Pedestrian Detection from a Moving Vehicle

D. M. Gavrila

Presented by Chia-Chih Chen

Overview (1)

- Goal:
  A working system for pedestrian detection on-board a moving vehicle
- Difficulties:
  1) highly cluttered BG
  2) wide range of object appearances
  3) appear rather small in low-resolution images
  4) cameras are on a moving platform
  5) hard real-time requirements for vehicle application

Overview (2)

- Procedure:
  Step 1: Lock onto candidate solutions
  - shape matching using DT
  - hierarchical template structure

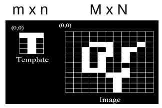
  Step 2: Verification
  - dismiss false-positives using RBF-based classification
  - introduce bias towards samples close to imaginary target using incremental boostrapping

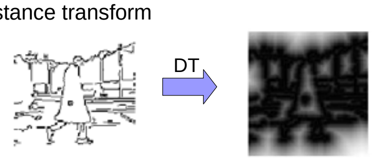
Basic Idea

- Our template T is an edge-map.
- Create edge map of image. This is our feature-image I.
- Slide T over I, until it somehow delivers the best match.

Chamfer Matching – Chamfer DT (1)

- Binary correlation
  - computational expensive
  - sensitive to noise
- Solution
  - smoothen the edges of the edge-image using distance transform



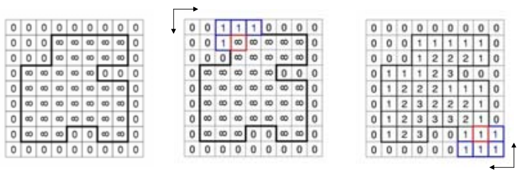


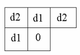
Chamfer Matching – Chamfer DT (2)

- Definition
  - converts a binary image into a intensity image
  - each pixel value denotes the Euclidean distance to the nearest feature pixel
- Properties
  - distance transform is a global transformation
  - the distance can be approximated using integer arithmetic in raster-scan faction

Chamfer Matching – Chamfer DT (3)

- Procedure
  - *Initialization* - *FW scan* - *BW scan*





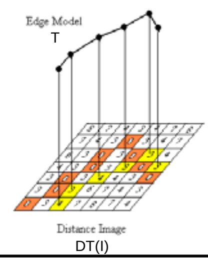
$V_{i,j}$ = min ($v_{i-1,j-1}+d_2$, $v_{i-1,j}+d_1$, $v_{i-1,j+1}+d_2$, $v_{i,j-1}+d_1$, $v_{i,j}$)

Chamfer Matching – Average Chamfer Distance

- Relevant T is translated & positioned over *DT(I)*
- *D(T, I)* is determined by the pixel values of *DT(I)* which lie under the pixels of translated T
- *T considered* match when *D(T,I)<θ*

$$D_{Chamfer}(T,I) = \frac{1}{|T|}\sum_{t \in T} \min_{i \in I} \|t - i\|$$

Ex: D(T, I) = 1/6*(4+3+4+3+3+3) = 3.33



Chamfer Matching – Template Hierarchy (1)

- Objective
  To organize templates hierarchically so that matching can be conducted efficiently
- Approach
  - group similar templates together and represent them by a "prototype" template and a distance
  - T are moved between groups so that *E* is minimized

$$E = \sum_{k=1}^{K} \max_{\mathbf{t}_i \in S_k} D(\mathbf{t}_i, \mathbf{p}_k^*)$$

Chamfer Matching – Template Hierarchy (2)

$t_1, t_2, \dots, t_N$
Partition size K
$D(t_i, p_k^*)$

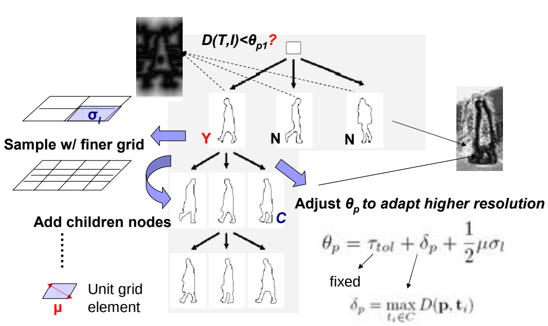
→ **Partitional Clustering** →

$\arg\min_{S_1, S_2, \dots, S_k} E$

K groups:
$\{S_1, S_2, \dots, S_k\}$
K prototypes:
$\{p_1, p_2, \dots, p_k\}$

- Stopping criterion: minimum E-value
  - tight grouping
  - lowers the distance threshold for matching
    ➔ decrease the number of locations to be considered

Chamfer Matching – Template Hierarchy (3)

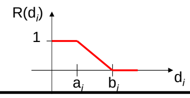


Verification

- Objective
  Verify candidate solutions found in the detection phase
- Given
  Candidate solutions w/ T_{id} and corresponding image locations
- Procedures
  - extract bounding box of the template matched
  - normalize the window for scale
  - employ RBF-based classification and incremental boostrapping

Verification
RBF network (1)

- RBF centers
  Apply agglomerative clustering in feature space of training data to find centers, $\vec{G}_i,\ i = 1,\ldots,N$ which are from K classes c_k, $k = 1 \ldots K$ (K = 1 in our case)
- Classify an unknown instance **ε**
  - the distance from **ε** to each RBF centers is calculated by $d_i = \|\varepsilon - G_i\|$
  - d_i is further transformed by R(•), which is controlled by a_i and b_i



Verification
RBF network (2)

- Classify an unknown **ε**
  - define by individual class likelihood: $\bar{P}_k = \sum_{i=1,\, c_i=k}^{N} R(d_i)$

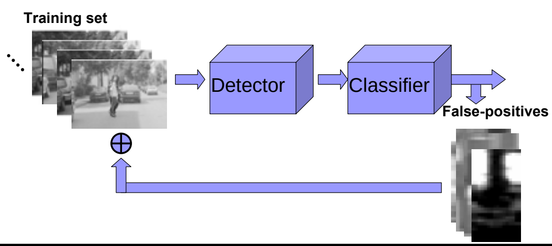
    total class likelihood: $S_{\bar{P}} = \sum_{k=1}^{K} \bar{P}_k.$

    normalized likelihood: $P_k = \begin{cases} \bar{P}_k / S_{\bar{P}} & \text{if } S_{\bar{P}} > 1 \\ \bar{P}_k & \text{if } S_{\bar{P}} \leq 1 \end{cases}$ $P_{reject} = 1 - S_{\bar{P}}$

  - **ε** is assigned to the class with highest P_k
  - **ε** is rejected if 1) P_{reject} > all P_k
    2) highest P_k is lower than a threshold t

Verification
- Incremental Boostrapping

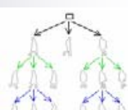
- Objective
  Train the RBF classifier to be more discriminant on the imaginary border of pedestrian class



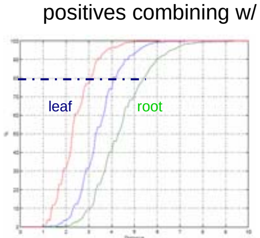
Results (1)

- Settings
  - 1250 distinct pedestrian shapes
  - 3-level hierarchy, 900 templates at leaf per scale
  - 5 scales were used
- Implementation improvements
  - oriented edge features
  - template subsampling
  - multi-stage segmentation thresholds
  - ground plane constraints

Results (2)

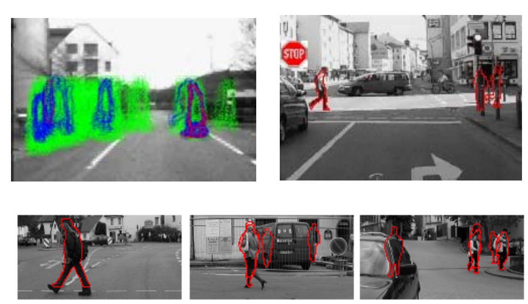


- Detection rates
  - 60-90% using Chamfer System alone
  - detect 85% pedestrians, conceding 10% false-positives combining w/ incremental boostrapping



- Cumulative distribution
  - average Chamfer distance values from root to the correct leaf
  - used to determine θ_p at different T levels

Results (3)



Conclusions

- Advantages

  coarse-to-fine approach:
  WxHxK ➔ reduce WxH, reduce K
- Problems

  - depends on reasonable segmentation
  - effective at limited scales
  - partial occluded pedestrian, night scenes
- Improvements
  - multi-modal shape tracker
  - SVM

References

- DM Gavrila, “Pedestrian detection from a moving vehicle," Proc. 6th European Conf. on Computer Vision, 2000
- U. Kressel, F. Lindner, C. W¨ohler, and A. Linz. "Hypothesis verification based on classification at unequal error rates," Proc. of ICANN, 1999
- C. Papageorgiou and T. Poggio. "A pattern classification approach to dynamical object detection," Proc. of the International Conference on Computer Vision, Kerkyra, Greece, 1999