Abstract

Population of old generation that live alone is growing in most countries. Surveillance systems help them stay home and reduce the burden on the healthcare system. Automatic visual surveillance systems have advantages over wearable devices. They extract features from video sequences and use them for event classification. But these features are dependent on the position of cameras relative to the person. Therefore they need multi-camera for more accuracy that increases cost and complexity. In this paper we propose using silhouette area combined with inclination angle as robust features that can be measured using only one camera with an arbitrary direction. Through rigorous simulations on a publicly available dataset the error rate of the system is found to be less than 1%.

Proposed Method

Three general steps in automatic fall detection systems:
1. Separation of moving objects form stationary background.
2. Extraction of features form the moving objects.
3. Decision making based on the extracted features.

A. Background Separation

- Moving objects should be separated from stationary parts of the image.
- A simple and fast method: Running Gaussian Average. Background value of each pixel is estimated and updated.
  \[ b_I = g_I + (1 - g) b_{I-1} \]  

- \( b_I \) is the background value and \( g_I \) is pixel’s current value for frame \( I \).
- \( g \) is a real number between 0 and 1 that determines the updating speed.
- At each frame if \( |b_I| > T_B \) then the pixel belongs to foreground.
- Noise removal: erosion step, using a disk structuring element with radius of \( I \).
- “Silhouette”: Output of background subtraction method, a binary image. Moving objects are white and stationary parts are black.
- Drawback of Equation (1): unnecessarily updates all pixels. An unwanted “shadow” in the silhouette follows the moving object.
- Shadow size: proportional to motion speed.
- An object becoming stationary \( \rightarrow \) its silhouette will gradually fade out.

B. Feature Extraction

- Commonly used feature for fall detection: person’s orientation.
- Silhouette approximated by an ellipse with same second moment
- Orientation: view-point sensitive, no difference between falling and sitting down.

- Characteristics of a fall:
  1. Suddens increase in motion speed
  2. Surrounding objects occasionally shake
  3. Severely injured person remains motionless on the ground.
- Variation of silhouette area reveals all of the above characteristics
- Sudden movement: Increase in person’s silhouette and shadow area.
- Shaken surroundings: appear in the silhouette and increase the area.
- After fall: stationary person’s silhouette area shrinks to zero.
- Variations of silhouette area: robust to view direction.

C. Classification

- Least Square Support Vector Machines (LS-SVM) for separating “fall” from “walk”.
- Classification in two phases:
  1. Training phase: N training samples of the form \( \{x_i, y_i\} \) where \( x_i \) is p-dimensional input vector and \( y_i \) is labeled with 1 and -1. The optimal hyperplane to separate these samples is found by solving matrix of Equation (2).
  \[
  \begin{bmatrix}
  0 & I^T & C & I^T y \\
  1 & K(x_i, x) & C & 0
  \end{bmatrix}
  \begin{bmatrix}
  \alpha \\
  \beta \\
  \rho
  \end{bmatrix}
  = 0
  \]
- \( \alpha \) is a normal vector of length L and \( \beta \) is the offset of hyperplane. Also, \( \delta \) is vector of labels of length N and \( j \) is a vector of \( 1 \)’s length N. Here \( C \) is margin parameter and \( I \) shows identity matrix. \( K(x_i, x) \) is the kernel function that maps input training samples to higher dimensions (in order to better separability).
- Classification phase: new test vector \( x \) is classified using (3) based on the sign of \( y(x) \).

\[ y(x) = \sum_1^\alpha K(x_i, x) + \beta \]  

Experimental Results

- 130 samples selected from dataset including “falling” and “walking”.
- Each sample is represented with two feature vectors of orientation variations and silhouette area variations.
- Feature-vector calculated in 200 consecutive frames. For noise reduction, one of 5 consecutive elements using median operator was selected to obtain 40-element vector.
- Background separation: empirically selected updating speed of \( \gamma = 0.02 \) and constant foreground threshold of \( T_B=30 \).
- Classification step: margin parameter of C=100 was selected.
- 10-fold cross validation was used.
- Orientation feature vector produces high error rate.

**TABLE I: Classification Results based on Orientation**

<table>
<thead>
<tr>
<th>SVM Kernel</th>
<th>Linear</th>
<th>Polynomial</th>
<th>RBF</th>
<th>Fall (x=1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>95</td>
<td>65</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>TN</td>
<td>95</td>
<td>65</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>FP</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>FN</td>
<td>5</td>
<td>False Positive</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>95.70</td>
<td>95.70</td>
<td>95.70</td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td>90.09</td>
<td>90.09</td>
<td>90.09</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>90.89</td>
<td>90.89</td>
<td>90.89</td>
<td></td>
</tr>
<tr>
<td>Error Rate</td>
<td>9.11</td>
<td>9.11</td>
<td>9.11</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE II: Classification Results based on Silhouette Area**

<table>
<thead>
<tr>
<th>SVM Kernel</th>
<th>Linear</th>
<th>Polynomial</th>
<th>RBF</th>
<th>Fall (x=1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>95</td>
<td>65</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>TN</td>
<td>95</td>
<td>65</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>FP</td>
<td>5</td>
<td>False Positive</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>FN</td>
<td>5</td>
<td>False Positive</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>95.72</td>
<td>95.72</td>
<td>95.72</td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td>90.61</td>
<td>90.61</td>
<td>90.61</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>90.21</td>
<td>90.21</td>
<td>90.21</td>
<td></td>
</tr>
<tr>
<td>Error Rate</td>
<td>9.79</td>
<td>9.79</td>
<td>9.79</td>
<td></td>
</tr>
</tbody>
</table>

Each sample is classified correctly at least by one of the above classifiers.
- Outputs of orientation based classifier and area based classifiers are fed into a third classifier to form a new 2-dimensional classifier, and area based classifier (with polynomial kernel). Figure 4 shows well separation of samples in 2D domain.
- Classification of 130 samples with proposed combined system produced only one False Positive (false alarm). No single fall event was missed \( \rightarrow \) sensitivity = 100%.

**Conclusion**

- Suitable features should be extracted in automatic visual fall detection systems.
- We exploited a drawback in one simple background subtraction method.
- Variations of silhouette area reveals a fall event, independent from view direction.
- Combination of two features (silhouette area and orientation) improved classification results.