

Possibilistic C-Means In Scene Matching

Ozy Sjahputera

Dept. Of Pathology and Anatomical Sciences
University of Missouri-Columbia
sjahputerao@missouri.edu

James M. Keller

Dept. of Electrical and Computer Engineering
University of Missouri-Columbia
kellerj@missouri.edu

Abstract

Determining if two images acquired at different times and under different viewing conditions contain the same scene is a difficult problem in computer vision. We demonstrate an approach that utilizes spatial relationships among the objects in the two scenes that ultimately produces a mapping of objects from one view to the other, and as a bonus, recovers the viewing transformation parameters. The core of the system relies on capturing spatial relationship information through Force Histograms, affine-invariant image descriptors. Object mapping across images is performed by finding the best correspondence map (FMAP) between force histograms in the two images. The major problem is that the number of potential FMAPS is huge, even for modest numbers of scene objects. Hence, some optimization is required. Similar feature vectors are observed from F-histogram matching defined in the best correspondence map. Therefore, dense regions in the feature space are suspected to contain these vectors. Possibilistic C-Means clustering (PCM) is used to find these dense regions. The centroids of these dense regions are used to generate an FMAP using a nearest-neighbor like approach. The best FMAP is selected and translated into an object map identifying the correspondences between objects in the two images.

Keywords: Scene Matching, Spatial Relations, Force Histograms, Possibilistic C-Means, Clustering.

I Introduction

Scene analysis is among the most complex problems in computer vision involving areas such as image segmentation, object recognition, and spatial reasoning. Scene matching is the process of assessing the degree of similarity of two images, which often includes the identification and/or mapping of objects. Object information such as shape, size, color, and texture are used as features in several scene matching approaches. However, objects may appear different when captured from different viewing geometries and/or at different times of day or year under different circumstances, thus having different feature values.

Besides using objects features, humans also use spatial reasoning to identify scenes. Humans are very good at identifying a scene from different orientations; we apply “transformations” to objects spatial relationships to adapt them to our current viewing orientation. Thus, object spatial relationships offer a set of features with potential for scene matching. A recent technique for studying spatial relationship is the *force histogram* (F-histogram) method [1]. This method is built on solid mathematical foundations and it supercedes the *angle histogram* method introduced earlier [2]. Following the survey by Brown [3], image registration methods can be loosely divided into four categories based on matching element types such as the frequency-domain based algorithms (e.g. the Fourier-Mellin with log-polar transform [4][5]), low level image features (e.g. edges and points [6][7]), direct pixel correlation [8], and high level features related to objects [9][10][11][12].

This work extends [11] and [12], and is detailed in [13]. Previously, the best histogram maps were found exhaustively and no system was in place to translate the histogram map into a mapping between objects. We now formalize the definition of the F-histogram map (FMAP) and its properties. We show that the FMAP search space is too large for exhaustive search. We use the *Possibilistic C-*

Means (PCM) method [14][15] to detect dense regions in the search space suspected of containing good FMAPs. We propose a protocol to translate FMAP into an object correspondence map (OMAP). Matching performance of this approach is compared against that of the Fourier-Mellin approach.

II Background

The F-histogram method was proposed by Matsakis in [1]. The relative position of a 2D object A (argument) from object B (referent) is captured by the function $F^{AB} (\mathbb{R} \rightarrow \mathbb{R}_+)$ where $F^{AB}(\theta)$ is the total scalar resultant of elementary forces that support the proposition “ A is in direction θ of B ”. These forces are exerted by the points of A on those of B , and each tends to move B in direction θ (Figure 1(a)). Let d be the distance between points in A and B , and d^r is the elementary force between the points, then F is denoted by F_r . F_0 and F_2 are called the *constant* and *gravitational* F-histograms, respectively. F_0 provides a global view of the situation; the closest parts and the farthest parts of the objects are considered equally, whereas F_2 focuses on the closest parts. We employ F_0 in this work.

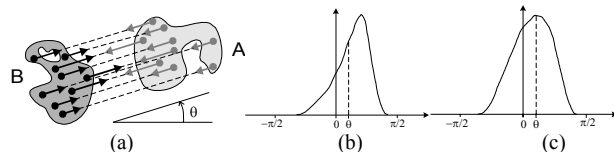


Figure 1 (a) $F^{AB}(\theta)$ as scalar resultant of forces (black arrows), (b) and (c) are F_0 and F_2 histograms.

The F-histogram method is built on solid mathematical foundations that include a set of geometric properties *stre* (stretch), *dil* (dilation), and *rot* (rotation); additionally, the force histogram is translational invariant. These properties allow F-histograms to be used as affine-invariant descriptors for spatial relations between objects [12]. Let AB be an object pair represented by F^{AB} . Let $A'B'$ be the results of affine-transformations (rotation, scaling, translation, and stretching) on AB . Under ideal image condition (2D objects, vector data, orthographic projection), we can match $rot(dil(stre(F^{AB})))$ to $stre(F^{A'B'})$ using the fuzzy histogram similarity measure $\sigma = \mu_s(rot(dil(stre(F^{AB}))), stre(F^{A'B'}))$, where $\mu_s(h_1, h_2) = \frac{\sum \min(h_1(\theta), h_2(\theta))}{\sum \max(h_1(\theta), h_2(\theta))}$, and $\sigma \in [0, 1]$. We use a simplified form of $\mu(F^{AB}, F^{A'B'})$ proposed in [12] where parameters used by *stre* (t &

t'), *dil* (ℓ) and *rot* (ρ) are calculated from F-histograms (ℓ and ρ) or selected from several possible values (t and t') [12]. The set of (ρ, ℓ, t, t') that maximizes σ is denoted as (R, L, T, T') and is returned as $\mu(F^{AB}, F^{A'B'}) = (\sigma, R, L, T, T')$.

In order to assess the value of our approach, we compare it to the main image-based matching technique, the Fourier-Mellin (FM) method. Let f_1 and f_2 be images, where f_2 is a scaled, rotated and translated variant of f_1 ; F_1 and F_2 are the Fourier spectral of f_1 and f_2 . The FM method is based on *cross-power spectrum* (CPS) which, in the absence of noise, reduces to phase difference:

$$CPS(F_1, F_2) = \frac{F_1(u, v)}{|F_1(u, v)|} \cdot \frac{F_2^*(u, v)}{|F_2^*(u, v)|} = e^{j(\phi_1(u, v) - \phi_2(u, v))} \quad (1)$$

The phase difference is related to the Fourier transform translation property where $f_2(x, y) = f_1(x - x_0, y - y_0)$ and $F_2(u, v) = F_1(u, v) e^{-j2\pi(u x_0 + v y_0)}$, such that $e^{-j2\pi(u x_0 + v y_0)} = e^{-j(\phi_1(u, v) - \phi_2(u, v))}$. The inverse Fourier transform of $e^{-j2\pi(u x_0 + v y_0)}$, ideally, is zero everywhere except for a δ -function (impulse) located at (x_0, y_0) . Rotational and scale invariance is obtained by first applying the *log-polar* transformation to the Fourier magnitudes M_1 and M_2 to obtain translation relationship in log-polar coordinate system. The FM method has no provisions for handling image stretching due to tilt angle difference. Details on the FM method can be found in [4][5].

III Scene Matching Based On Object Spatial Relationships (SMOSAR)

III.1 Object Spatial Relationships as Image Descriptors

Let S and S' be scene images, $S = \{O_i | i = 0, \dots, N\}$ and $S' = \{O_j | j' = 0, \dots, N'\}$; O_i and O_j are *segmented* and *uniquely labeled* objects. We assume $S \subseteq S'$ ($N \leq N'$), i.e. $\forall O_i \in S, \exists O_j \in S'$ such that object correspondence $O_i \rightarrow O_j$ is true. We denote S and S' the *template* and *target* images. The spatial relationship between two objects in an image is assessed using the F-histogram method. F-histograms from all legal object pairs in the image are used as *image descriptors*.

Let $O_a O_b$ be an object pair where O_a is the *argument* and O_b is the *referent* objects observing the constraints $a < b$ (image S) and $a' \neq b'$ (image S'). Thus, S and S' have $N(N-1)/2$ and $N'(N'-1)$

object pairs. Let F^{ab} and $F^{c'd'}$ be the F-histograms of $O_a O_b$ and $O_c O_d$. The matching proposition $F^{ab} \rightarrow F^{c'd'}$ supports the object matching of $O_a \rightarrow O_c$ and $O_b \rightarrow O_d$. The matching proposition $F^{ba} \rightarrow F^{d'c'}$ supports the same object matching as $F^{ab} \rightarrow F^{c'd'}$. The constraint $a < b$ in image S eliminates this unnecessary redundancy. We denote object pairs $O_c O_d$ and $O_d O_c$ as *duals*, and their F-histograms $F^{c'd'}$ and $F^{d'c'}$ are identical in shape and magnitude, but differ in phase by 180° . In what follows, we convert the double index (a,b) into a linear index as described in [13][16].

Let Q and Q' be the numbers of legal object pairs in S and S' , $Q = N(N-1)/2$ and $Q' = N'(N'-1)$. Then F-histogram sets $H = \{F^{(k)} | k = 0, \dots, Q-1\}$ and $H' = \{F^{(k')} | k' = 0, \dots, Q'-1\}$ are the image descriptors for S and S' , respectively. These image descriptors are used as *matching elements* in our scene matching algorithm. No further image operation is to be performed on S and S' .

III.2 Image Descriptors Matching

Each potential match of a histogram from one view to the other provides a 5 dimensional vector that we use to build object correspondence. Let $R_H = H \times H'$ whose elements are F-histogram pairs $\{F^{(k)}, F^{(k')}\}$ representing $F^{(k)} \rightarrow F^{(k')}$. Now, $\mu(F^{(k)}, F^{(k')})$ gives a 5-tuple output (σ, R, L, T, T') . Define $R_F = \mu(R_H)$ as a set of all 5-tuple outputs from all $F^{(k)} \rightarrow F^{(k')}$ in R_H . Suppose S and S' capture the same scene and there is a bijective object mapping between the two images. Let $F^{(m)}, F^{(n)} \in H$, and $F^{(r)}, F^{(s')} \in H'$. $F^{(m)} \rightarrow F^{(r)}$ and $F^{(n)} \rightarrow F^{(s')}$ are evaluated as $\mu(F^{(m)}, F^{(r)}) = (\sigma_{mr}, R_{mr}, L_{mr}, T_{mr}, T'_{mr})$ and $\mu(F^{(n)}, F^{(s')}) = (\sigma_{ns}, R_{ns}, L_{ns}, T_{ns}, T'_{ns})$. From [12], $(\sigma_{mr}, R_{mr}, L_{mr}, T_{mr}, T'_{mr}) = (\sigma_{ns}, R_{ns}, L_{ns}, T_{ns}, T'_{ns})$ if $F^{(m)} \rightarrow F^{(r)}$ and $F^{(n)} \rightarrow F^{(s')}$ are correct histogram correspondences (ideal conditions).

The outputs of $\mu(F^{(m)}, F^{(r)})$ and $\mu(F^{(n)}, F^{(s')})$ can be viewed as vectors. The distance between vectors reflects the agreement of $F^{(m)} \rightarrow F^{(r)}$ and $F^{(n)} \rightarrow F^{(s')}$. Since σ, R, L, T , and T' have different value ranges and units, we normalize them as in [13][16]. We define $\bar{x}_{kk'} = [\hat{\sigma}_{kk'}, \hat{R}_{kk'}, \hat{L}_{kk'}, \hat{T}_{kk'}, \hat{T}'_{kk'}]^t$, and $X = \{\bar{x}_{kk'}\}$ for $k = 0, \dots, Q-1$ and $k' = 0, \dots, Q'-1$.

III.3 F-Histogram Map (FMAP)

For $S \subseteq S'$, $S \rightarrow S'$ is defined by a *one-to-one* F-histogram map (FMAP) from H to H' , $\text{FMAP} \subset R_H$. FMAP is bijective for $S = S'$ as illustrated in Figure 2. Here, S and S' capture the same objects ($N = N' = 3$, $Q = 3$, $Q' = 6$) from different orientations. The objects are labeled consistently (O_i in S actually corresponds to O_i in S'). Here, $R_H = H \times H'$ contains 18 possible histogram correspondences. Not all subsets of R_H with Q elements can be used as FMAP. Even though we need to include the dual correspondences in H' , we cannot utilize both primary and dual element in the match [13].

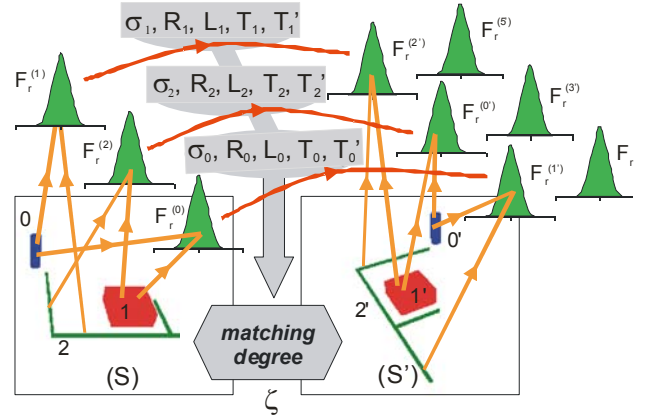


Figure 2 Histogram matching and the calculation of FMAP matching degree (ζ).

The heart of the algorithm to construct a good object correspondence is to search for the proper histogram assignments, i.e., find a good FMAP. In our approach, the first element $F^{(k1)} \rightarrow F^{(k1')}$, $F^{(k1')}$ is one of Q' F-histograms in H' . For the next element $F^{(k2)} \rightarrow F^{(k2')}$, $F^{(k2')}$ is one of $(Q' - 2)$ remaining F-histograms in H' ; the dual of $F^{(k1')}$ cannot be the matching candidate for $F^{(k2)}$ (FMAP restriction on histogram duals). Thus, total $\text{FMAP} = \prod_{i=1}^Q (Q' - 2(i - 1))$ or $((Q'/2)! / ((Q'/2) - Q)!) 2^Q$. For images with 5 objects ($N = N' = 5$, $Q = 10$, $Q' = 20$) we have 3,715,891,200 FMAPs! Even by excluding the dual histograms in H' (essentially assuming consistent object labeling), the search space is reduced only to $((Q'/2)! / ((Q'/2) - Q)!)$. An exhaustive search over the reduced search space is denoted as *semi-exhaustive* search. For images with 5 objects, the semi-exhaustive search implemented in ANSI-C takes about 17 seconds on a 2 GHz Pentium-M laptop computer with 1 GB of RAM. The full-exhaustive search takes about $17 \times 2^Q \approx$

17,000 seconds. Clearly, some efficient method is required to reduce the search time. In this paper, we exploit the mode seeking ability of the Possibilistic C-Means for this search.

III.4 FMAP Matching Degree

The FMAP matching degree is a weighted average:

$$\zeta = \frac{w_\sigma \cdot \zeta^\sigma + w_R \cdot \zeta^R + w_L \cdot \zeta^L + w_T \cdot \zeta^T}{w_\sigma + w_R + w_L + w_T} \quad (2)$$

$\zeta \in [0, 1]$, where $w_i \in [0, 1]$ are weights for each features $\zeta^\sigma, \zeta^R, \zeta^L$, and ζ^T . These features represent the degree of consistency of σ, R, L, T and T' from all $F^{(k)} \rightarrow F^{(k')} (\mu(F^{(k)}, F^{(k')}))$ defined in FMAP. For a perfect FMAP, under ideal conditions, (σ, R, L, T, T') should be identical for all $F^{(k)} \rightarrow F^{(k)}$; $\zeta^\sigma = \zeta^R = \zeta^L = \zeta^T = 1$. The ideal conditions do not apply to real world images (raster data, 3-D, perspective). So $\zeta^\sigma, \zeta^R, \zeta^L$, and ζ^T may be < 1 . These values are calculated using appropriate averages [13]. The FMAP matching degree ζ is illustrated in Figure 2 for $N = N' = 3$ with $\text{FMAP} = \{F^{(0)} \rightarrow F^{(0')}, F^{(1)} \rightarrow F^{(1')}, F^{(2)} \rightarrow F^{(2')}\}$ or $\{0', 1', 2'\}$.

III.5 FMAP Generator (FMG) – Nearest Neighbor Method

We define a neighborhood in X centered on \bar{y} (seed point). Our approach is to grow the FMAP using a forward sequential search of the nearest neighbors starting from \bar{y} as shown below:

1. Let FMAP be an empty 1-D array with Q elements.
2. Calculate $d(\bar{y}, \bar{x}_{kk'})$ for all $\bar{x}_{kk'}$ in X .
3. Sort $d(\bar{y}, \bar{x}_{kk'})$ in ascending order. Let $\bar{x}_{kk'}^{(h)}$ be the sorted vector where h is the sorted order, $h = 0, \dots, (QQ'-1)$. Initialize $h = 0$.
4. **Do While** FMAP is not full.
 - 4.1 **If** $\bar{x}_{kk'}^{(h)} (F^{(k)} \rightarrow F^{(k')})$ satisfies FMAP properties **Then**
 - 4.1.1 Add $F^{(k)} \rightarrow F^{(k')}$ to FMAP, i.e., $\text{FMAP}[k] = k'$. **End If (4.1)**
 - 4.2 $h \leftarrow h + 1$
- End While (4)**

III.6 FMAP Search Using PCM

As discussed before, if $F^{(m)} \rightarrow F^{(r')}$ and $F^{(n)} \rightarrow F^{(s')}$ are correct histogram correspondences, then their feature vectors $[\hat{\sigma}_{mr'}, \hat{R}_{mr'}, \hat{L}_{mr'}, \hat{T}_{mr'}, \hat{T}'_{mr'}]^t$ and $[\hat{\sigma}_{ns'}, \hat{R}_{ns'}, \hat{L}_{ns'}, \hat{T}_{ns'}, \hat{T}'_{ns'}]^t$ should be identical (under ideal conditions), or close. Therefore, it is reasonable to expect that dense regions in the feature space have good potential to produce good FMAPs. The “mode-seeking” property of PCM makes it a good

tool to find dense regions. We view dense regions as clusters, each with its own cluster prototype. Unlike FCM, PCM is not sensitive toward the choice of the number of clusters; if the number of natural clusters is unknown, users can “over-specify” the cluster number and, upon convergence, multiple cluster prototypes may refer to the same “natural” cluster. During each iteration of PCM, we generate an FMAP from each cluster using the FMG algorithm with the current cluster prototype used as the seed point \bar{y} . We keep the FMAP having the highest ζ . Cluster prototypes in PCM are initialized using K vectors $\bar{x}_{kk'}$ in X with the highest $\hat{\sigma}_{kk'}$; K is the number of clusters in PCM. Thus, in the first iteration, $F^{(k)} \rightarrow F^{(k')}$ (represented by $\bar{x}_{kk'}$) is the first element added to the FMAP. This initialization technique is motivated by our observation that the correct correspondence $F^{(k)} \rightarrow F^{(k')}$ tends to produce high $\hat{\sigma}_{kk'}$. PCM is terminated if one of the following convergence criteria is met: 1) maximum number of iterations (t_{max}) is exceeded, 2) all cluster prototypes are relatively constant (within ϵ) for the last I_ϵ iterations.

Once an FMAP is selected, we convert it to the corresponding mapping between object labels (OMAP). Confidence values for object correspondence are given in an $N \times N'$ matrix called OCCM ($N = |S|, N' = |S'|$) with OC_{ij} be a matrix element of confidence value for object correspondence $O_i \rightarrow O_{j'}$, $O_i \in S, O_{j'} \in S'$. Histograms $F^{(k)}$ and $F^{(k')}$ represent object pairs $O_a O_b$ and $O_a' O_{b'}$. FMAP element $F^{(k)} \rightarrow F^{(k')}$ implies $\{O_a, O_b\} \rightarrow \{O_{a'}, O_{b'}\}$ which results in four possible object correspondences: $\{O_a \rightarrow O_{a'}, O_b \rightarrow O_{b'}, O_a \rightarrow O_{b'}, O_b \rightarrow O_{a'}\}$. In [13], we showed that each object correspondence can be supported by up to $(N - 1)$ histogram correspondences $F^{(k)} \rightarrow F^{(k')}$, each contributing its $\sigma_{kk'}$ to make up the object correspondence confidence. Therefore, $\sigma_{kk'}$ from $F^{(k)} \rightarrow F^{(k')}$ is added to the four OCCM elements ($OC_{aa'}, OC_{ab'}, OC_{ba'},$ and $OC_{bb'}$) that it supports. Normalizing OCCM with $(N - 1)$ brings the confidence values to $[0, 1]$. If a perfect FMAP is obtained, all the correct object correspondences will be supported by exactly $(N - 1)$ histogram correspondences. Indeed, what is interesting to note is that even if some mistakes are made on the FMAP assignments, the correct object maps may still be

identified thanks to this redundancy of information. Many experiments and further details are found in [13][16].

IV Experimental Results & Discussions

IV.1 Data and Methods

In Table 1, four image groups (Group1 – Group4) are divided into two data sets: *synthetic* and *real*. Images in a group capture the same objects from different orientations. *Synthetic* set contains 28 scene matching problems and *real* set contains 39 problems. Synthetic images in Group1 (created using computer graphics) replicate real powerplant scenes in Group2. Groups 2 and 3 are created by applying the *pseudo-intensity* filter [9] to the range data from laser radar (LADAR) provided by the Naval Air Warfare Center (NAWC), China Lake, CA. Group4 images are captured using a digital camera. Images in Groups 2 and 3 are hand-segmented, images in Group 4 are automatically segmented using Otsu-like thresholding. Identical cans are used in Group 4. Image examples are shown in Figure 3.

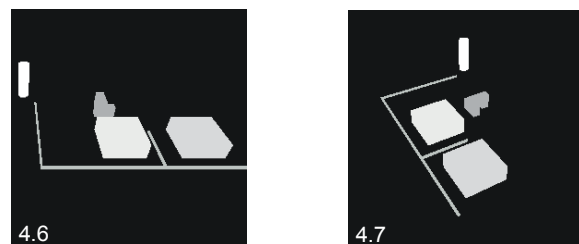
Table 1 Scene data

| | Data Set | Scenes | # Objs | # Imgs | # Img Pair |
|---|-----------|------------|--------|--------|------------|
| 1 | Synthetic | Powerplant | 5 | 8 | 28 |
| 2 | Real | Powerplant | 5 | 3 | 3 |
| 3 | Real | Powerplant | 4 | 6 | 15 |
| 4 | Real | Cans | 5 | 7 | 21 |

We use the constant force histogram (F_0) as we did in [12]. For PCM, we use 5 clusters, $\eta_j = 1 (\forall j)$, $q = 2$, $I_\varepsilon = 10$, $\varepsilon = 0.01$, $t_{max} = 100$, with *Euclidean* distance. The following performance measures were calculated: M1 (number of correct histogram correspondences), M2 (perfect FMAP), M3 (number of correct object correspondences), and M4 (perfect OMAP). Results are compared to the Fourier-Mellin method. Weights $W_\sigma = 1.0$, $W_R = 1.0$, $W_L = 0.821$, and $W_T = 0$ in Equation (2) are estimated from individual features performance using semi-exhaustive FMAP search on the synthetic data set [13]. Experiments are performed on a 2 GHz Pentium-M laptop computer with 1 GB of RAM.

Corresponding objects in images within a group are labeled identically to identify correct histogram and object correspondences easily. However, non-consistent object labeling does not affect SMOSAR performance as shown in Figure 4. Here, we found the perfect FMAP = $\{F^{(0)} \rightarrow F^{(7)}, F^{(1)} \rightarrow F^{(5)}, F^{(2)} \rightarrow F^{(9)}, F^{(3)} \rightarrow F^{(2)}, F^{(4)} \rightarrow F^{(0)}, F^{(5)} \rightarrow F^{(10)}\}$ with ζ

= 0.79 as the best FMAP. We found OMAP = $\{O_0 \rightarrow O_2, O_1 \rightarrow O_0, O_2 \rightarrow O_3, O_3 \rightarrow O_1\}$ which is the correct object map. More details about object labeling independence are found in [13].



Group1: synthetic powerplant images



Group2: real powerplant images with 5 objects (LADAR)



Group3: real powerplant images with 4 objects (LADAR)



Group4: real images of identical tuna cans

Figure 3 Examples of images from Group1, Group2, Group3, and Group4.

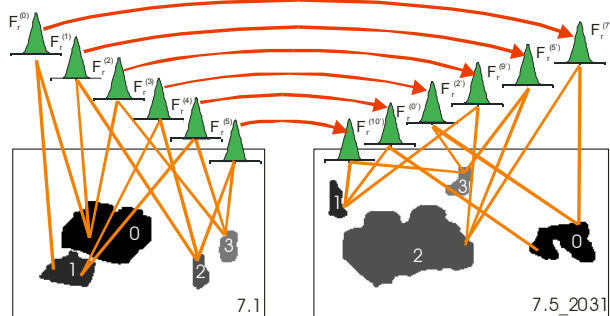


Figure 4 F-histogram map with inconsistent object labels.

IV.2 Results

The experimental results on synthetic and real world data using the PCM, semi-exhaustive search and FM method are given in Table 2. For synthetic scenes, 100% of perfect OMAP is recovered by both PCM method and semi-exhaustive search (M3 = M4 = 1.0). However, the PCM recovers only 43%

