection problem. To avoid the pitfalls of learning the relevant probability distributions from training data, we proposed

a method that generates the required probability distribution from a single pair of images. The method computes the conditional probability distribution based on the notion of contour affinity and effectively captures the expectation that objects have regular shapes and continuous boundaries. We then computed the mutual information between the features extracted from both sensors. Finally, we employed a new scheme to reliably obtain a subset of features from the secondary sensor that have the highest mutual information with the provided object contours. The final fused result is obtained by overlaying the selected contours from both domains. The final contours are then complete and filled to create silhouettes.

Our approach was tested in a video surveillance setting, using co-located thermal and color cameras. The fusion algorithm improved object segmentation performance over using either sensor alone. Experiments were conducted using a set of over 200 manually segmented object regions, and were evaluated using the F-measure of Precision and Recall. The segmentation result of the fusion algorithm yielded an F-measure of 0.77, better than those obtained from detection results of either sensor used independently. The proposed algorithm was also compared to other fusion approaches, a low-level technique [7] and another mid-level technique [10], and was shown to produce comparable (or better) results while requiring lesser computational resources.

In the future, we plan to extend the method to enable two-way information flow in our fusion pipeline. Such an approach would potentially enable the final segmentation to be built up incrementally, such that, in each iteration the segmentation from one sensor would seed feature selection in the other, and so on. We would also like to investigate the robustness of our feature representation to translation and rotation of the sensors. This would potentially enable our approach to withstand larger errors in image registration across the sensors.

## 8 Acknowledgements

This research was supported in part by the National Science Foundation under grant No. 0428249. A shorter version of this paper appeared in the 2006 *IEEE Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum* [31].

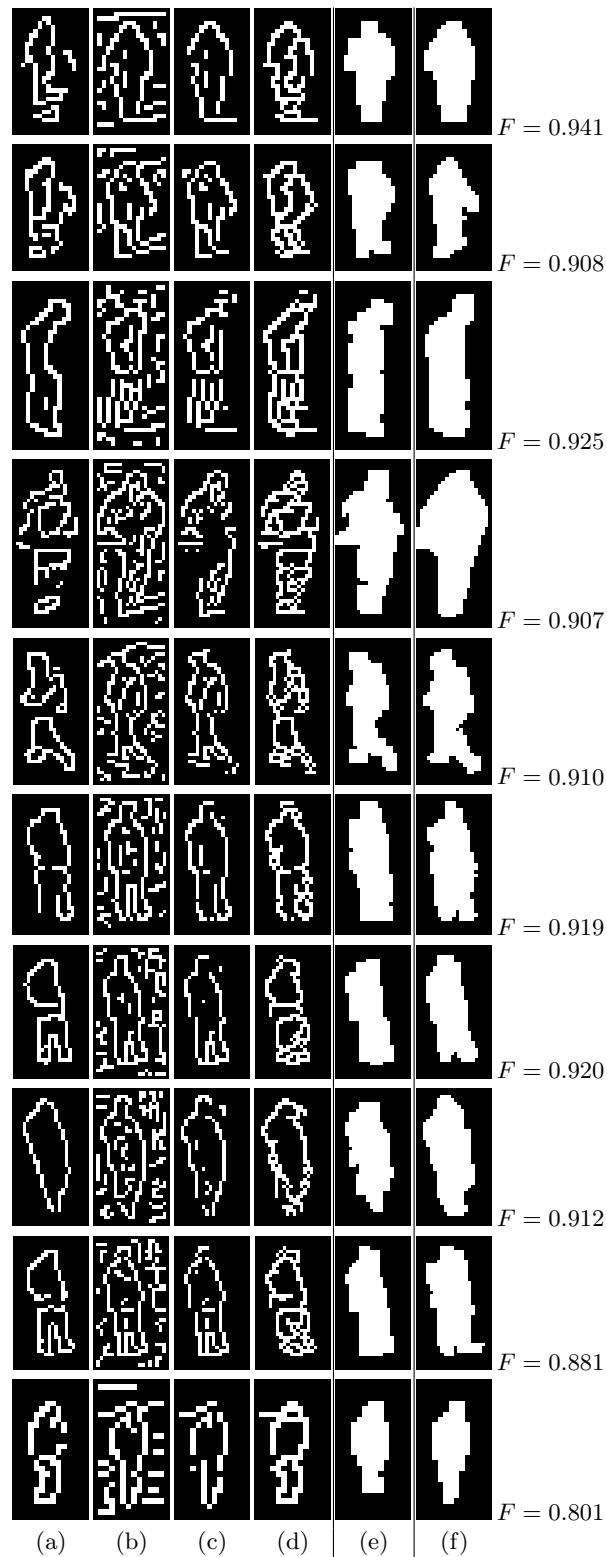
## References

1. T. Bakert and P. Losiewicz. Force aggregation via bayesian nodal analysis. In *Proceedings of Information Technology Conference*, 1998.
2. P. Bogler. Shafer-dempster reasoning with applications to multisensor target identification systems. In *IEEE Transactions on System, Man, and Cybernetics*, volume 17, pages 968–977, 1987.
3. D. Borghys, P. Verlinde, C. Perneel, and M. Achery. Multi-level data fusion for the detection of targets using multi-spectral image sequences. In *SPIE Optical Engineering, special issue on Sensor Fusion*, volume 37, pages 477–484, 1998.

4. P.J. Burt and R.J. Kolczynski. Enhanced image capture through fusion. In *Proc. Comp. Vis. and Pattern Rec.*, pages 173–182, 1993.
5. T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.
6. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. Int. Conf. Comp. Vis.*, pages 886–893, 2005.
7. S. Das and W. Krebs. Sensor fusion of multi-spectral imagery. *Electronics Letters*, 36:1115–1116, 2000.
8. B. Dasarathy. *Decision Fusion*. IEEE Computer Society Press, 1994.
9. J. Davis and V. Sharma. Background-subtraction in thermal imagery using contour saliency. *Int. J. of Comp. Vis.*, 71(2):161–181, 2007.
10. J. Davis and V. Sharma. Background-subtraction using contour-based fusion of thermal and visible imagery. *Computer Vision and Image Understanding*, 106(2-3):162–182, 2007.
11. R. Delaunoy, J. Verly, and D. Dudgeon. Pixel-level fusion using interest images. In *Proc. of the 4th National Symposium on Sensor Fusion*, volume 1, pages 29–41. IRIA (ERIM), 1991.
12. Francis Corbett et. al. Fused atr algorithm development for ground to ground engagement. In *Proceedings of the 6th National Sensory Symposium*, volume 1, pages 143–155, 1993.
13. D. A. Fay et al. Fusion of multi-sensor imagery for night vision: Color visualization, target learning and search. In *3rd International Conference on Information Fusion*, pages TuD3-3 –TuD3-10, 2000.
14. M. Hinman. Some computational approaches for situation assessment and impact assessment. In *Proceedings of the Fifth International Conference on Information Fusion*, pages 687–693, 2002.
15. N. Kwak and C. Choi. Input feature selection by mutual information baed on parzen window. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 24(12):1667–1671, 2002.
16. L. Lazofson and T. Kuzma. Scene classification and segmentation using multispectral sensor fusion implemented with neural networks. In *Proceedings of the 6th National Sensor Symposium*, volume 1, pages 135–142, 1993.
17. J. Lewis, R. O’Callaghan, S. Nikolov, D. Bull, and C. Cangarajah. Region-based image fusion using complex wavelets. In *International Conference on Information Fusion*, pages 555–562, 2004.
18. H. Li, B. S. Manjunath, and S. K. Mitra. Multisensor image fusion using the wavelet transform. In *Graphical Model and Image Processing*, volume 57, pages 234–245, 1995.
19. N. Li, S. Dettmer, and M. Shah. Visually recognizing speech using eigensequences. In *Motion-Based Recognition*, pages 345–371. Kluwer Academic, Dorcrecht, 1997.
20. J. Lowrance, T. Garvey, and T. Strat. A framework for evidential reasoning system. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, pages 896–901, 1986.
21. K. Mikolajczyk, A. Zisserman., and C. Schmid. Shape recognition with edge-based features. In *Brit. Mach. Vis. Conf.*, pages 779–788, 2003.
22. B. Park and J. Marron. Comparison of data-driven bandwidth selectors. *J. of Amer. Stat. Assoc.*, 85(409):66–72, 1990.
23. M. Pavel, J. Larimer, and A. Ahumada. Sensor fusion for synthetic vision. In *AIAA Conference on Computing in Aerospace 8*, 1991.



24. H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance and min-redundancy. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 27(8):1226–1238, 2005.
25. V. Petrovic and C. Xydeas. Gradient-based multiresolution image fusion. *IEEE Transactions on Image Processing*, 13(2):228–237, 2004.
26. G. Piella. A region-based multiresolution image fusion algorithm. In *Information Fusion*, pages 1557–1564, 2002.
27. C. Ramac, M. Uner, P. Varshney, M. Alford, and D. Ferris. Morphological filters and wavelet-based image fusion for concealed weapons detection. *Proc. of SPIE*, 3376:110–119, 1998.
28. R. Raskar, A. Llie, and J. Yu. Image fusion for context enhancement and video surrealism. In *Non-Photorealistic Animation and Rendering*, pages 85–94. ACM, 2004.
29. C. Van Rijsbergen. *Information Retrieval*. Dept. of Computer Science, University of Glasgow, second edition, 1979.
30. P. Scheunders. Multiscale edge representation applied to image fusion. In *Wavelet applications in signal and image processing VIII*, pages 894–901, 2000.
31. V. Sharma and J. Davis. Feature-level fusion for object segmentation using mutual information. In *IEEE Int. Wkshp. on Object Tracking and Classification Beyond the Visible Spectrum*, 2006.
32. E. Sharon, A. Brandt, and R. Basri. Completion energies and scale. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 22(10):1117–1131, 2000.
33. Diego A. Socolinsky and Lawrence B. Wolff. A new visualization paradigm for multispectral imagery and data fusion. In *Proc. Comp. Vis. and Pattern Rec.*, pages 319–324, 1999.
34. A. Toet. Heirarchical image fusion. *Machine Vision and Applications*, 3:1–11, 1990.
35. K. Torkkola. Feature extraction by non-parametric mutual information maximization. 3:1415–1438, 2003.
36. P. Varshney. *Distributed Detection and Data Fusion*. Springer Verlag, 1996.
37. L. Williams and D. Jacobs. Stochastic completion fields: a neural model of illusory contour shape and salience. *Neural Computation*, 9(4):837–858, 1997.
38. M Kokar Z Korona. Model-based fusion for multisensor target recognition. In *Proc. SPIE*, volume 2755, pages 178–189, 1996.
39. Z. Zhang and R. Blum. Region-based image fusion scheme for concealed weapon detection. 1997.



**Fig. 12.** Examples of fusion results. (a) Contours detected from thermal domain (Set T). (b) Contours present in the visible domain. (c) Contours selected from (b) (Set V). (d) Overlay of contours from (c) on (a) (Set TV). (e) Segmentation obtained after completing and filling (d). (f) Manually segmented object regions, and corresponding F-measure values (on comparison with (e)).