Progressive Learning for Interactive Surveillance Scenes Retrieval

Jérôme Meessen

Multitel Video Analysis workshop, May 30, 2008

Communication and Remote Sensing Lab.
Ecole polytechnique de Louvain
Outline

1. Introduction

2. Multiple-instance SVM for Interactive Retrieval
   - Overview
   - Multiple-instance framework and adaptive similarity measure
   - Active Learning
   - Experiments

3. Conclusions
1. Introduction

2. Multiple-instance SVM for Interactive Retrieval

3. Conclusions
Context and objective

Searching surveillance video

- Increasing interest
- Mainly for retrieving forensic evidences
- Today: need a priori knowledge or metadata about target events
- No existing fully content-based solution

Objective

- User queries, examples
- Similarity to stored content
- Results presentation
- User feedback

Adapted to surveillance video frames
Context and objective

Searching surveillance video

- Increasing interest
  - Mainly for retrieving forensic evidences
  - Today: need a priori knowledge or metadata about target events
  - No existing fully content-based solution
Searching surveillance video

- Increasing interest
- Mainly for retrieving forensic evidences
- Today: need a priori knowledge or metadata about target events
- No existing fully content-based solution
Context and objective

Searching surveillance video

- Increasing interest
- Mainly for retrieving forensic evidences
- Today: need a priori knowledge or metadata about target events
- No existing fully content-based solution
Context and objective

Searching surveillance video

- Increasing interest
- Mainly for retrieving forensic evidences
- Today: need a priori knowledge or metadata about target events
- No existing fully content-based solution
Context and objective

Searching surveillance video
- Increasing interest
- Mainly for retrieving forensic evidences
- Today: need a priori knowledge or metadata about target events
- No existing fully content-based solution

Objective
- An interactive content-based image retrieval system:
  - User queries, examples
  - Similarity to stored content
  - Results presentation
  - User feedback
- Adapted to surveillance video frames
Context and objective

Searching surveillance video
- Increasing interest
- Mainly for retrieving forensic evidences
- Today: need a priori knowledge or metadata about target events
- No existing fully content-based solution

Objective
- An interactive content-based image retrieval system:
  - User queries, examples
  - Similarity to stored content
  - Results presentation
  - User feedback
- Adapted to surveillance video frames
Context and objective

Searching surveillance video
- Increasing interest
- Mainly for retrieving forensic evidences
- Today: need a priori knowledge or metadata about target events
- No existing fully content-based solution

Objective
- An interactive content-based image retrieval system:
  - User queries, examples
  - Similarity to stored content
  - Results presentation
  - User feedback
- Adapted to surveillance video frames
Outline

1. Introduction

2. Multiple-instance SVM for Interactive Retrieval
   - Overview
   - Multiple-instance framework and adaptive similarity measure
   - Active Learning
   - Experiments

3. Conclusions
System overview

An iterative process:
- User labels frames: “relevant” vs “irrelevant”
- Training of a soft margin SVM, including relevance feedback
- Class prediction
- Selection of new examples to present: active learning

[MILES06], [Hastie04], [Rui98], [Cox00], [Smeulders00],...
Issues raised by (among others) surveillance context

1. Multiple-instance problem
   - Labels to frames
   - Features for foreground objects
     ⇒ Ambiguity on the objects and the features

2. Rare events
   ⇒ Pay attention to positive examples

3. User typically labels only a few examples at each round
Issues raised by (among others) surveillance context

1. Multiple-instance problem
   - Labels to frames
   - Features for foreground objects
     ⇒ Ambiguity on the objects and the features

2. Rare events
   ⇒ Pay attention to positive examples

3. User typically labels only a few examples at each round
Issues raised by (among others) surveillance context

1. Multiple-instance problem
   - Labels to frames
   - Features for foreground objects
     ⇒ Ambiguity on the objects and the features

2. Rare events
   ⇒ Pay attention to positive examples

3. User typically labels only a few examples at each round
Issues raised by (among others) surveillance context

1. Multiple-instance problem
   - Labels to frames
   - Features for foreground objects
     ⇒ Ambiguity on the objects and the features

2. Rare events
   ⇒ Pay attention to positive examples

3. User typically labels only a few examples at each round
Issues raised by (among others) surveillance context

1. Multiple-instance problem
   - Labels to frames
   - Features for foreground objects
     ⇒ Ambiguity on the objects and the features

2. Rare events
   ⇒ Pay attention to positive examples

3. User typically labels only a few examples at each round
**MILES: a framework for Multiple-instance SVM classification**

Frames \(= \text{Bags} (B_i)\), Objects \(= \text{instances} \ (x^k)\)

2 steps:

1. Map the bags in the *instance space*, using similarity between the training bags and their instances:
   
   \[
   s(B_i, x^k) = \max_j \exp \left( -\frac{d(x^k, x^j)}{\sigma^2} \right)
   \]

   Bag coordinates:
   
   \[
   m(B_i) = [s(B_i, x^1), s(B_i, x^2), \ldots, s(B_i, x^n)]
   \]

2. Build a linear classifier: \(y = \text{sign}(w^Tm + b)\)

Frames = Bags ($B_i$), Objects = instances ($x^k$)

2 steps:

1. Map the bags in the instance space, using similarity between the training bags and their instances:

   \[ s(B_i, x^k) = \max_j \exp \left( -\frac{d(x^k, x_j^i)}{\sigma^2} \right) \]

   Bag coordinates:
   \[ m(B_i) = [s(B_i, x^1), s(B_i, x^2), \ldots, s(B_i, x^n)] \]

2. Build a linear classifier: \( y = \text{sign}(w^T m + b) \)

Frames = Bags \( (B_i) \), Objects = instances \( (x^k) \)

2 steps:
1. Map the bags in the instance space, using similarity between the training bags and their instances:
   
   \[
   s(B_i, x^k) = \max_j \exp \left( -\frac{d(x^k, x^j_i)}{\sigma^2} \right)
   \]
   
   Bag coordinates:
   
   \[
   m(B_i) = [s(B_i, x^1), s(B_i, x^2), \ldots, s(B_i, x^n)]
   \]

2. Build a linear classifier:
   \[
   y = \text{sign}(w^T m + b)
   \]

**MILES: a framework for Multiple-instance SVM classification**

Frames = Bags \( (B_i) \), Objects = instances \( (x^k) \)

2 steps:

1. **Map the bags in the instance space**, using similarity between the training bags and their instances:
   - \( s(B_i, x^k) = \max_j \exp \left( -\frac{d(x^k, x^j)}{\sigma^2} \right) \)
   - Bag coordinates:
     \[ m(B_i) = [s(B_i, x^1), s(B_i, x^2), \ldots, s(B_i, x^n)] \]

2. **Build a linear classifier**:
   \[ y = \text{sign}(w^T m + b) \]

Maximum margin classifier with slack variables
Training the SVM

SVM: Look for a hyperplane with largest margin between $l^+$ positive and $l^-$ negative ex:

$$
\begin{align*}
\min_{\lambda} & \quad \sum_{k=1}^{n} |w_k| + \mu \sum_{i=1}^{l^+} \xi_i + (1 - \mu) \sum_{j=1}^{l^-} \eta_j \\
\text{s.t.} & \quad \left[ w^T m_i + b \right]_{l^+} + \xi_i \geq 1 \\
& \quad \left[ w^T m_j - b \right]_{l^-} + \eta_j \geq 1
\end{align*}
$$

Defines non-zero $w^*$ and their support instances.

Training the SVM

SVM: Look for a hyperplane with largest margin between $I^+$ positive and $I^-$ negative ex:

$$\min \lambda \sum_{k=1}^{n} |w_k| + \mu \sum_{i=1}^{l^+} \xi_i + (1 - \mu) \sum_{j=1}^{l^-} \eta_j$$

s.t. $[w^T m^+_i + b] + \xi_i \geq 1$ and $-[w^T m^-_j + b] + \eta_j \geq 1$

Training the SVM

SVM: Look for a hyperplane with largest margin between $I^+$ positive and $I^-$ negative ex:

\[
\min \lambda \sum_{k=1}^{N} |w_k| + \mu \sum_{i=1}^{I^+} \xi_i + (1 - \mu) \sum_{j=1}^{I^-} \eta_j
\]

s.t. $[w^T m_i^+ + b] + \xi_i \geq 1$ and $-[w^T m_j^- + b] + \eta_j \geq 1$

$\Rightarrow$ Defines non-zero $w_k^*$ and their support instances.

Training the SVM

SVM: Look for a hyperplane with largest margin between $l^+$ positive and $l^-$ negative ex:

$$\min \lambda \sum_{k=1}^{n} |w_k| + \mu \sum_{i=1}^{l^+} \xi_i + (1 - \mu) \sum_{j=1}^{l^-} \eta_j$$

s.t. $[w^T m^+_i + b] + \xi_i \geq 1$ and $-[w^T m^-_j + b] + \eta_j \geq 1$

$\Rightarrow$ Defines non-zero $w_k^*$ and their support instances.


Parameters

- $\mu \cdot l^+ = (1 - \mu)l^-$
Weighted distance between instances

\[ d(x^k, x^j) : \]
- Each visual feature \( v \) has a weight: \( u_v \),
- Computed from the positive and negative class moments \( \mu^+_v, \mu^-_v, (\sigma^+_v)^2 \) and \( (\sigma^-_v)^2 \)
- Problem: computation of these moments in our MI framework?
Weighted distance between instances

In \( d(x^k, x^j_i) \):
- Each visual feature \( v \) has a weight: \( u_v \),
- Computed from the positive and negative class moments \( \mu_v^+, \mu_v^-, (\sigma_v^+)^2 \) and \( (\sigma_v^-)^2 \)
- Problem: computation of these moments in our MI framework?
Weighted distance between instances

\[ d(x^k, x^j) : \]

- Each visual feature \( \nu \) has a weight: \( u_{\nu} \),
- Computed from the positive and negative class moments \( \mu_+^\nu, \mu_-^\nu, (\sigma_+^\nu)^2 \) and \( (\sigma_-^\nu)^2 \)
- Problem: computation of these moments in our MI framework?
Weighted distance between instances

In \( d(x^k, x^j_i) \):

- Each visual feature \( v \) has a weight: \( u_v \),
- Computed from the positive and negative class moments \( \mu_v^+, \mu_v^-, (\sigma_v^+)^2 \) and \( (\sigma_v^-)^2 \)
- Problem: computation of these moments in our MI framework?

Positive moments

- \( PS \): using support instances
- \( PB \): based on a sub-set of “very positive” instances
- \( PC \): Sequential combination of \( PB \) and \( PS \)
Weighted distance between instances

In $d(x^k, x^j_i)$:

- Each visual feature $v$ has a weight: $u_v$,
- Computed from the positive and negative class moments $\mu_v^+, \mu_v^-, (\sigma_v^+)^2$ and $(\sigma_v^-)^2$
- Problem: computation of these moments in our MI framework?

Positive moments
- $PS$: using support instances
- $PB$: based on a sub-set of “very positive” instances
- $PC$: Sequential combination of $PB$ and $PS$

Negative moments
- $NS$: using support instances
- $NA$: all instances from negative bags
Active learning Strategies

1. “Most ambiguous”: closest to decision plane
2. Exploration of the space: using random selection
3. New strategy
   \[ E_{\text{new}} = \arg \max_i w_i T_m i + b \]
   ...but not identical to training data:
   \[ \max_l \min_j (B_e, x_l, j) < \theta \]
Active learning Strategies

1. “Most ambiguous”: closest to decision plane
2. Exploration of the space: using random selection

\[
\text{Expected positive examples...} \quad \text{arg max}_i w^* T_m i + b
\]

...but not identical to training data:
\[
\max_{i} \min_{j} (B_e, x_l, j) < \vartheta_2
\]
Active learning Strategies

1. “Most ambiguous”: closest to decision plane
2. Exploration of the space: using random selection

New strategy

Expected positive examples...

$$i_e = \arg \max_i w^T m_i + b$$

...but not identical to training data:

$$\max_i \min_j s(B_e, x^{l,j}) < \varphi_2.$$
Combination of AL strategies

- Goal: maximise knowledge acquisition by on-line combination of the strategies
- Existing criterion for the exploration probability [Osugi 05]:
  - Based on the classifier hypothesis change: \( d(h_t, h_{t-1}) = \frac{C_t}{U_t} \)
Combination of AL strategies

- Goal: maximise knowledge acquisition by on-line combination of the strategies
- Existing criterion for the exploration probability [Osugi 05]:
  - Based on the classifier hypothesis change: 
    \[ d(h_t, h_{t-1}) = \frac{C_t}{U_t} \]
Combination of AL strategies

- **Goal:** maximise knowledge acquisition by on-line combination of the strategies
- **Existing criterion for the exploration probability [Osugi 05]:**
  - Based on the classifier hypothesis change: \( d(h_t, h_{t-1}) = \frac{C_t}{U_t} \)

**Proposed criterion**

- Check the similarity between new examples and current hypothesis
- **Ex:** new positive examples are informative if, either:
  - The average \( s^+_{t,i} = \max_k s(B^+_{t,i}, x^k) \ (\forall k \mid w^*_k, t-1 > 0) \) is low
  - The average \( s^-_{t,i} = \max_k s(B^+_{t,i}, x^k) \ (\forall k \mid w^*_k, t-1 < 0) \) is high
Combination of AL strategies

- Goal: maximise knowledge acquisition by on-line combination of the strategies
- Existing criterion for the exploration probability [Osugi 05]:
  - Based on the classifier hypothesis change: \( d(h_t, h_{t-1}) = \frac{C_t}{U_t} \)

**Proposed criterion**
- Check the similarity between new examples and current hypothesis
- Ex: new positive examples are informative if, either:
  - The average \( s^+_{t,i} = \max_k s(B^+_{t,i}, x^k) \) (\( \forall k \ | \ w^*_{k,t-1} > 0 \)) is low
  - The average \( s^-_{t,i} = \max_k s(B^-_{t,i}, x^k) \) (\( \forall k \ | \ w^*_{k,t-1} < 0 \)) is high

**One important advantage!**
- Our criterion is computed before re-training the SVM
  \( \Rightarrow \) Could condition the expensive training of the classifier
Data sets

Synthetic data
- Using 2-D Gaussian distributions
- 3 types of bags:
  - 1 instance in 1 of them: “OR”,
  - 1 instance in each of them: “AND”,
  - only randomly scattered instances: “Random”
Data sets

Synthetic data

- Using 2-D Gaussian distributions
- 3 types of bags:
  - 1 instance in 1 of them: “OR”
  - 1 instance in each of them: “AND”
  - only randomly scattered instances: “Random”

Real-world data: IEEE PETS 2006:

- Scenario 1: Activity in “A”
- Scenario 2: Activity in “A&B”
PETS: Distribution of instances of scenario A

![Image of a train station with marked areas A and B]

![Graph showing instances distribution for Pets4all]
Distance weighting: results on synthetic data

\[ \lambda = 0.2 \] \[ \lambda = 0.3 \] \[ \lambda = 0.4 \]
Distance weighting: results on real data

$\lambda = 0.1$

$\lambda = 0.2$

Conclusion

Feature weighting improves the results

Best performing method: PC+NA
Distance weighting: results on real data

\[ \lambda = 0.1 \]

\[ \lambda = 0.2 \]

**Conclusion**

- Feature weighting improves the results
- Best performing method: \( PC + NA \)
Active learning on synthetic data

10% Pos.

“OR” Negative

Random Negative

30% Pos.
Active learning on real data

Scenario A

Scenario A&B

Conclusion

MostPositive: useful in early stages, quickly underperforming

Proposed transition criterion: globally outperforming others on PETS data
Active learning on real data

Scenario A

Scenario A&B

Conclusion

- **MostPositive**: useful in early stages, quickly underperforming
- Proposed transition criterion: globally outperforming others on PETS data
Outline

1. Introduction
2. Multiple-instance SVM for Interactive Retrieval
3. Conclusions
Summary of the contributions

Classification method adapted to real-life video surveillance constraints

- Multiple-instance framework
- Rarity of the target scenes
- Minimal load on the user
Thanks for your attention!