Feature-level Fusion for Object Segmentation using Mutual Information [*]

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

Abstract

We present a new feature-level image fusion technique for object segmentation based on mutual information. Using object regions roughly detected from one sensor as input, the proposed technique extracts relevant information from another to complete the 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. Results are presented for segmenting people from background, and quantitatively evaluated.

1. Introduction

In vision applications, such as video surveillance and automatic target recognition, imaging sensors of different modality are often used. 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.

We adopt here a more "goal-oriented" view of image fusion than is traditionally used. Instead of improving the context (or information) present in a scene, we consider the task of using image fusion to improve the estimation of the shape (as defined by a silhouette, or a boundary) of an object. We start with an initial detection of object regions from one sensor (*A*). The proposed algorithm then takes these detected regions as input, and extracts relevant information from the other sensor (*B*) in an attempt to produce a better segmentation of the object. The initial detections can be obtained from either sensor. Depending on the application, factors such as persistence, signal-to-noise ratio, and the availability and complexity of the detection scheme can influence this choice. The proposed fusion technique enables a user to benefit from the presence of two imaging modalities at the cost of providing an initial segmentation from either one.

Image fusion algorithms can be broadly classified into low-, mid-, and high-level techniques based on their position in the information processing pipeline. Our proposed algorithm is a goal-oriented mid-level fusion technique that utilizes contour-based features. As will be shown, these features allow a description of shape from partially detected object regions. We use these features within a mutual information framework to extract relevant (redundant and complementary) information across sensors. Deriving such information directly from raw pixel intensities or similar low-level fusion cues would be difficult. High-level fusion techniques usually employ voting schemes to combine results after independently processing each input. Such high-level techniques are not well suited to situations when only *two* input channels exist. Further, the cost of obtaining detection results independently in each channel can be considered an overhead.

We start by extracting contour features from the input image regions of both sensors (assumed to be co-registered). We extend the notion of affinity, originally defined to measure the smoothness of the curve joining two edge elements [17], to contours. 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.

We demonstrate the approach for a video surveillance application using a thermal and color camera as the two in-

[*] Appears in *IEEE Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum*, New York, NY, June 22, 2006.

put sensors. Based on manually segmented object regions we show the efficacy of the proposed method by comparing segmentation performance using the fused result over using either input sensor independently.

2. Related Work

Image fusion techniques have a long history in vision. Gradient-based techniques include defining first-order contrasts in high dimensions [14] and examining gradients at multiple resolutions [13]. Several region-based multi-resolution algorithms have been proposed such as the pyramid approaches of [15, 10] and the wavelet-based approach of [7]. Other biologically motivated techniques [5] have also been proposed. Most of these fusion techniques aim at enhancing the information content of the scene, to ease and improve human visual analysis. In contrast, the method we propose is designed specifically to enhance the capabilities of an automatic vision-based detection system.

Recently [4] proposed a fusion algorithm also designed with a similar aim. However, their fusion technique was specific to a thermal and color camera, and required background modeling in both domains. Our proposed method is an improvement on both counts, in that it is not tied to any particular combination of sensors and it 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 detection in 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 [2, 8]. 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 $c = [ep_1, ep_2, \theta, E_{mag}]$, where ep_1 and ep_2 are the coordinates of the two end-points, θ is the quantized orientation, 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. The shape of the imaged object is implicitly captured by the magnitude (E_{mag}), position (ep_1, ep_2), and orientation (θ) of the contour features.

4. Estimating Feature Relevance

Having extracted contour features, our goal 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 [1]. 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 [6, 11, 16].

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

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

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 (or uncertainty) 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

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

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

$$I(X;Y) = H(X) - H(X|Y) \tag{3}$$

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 [16, 11].

Nevertheless, a typical approach would be to estimate these distributions using a large training data-set consisting

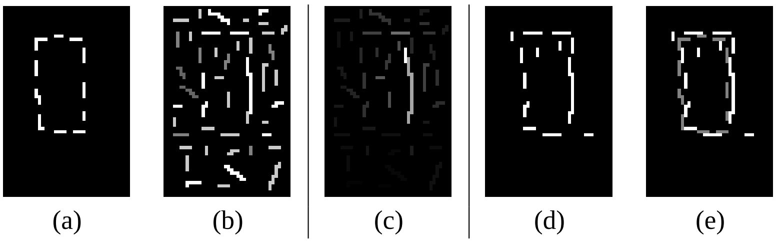

Figure 1: Toy example illustrating the relevant processing stages. (a) Detected object contours from sensor *A*. (b) Contours obtained from sensor *B*. (c) Relative affinity values of contours in (b) with respect to a contour (shown in white) from (a). (d) Set of contours selected from (b). (e) Overlay of contours from (a), shown in gray, with the selected contours (d).

of manually segmented objects imaged using the sensors in question. The difficulty of generating such a data-set 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 [9] 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. 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. 1(a) and (b) represent the thinned gradient magnitudes of a rectangular box imaged using two sensors. Let Fig. 1(a) represent the object contours from sensor *A*, and Fig. 1(b) the set of contours obtained from sensor *B*.

Visualizing the contour features extracted from sensor *A* in image-space, as in Fig. 1(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 *A*.

Perceptual (and computational) figure completion is a very active field of research. For the purpose of figure completion, several studies, such as [17], 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. We borrow this notion of affinity and adapt it to deal with contours of finite size instead of the dimensionless edge elements used in the literature.

Consider a pair of contours c_1 and c_2. Hypothesize a curve connecting c_1 and c_2 such that the contours lie completely along it. Any such connection would join one of the end-points of c_1 to an end-point of c_2. Based on which two end-points are connected, all such curves fall into one of four categories. Consider one such curve 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. Let θ_1 denote the angle between this vector and the outward pointing unit vector at the end-point of c_1, directed along the tangent to c_1. Let θ_2 denote the angle from c_2, analogous to θ_1. Finally, let r denote the Euclidean distance between the two end-points of c_1 and c_2. These quantities, θ_1, θ_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. The affinity for a particular curve set is defined as

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

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 σ_r, σ_t, and σ_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, β, and Δ, 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 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 1(c) shows the relative difference in affinity between the short contour shown in white (selected from Fig. 1(a)) and the other contours (from Fig. 1(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 $f_3 = 0$, since 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

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

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

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

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

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

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 [11], which is the mean of all mutual information values between contour features $c_j \in C_2$ and S_1

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

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 [11, 6], 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.

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

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

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.

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 [11, 6], 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 [11]. 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:

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 $I^j \leq I_{full}$

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. The algorithm presented above 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 ($> I_{full}$).

I_{full} can be considered to be the minimum mutual information required between S_1 and S_2. Using I_{full} as the threshold, however, can sometimes result in only a few c_j being selected due to inaccuracies in the estimation of the pdfs. A better threshold can be obtained in practice. If we observe the sequence of mutual information values I_j in descending order (e.g. Fig. 2), there is often a sharp drop (corresponding to the separation of object and non-object contours) in the mutual information at some value $I_j = I_T$ 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. Figure 2 shows the mutual information values I^j for a pair of input images. 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. In practice, the subset of contours below I_T form a better solution than using just the contours below I_{full}. Under conditions of very high clutter, the profile of mutual information values may not contain a point with a distinctly large drop. However, the described heuristic still provides a reasonable separation of object/non-object contours in this case.

The result of the overall contour selection process, applied to the example problem of Fig. 1(a) and (b), is shown in Fig. 1(d). As can be seen, apart from the stray contour in the bottom right, and a few internal contours, the subset of contours that is selected is reasonable. Figure 1(e) shows the contours from sensor *A* (Fig. 1(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.

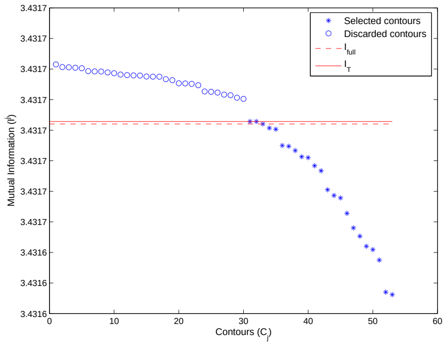

Figure 2: Variation of mutual information values (I^j) for different C_2^j sorted in descending order. I_T (solid line) forms a better threshold than I_{full} (dashed line) in practice.

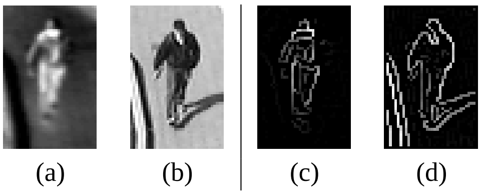
(a) (b) (c) (d)

Figure 3: An example input. (a) Thermal sub-image. (b) Visible sub-image. (c) Initial object contours detected from (a). (d) Thinned gradient magnitudes from (b).

6. Experiments

To test our approach for feature-level fusion, we consider a video surveillance scenario that employs a pair of co-located and registered cameras. This enables us to evaluate the ability of our fusion approach to improve the shape segmentation of objects in typical urban surveillance scenarios. The two sensors used are a ferroelectric thermal camera (Raytheon 300D) and a color camera (Sony TRV87 Handycam). We show an example of a typical image pair, cropped to a person region, in Fig. 3(a) and (b). Note that all the color images are converted to gray-scale.

To choose the “reference” sensor (*A*), we considered the nature of the application and the ease of obtaining an initial detection. The need for persistence in a surveillance application, and the ease of background modeling in the relatively stable thermal domain [3], prompted us to choose the thermal camera as sensor *A*. We employ the contour-based background subtraction scheme using Contour Saliency Maps [3] 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. 3(c) and (d) we show an example of the sub-image pair corresponding to the image regions shown in Fig. 3(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 can be used to ensure that all contours are 1-pixel thick.

We show several examples of the fusion results in Fig. 6. All images have been shown in binary to improve clarity. Figure 6(a) shows the detected contours obtained from the thermal domain. The Fig. 6(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. 6(c). Figure 6(d) shows the final fused result obtained by overlaying Fig. 6(c) with Fig. 6(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 or contour-based segmentation methods 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 39 images-pairs chosen at random 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.

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

- Set T: initially detected from the thermal sensor,
- Set V: subset selected from the visible sensor,
- 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 segmentation and figure completion exist. For the purpose of this evaluation, we make use of the method suggested in [3] to complete and fill the shape.

The set of 39 image-pairs generated a total of 65 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 segmentation for Set TV are shown in Fig. 6(e). To enable a visual assessment of the segmentation result, we show in Fig. 6(f) the manual segmentation of the image regions.

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

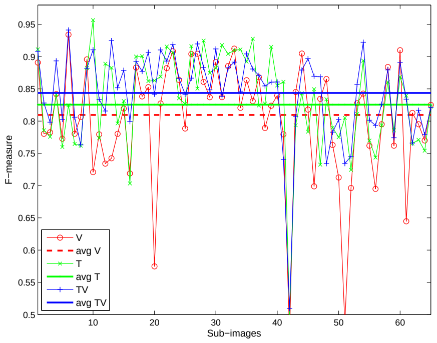

Figure 4: Comparison of F-measure.

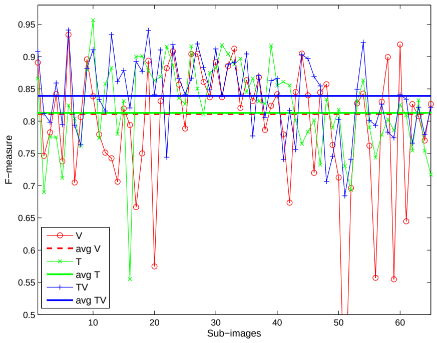

Figure 5: F-measure comparison using subset of Set T.

combine these values into a single measure of performance using the F-measure [12], which is the harmonic mean of Precision and Recall. The higher the F-measure, the better the performance.

In Fig. 4 we present the F-measures evaluated for the three different scenarios, over all the sub-images. The bold horizontal lines represent the average F-measure ($\overline{F}$) over all the sub-images for each input set of contours ($\overline{F}_T = 0.8254$, $\overline{F}_V = 0.8108$, $\overline{F}_{TV} = 0.8436$). As clearly shown in the figure, Set TV, the result of the fusion algorithm, yields the best overall segmentation result. The improvement over Set T, which is the detection result from the thermal sensor, is 2.2%. It is interesting also to note that Set V, the set of contours selected by the fusion algorithm from the visible domain, also performed comparably to the thermal domain (lower than Set T by 1.8%).

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. We perform the same experiment as before, only this time we use a subset of Set T by randomly discarding 10% 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*. The results of this experiment are shown in Fig. 5. As in Fig. 4, the bold lines represent the average F-measure generated from each set ($\overline{F}_T = 0.8127$, $\overline{F}_V = 0.8106$, $\overline{F}_{TV} = 0.8399$). Once again, Set TV, the fusion result, yields the best segmentation. The improvement over Set T is 3.4%, slightly higher than earlier. Also, the segmentation generated from Set V is very similar (lower by 0.2%) to Set T.

Comparing these results to that of the previous experiment, the interesting point to note is the relative changes in the F-measure across Sets T, V, and TV. As expected, discarding contours adversely affects the segmentation results achieved from Set T. However, Sets V and TV hardly show a decrease in performance. In fact, the drop in segmentation performance for Sets V and TV is only around 20% of the decrease seen in Set T. These observations lend further credibility to the proposed fusion scheme. In particular, this shows that the algorithm is (at least to some extent) able to extract information from the other sensor to compensate for the impoverished input from one sensor.

7. Summary

We presented a new, feature-level fusion technique for object segmentation based on mutual information. 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 complex enough to implicitly capture object shape and 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 generatesthe 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.

Our approach was tested in a video surveillance setting, using co-located thermal and color cameras. Using over 65 manually segmented object regions, the result of the fusion algorithm yielded better segmentation results than those obtained from detection results of any one sensor alone.

In the future we plan to extend the method to extract information from *both* sensors simultaneously. We would also like to investigate the robustness of our feature representation to translation and rotation of the sensors.

References

[1] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. Int. Conf. Comp. Vis.*, pages 886–893, 2005.

[3] J. Davis and V. Sharma. Robust background-subtraction for person detection in thermal imagery. In *IEEE Int. Wkshp. on Object Tracking and Classification Beyond the Visible Spectrum*, 2004.

[4] J. Davis and V. Sharma. Fusion-based background subtraction using contour saliency. In *IEEE Int. Wkshp. on Object Tracking and Classification Beyond the Visible Spectrum*, 2005.

[5] 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.

[6] 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.

[7] 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.

[8] K. Mikolajczyk, A. Zisserman., and C. Schmid. Shape recognition with edge-based features. In *Brit. Mach. Vis. Conf.*, pages 779–788, 2003.

[9] B. Park and J. Marron. Comparison of data-driven bandwidth selectors. *J. of Amer. Stat. Assoc*, 85(409):66–72, 1990.

[10] M. Pavel, J. Larimer, and A. Ahumada. Sensor fusion for synthetic vision. In *AIAA Conference on Computing in Aerospace 8*, 1991.

[11] 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.

[12] C. Van Rijsbergen. *Information Retrieval*. Dept. of Computer Science, University of Glasgow, second edition, 1979.

[13] P. Scheunders. Multiscale edge representation applied to image fusion. In *Wavelet applications in signal and image processing VIII*, pages 894–901, 2000.

[14] D. Socolinsky and L. Wolff. A new visualization paradigm for multispectral imagery and data fusion. In *Proc. Comp. Vis. and Pattern Rec.*, pages 319–324, 1999.

[15] A. Toet. Heirarchical image fusion. *Machine Vision and Applications*, 3:1–11, 1990.

[16] K. Torkkola. Feature extraction by non-parametric mutual information maximization. *J. of Machine Learning Research*, 3:1415–1438, 2003.

[17] L. Williams and D. Jacobs. Stochastic completion fields: a neural model of illusory contour shape and salience. *Neural Computation*, 9(4):837–858, 1997.

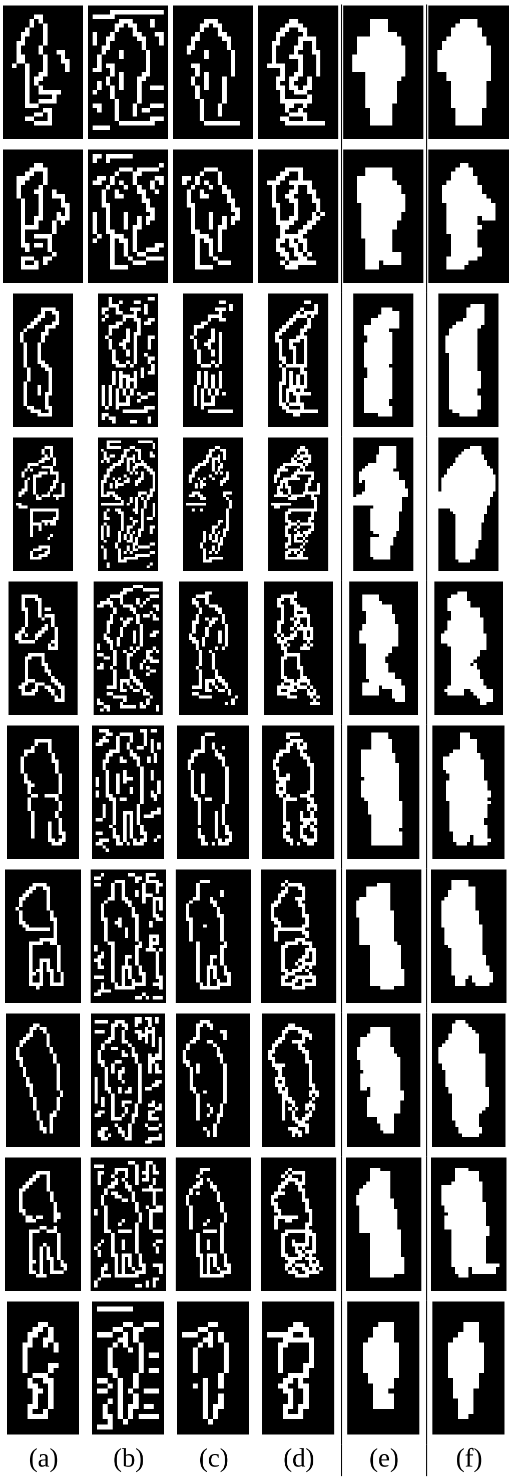

Figure 6: Examples of fusion results. (a) Contours detected from thermal domain (Set T). (b) Contours present in the visible domain. (c) Contours selected from (b) (Set V). (d) Overlay of contours from (c) on (a) (Set TV). (e) Segmentation obtained after completing and filling (d). (f) Manually segmented object regions.